



INTERNATIONAL JOURNAL OF CENTRAL BANKING

Asset Purchase Programs and Financial Markets:
Lessons from the Euro Area

Carlo Altavilla, Giacomo Carboni, and Roberto Motto

The Single Resolution Fund and the Credit Default Swap:
What Is the Coasian Fair Price of Their Insurance Services?

Anna Naszodi

Monetary Policy Transmission via Loan Contract Terms
in the United States

Esteban Argudo

International Trade Finance and the Cost Channel of Monetary
Policy in Open Economies

Nikhil Patel

Policy Performance and the Behavior of Inflation Expectations

Eda Gülşen and Hakan Kara

Which Credit Gap Is Better at Predicting Financial Crises?

A Comparison of Univariate Filters

Mathias Drehmann and James Yetman

The Impact of Regime Change on the Influence of the Central
Bank's Inflation Forecasts: Evidence from Japan's Shift
to Inflation Targeting

*Masazumi Hattori, Steven Kong, Frank Packer,
and Toshitaka Sekine*

On the Optimal Labor Income Share

Jakub Growiec, Peter McAdam, and Jakub Mućk



Asset Purchase Programs and Financial Markets: Lessons from the Euro Area <i>Carlo Altavilla, Giacomo Carboni, and Roberto Motto</i>	1
The Single Resolution Fund and the Credit Default Swap: What Is the Coasian Fair Price of Their Insurance Services? <i>Anna Naszodi</i>	49
Monetary Policy Transmission via Loan Contract Terms in the United States <i>Esteban Argudo</i>	85
International Trade Finance and the Cost Channel of Monetary Policy in Open Economies <i>Nikhil Patel</i>	117
Policy Performance and the Behavior of Inflation Expectations <i>Eda Gülşen and Hakan Kara</i>	179
Which Credit Gap Is Better at Predicting Financial Crises? A Comparison of Univariate Filters <i>Mathias Drehmann and James Yetman</i>	225
The Impact of Regime Change on the Influence of the Central Bank's Inflation Forecasts: Evidence from Japan's Shift to Inflation Targeting <i>Masazumi Hattori, Steven Kong, Frank Packer, and Toshitaka Sekine</i>	257
On the Optimal Labor Income Share <i>Jakub Growiec, Peter McAdam, and Jakub Mućk</i>	291

Copyright © 2021 by the Association of the International Journal of Central Banking. All rights reserved. Brief excerpts may be reproduced or translated provided the source is cited. Consult www.ijcb.org for further information.

The *International Journal of Central Banking* is published quarterly (ISSN: 1815-4654). Online access to the publication is available free of charge at www.ijcb.org. Changes to hard copy subscriptions and orders should be placed by e-mail at editor@ijcb.org.

Renewals, claims, address changes, and requests for permission to reprint material from this journal should be addressed to:

International Journal of Central Banking
HS08.20
European Central Bank
Sonnemannstr. 22
D-60640 Frankfurt am Main
Germany
Phone: +4969134495894
E-mail: editor@ijcb.org

The views expressed in this journal do not necessarily represent the views of the Association of the International Journal of Central Banking or any of its members.

ISSN: 1815-4654

International Journal of Central Banking

Board of Directors

Chairman

Richard Clarida, *Federal Reserve Board*

Board Members

Elias Albagli, *Central Bank of Chile*
Ahmed Alkholifeh, *Saudi Arabian Monetary Authority*
David E. Altig, *Federal Reserve Bank of Atlanta*
Kartik Athreya, *Federal Reserve Bank of Richmond*
Jan Marc Berk, *The Nederlandsche Bank*
Claudio Borio, *Bank for International Settlements*
Jan Bruha, *Czech National Bank*
Paul Castillo, *Central Reserve Bank of Peru*
Lillian Cheung, *Hong Kong Monetary Authority*
Laurent Clerc, *Bank of France*
Andrew Colquhoun, *Monetary Authority of Singapore*
Jose Gabriel Cuadra Garcia, *Bank of Mexico*
Francisco G. Dakila Jr., *Bangko Sentral ng Pilipinas*
Michael Dotsey, *Federal Reserve Bank of Philadelphia*
Luci Ellis, *Reserve Bank of Australia*
Rosthom Fadli, *Bank of Algeria*
Eugenio Gaiotti, *Banca d'Italia*
Carlos Garriga, *Federal Reserve Bank of St. Louis*
Joseph Gruber, *Federal Reserve Bank of Kansas City*
Yong Ha, *Reserve Bank of New Zealand*
Philipp Hartmann, *European Central Bank*
Jonathan Heathcote, *Federal Reserve Bank of Minneapolis*
Beverly Hirtle, *Federal Reserve Bank of New York*
Elena Iorga, *National Bank of Romania*
Esa Jokivuolle, *Bank of Finland*
Signe Krogstrup, *Danmarks Nationalbank*
Michael Kumhof, *Bank of England*
Ana Christina Leal, *Bank of Portugal*

Carlos Lenz, *Swiss National Bank*
Ye Liu, *People's Bank of China*
Ken Matsushita, *Bank of Japan*
Andre Minella, *Central Bank of Brazil*
Emanuel Mönch, *Deutsche Bundesbank*
Gerard O'Reilly, *Central Bank of Ireland*
Eva Ortega, *Bank of Spain*
Michael D. Patra, *Reserve Bank of India*
Anna Paulson, *Federal Reserve Bank of Chicago*
Thórarinn G. Pétursson, *Central Bank of Iceland*
Trevor Reeve, *Federal Reserve Board*
Sigal Ribon, *Bank of Israel*
Kasper Roszbach, *Norges Bank*
Glenn Rudebusch, *Federal Reserve Bank of San Francisco*
Krislert Samphantharak, *Bank of Thailand*
Çağrı Sarıkaya, *Central Bank of Turkey*
Lawrence Schembri, *Bank of Canada*
Woon Shin, *Bank of Korea*
Ulf Söderström, *Sveriges Riksbank*
Ellis Tallman, *Federal Reserve Bank of Cleveland*
George Tavlas, *Bank of Greece*
Geoff Tootell, *Federal Reserve Bank of Boston*
Dobieslaw Tymoczko, *National Bank of Poland*
Hernando Vargas Herrera, *Banco de la República*
Rafael Wouters, *National Bank of Belgium*
Mine Yücel, *Federal Reserve Bank of Dallas*
Ksenia Yudaeva, *Central Bank of the Russian Federation*

Editorial Board

Managing Editor

Luc Laeven
European Central Bank

Co-editors

Klaus Adam
University of Mannheim
Tobias Adrian
International Monetary Fund
Boragan Aruoba
University of Maryland
Huberto Ennis
Federal Reserve Bank of Richmond

Linda S. Goldberg
Federal Reserve Bank of New York
Òscar Jordà
Federal Reserve Bank of San Francisco
Sharon Kozicki
Bank of Canada

Keith Kuester
University of Bonn
Elena Loutschina
UVA Darden School of Business
Steven Ongena
University of Zurich

Associate Editors

Patrick Bolton
Columbia University
Michael D. Bordo
Rutgers University
Mark Carey
Federal Reserve Board
Pierre Collin-Dufresne
École Polytechnique Fédérale de Lausanne
Guy Debelle
Reserve Bank of Australia
Douglas W. Diamond
University of Chicago
Graduate School of Business
Francis Diebold
University of Pennsylvania
Michael Dotsey
Federal Reserve Bank of Philadelphia

Darrell Duffie
Stanford University
Jordi Galí
Centre de Recerca en Economia Internacional (CREI)
Marvin Goodfriend
Carnegie Mellon University
Michael B. Gordy
Federal Reserve Board
Luigi Guiso
Einaudi Institute
Andrew G. Haldane
Bank of England
Takatoshi Ito
Columbia University
David Lando
Copenhagen Business School
Philip Lane
European Central Bank

Francesco Lippi
LUISS University
Carmen M. Reinhart
Harvard Kennedy School
Hélène Rey
London Business School
Jean-Charles Rochet
University of Zurich
Andrew K. Rose
University of California, Berkeley
Lars E.O. Svensson
Stockholm School of Economics
Jürgen von Hagen
University of Bonn
Ernst-Ludwig von Thadden
University of Mannheim
Tsutomu Watanabe
University of Tokyo

Advisory Board

Franklin Allen
Imperial College London
Charles Goodhart
London School of Economics

Hyun Song Shin
Bank for International Settlements
John Taylor
Stanford University

Kazuo Ueda
University of Tokyo
Carl E. Walsh
University of California
Michael Woodford
Columbia University

Asset Purchase Programs and Financial Markets: Lessons from the Euro Area*

Carlo Altavilla,^{a,b} Giacomo Carboni,^a and Roberto Motto^a
^aEuropean Central Bank
^bCEPR

We estimate the effects of the asset purchase program launched by the European Central Bank (ECB) in 2015 on euro-area bond yields and assess its transmission channels. Our identification strategy rests on exploiting market reactions to news about the size and maturity range of asset purchases and cross-sectional variations in security-level data on prices and purchased quantities. We find that ECB asset purchases amounting to 10 percent of euro-area GDP compress euro-area 10-year sovereign bond yields by around 65 basis points (“stock effects”), which is a sizable impact, also in light of the low financial distress prevailing at the time. Bonds more exposed to interest rate risk (duration risk channel) and with lower creditworthiness (credit risk channel) experienced the highest returns. Local supply channels, narrowly related to the intensity of purchases in targeted market segments, are estimated to play a more limited role. Our findings provide support to theories that posit how low financial distress, while weakening local supply channels, facilitates the transmission of quantitative easing beyond targeted segments. The implication is that asset purchases are a viable policy tool under both high and low financial distress although the transmission channels are different.

JEL Codes: E43, E44, E52, E65, G14.

*First version: November 2015. For helpful comments, we thank Bartosz Maćkowiak, Oreste Tristani, Michael Weber, and seminar participants from the Bank of England, Bank of Italy, Deutsche Bundesbank, European Central Bank, and International Monetary Fund. All errors and omissions are our own responsibility. The opinions in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank or the Eurosystem. Please address any comments to Carlo Altavilla (carlo.altavilla@ecb.europa.eu), Giacomo Carboni (giacomo.carboni@ecb.europa.eu), or Roberto Motto (roberto.motto@ecb.europa.eu).

1. Introduction

Within the academic and policymaking environment, there has recently been renewed debate about the efficacy of large-scale asset purchase programs, fueled in part by the launch of strategic reviews by major central banks around the world. In this debate, center stage has been taken by the question of whether asset purchases should become a more regular tool of policy stabilization. In essence, asset purchases would be called upon to support the conventional interest rate instrument more frequently than in the past, as the enduring low levels of the natural rate of interest imply a higher incidence of the lower bound for the nominal interest rate. So far, experience of asset purchases largely relates to their adoption in response to the global financial crisis of 2008–09 and its ramifications. With some important qualifications, during that time asset purchases were generally found to be effective in steering financing conditions and sustaining the economy going forward. Hence a natural question emerges regarding the effectiveness and transmission channels of large-scale asset purchases in situations other than financial crises.

The theory posits some form of financial frictions for asset purchases having direct effects on asset prices by inducing a change in the quantity and composition of financial assets held by the market, generally known as “portfolio balance effects.” In the absence of any frictions, asset purchases would affect yields only indirectly to the extent that they provide a signal of the path of future risk-free short-term rates (“signaling effect”). Because financial frictions are likely to be more binding at times of financial distress, the presumption is that portfolio balance effects would be more relevant during those periods. Empirically, there is open debate regarding the effects and transmission channels of large-scale asset purchases, as documented in the vast literature on programs carried out in major advanced economies since the outbreak of the global financial crisis. Specifically, two different strands of literature can be identified. The first strand finds sizable portfolio balance effects for those programs carried out in the aftermath of the collapse of Lehman Brothers.¹ These portfolio balance effects are estimated to operate

¹The early work on the Federal Reserve System’s first large-scale asset purchase program (LSAP1) by Gagnon et al. (2011) emphasizes primarily the

mainly via “narrow channels,” i.e., channels that are specific to the assets targeted by the program (“local supply channel”) with limited spillovers to nontargeted market segments (see, for instance, Krishnamurthy and Vissing-Jorgensen 2011; D’Amico and King, 2013; and McLaren, Banerjee, and Latto 2014). The second strand comprises studies that also find significant effects of asset purchases in less extreme financial conditions (see, for instance, Cahill et al. 2013; and Li and Wei 2013) and/or considerable pass-through effects on nontargeted assets, in the form of borrowing costs faced by businesses and households (Gilchrist, López-Salido, and Zakrajšek 2015) or the exchange rate (Rogers, Scotti, and Wright 2018). In essence, this strand of literature supports the view that asset purchases work via “broad channels.”

We aim to shed light on this issue by estimating the effects of the European Central Bank’s (ECB’s) Asset Purchase Programme (APP) on euro-area bond yields and assessing its transmission channels. The APP, which was announced in January 2015, may provide helpful insights, because the purchases took place under relatively low financial distress and good market functioning, particularly when compared with programs carried in the immediate aftermath of the global financial crisis. Also, a distinct novelty of the APP when compared with large-scale asset purchases carried out in the United States (LSAPs) and the United Kingdom is that it targets long-term sovereign securities spanning different degrees of creditworthiness.

Our identification strategy draws first on market reactions to policy announcements, notably in the form of high-frequency asset price responses to news about the size and maturity range of asset purchases. Here we exploit a distinct feature of the ECB’s communications in that information about size and maturity was released at different points during the January 2015 press conference. We also run a regression analysis that exploits the cross-sectional variations in

portfolio balance effects of the program and the associated compression in term premiums. For the United Kingdom, Joyce et al. (2011) and Breedon, Chadha, and Waters (2012) similarly find that the initial quantitative easing in 2009–10 significantly lowered government bond yields through portfolio balance effects. Christensen and Rudebusch (2012) find that changes in policy expectations appear to have played an important role in LSAP1 in the United States, while the declines in yields in the United Kingdom appeared to reflect reduced term premiums.

security-level data on prices and purchased quantities, using proxies for different quantitative easing (QE) channels, including duration risk, credit risk, and local supply channels. As conceptual guidance to identify these channels empirically and to interpret the results, we extend an illustrative model with bond supply effects à la Vayanos and Vila (2009) by considering bonds with different credit risk intensity, thereby reflecting the cross-country heterogeneity within the euro area.

We find economically significant effects of the APP working primarily via “broad channels”: bonds with longer duration and thus more exposed to interest rate risk (duration risk channel) and bonds with higher credit risk (credit risk channel) experienced the greatest returns. By contrast, local supply channels, narrowly related to the intensity of the ECB’s interventions in targeted market segments, are estimated to play a more limited role, taking into account both the event study and the security-level regressions. Quantitatively, our baseline estimates imply that ECB asset purchases amounting to 10 percent of euro-area gross domestic product (GDP) in 2015 (i.e., around €1.0 trillion) compress GDP-weighted euro-area 10-year sovereign bond yields by around 65 basis points. The duration risk channel accounts for the bulk of the impact on euro-area 10-year yields, while the credit risk channel accounts for around 15 percent of the impact and the local supply channel for around 7 percent. We estimate an additional 15 basis point decline in 10-year yields for less creditworthy sovereigns via the credit risk channel. These transmission channels are all part of the broadly defined portfolio balance effects, and hence they are distinct from the impact that the APP might have had by steering expected future short-term rates (signaling effect).

Our results on the relative importance of the different transmission channels, as well as the empirical literature described above, are consistent with the predictions of our illustrative model. Under heightened risk aversion, asset purchases push up bond prices by exerting demand pressures in targeted segments. However, because of market segmentation, asset purchases operate locally (local supply channels), with limited spillovers to nontargeted segments.²

²This interpretation is also supported by findings of the empirical literature on two earlier ECB programs announced in 2010 and 2012 at a time of market

Conversely, our illustrative model predicts that, under less extreme financial stress, investors are effective in diversifying the total amount of risk borne in their portfolios across market segments. The compression in premiums reflects the overall quantity of risk absorbed by the central bank and the exposure of the securities to risk factors. As a result, bond returns are higher for securities with longer duration and higher credit risk, in line with our findings. Similarly, Li and Wei (2013) find significant effects of asset purchases on U.S. longer-term rates via duration risk channels, on the basis of a term structure model estimated over a pre-crisis period of good market functioning. Likewise, on the basis of the same methodology developed by D'Amico and King (2013), Meaning and Zhu (2011) estimate the largest effects of LSAP2 on longer-dated securities through duration risk channels, in contrast with the relevance of local supply effects documented by D'Amico and King (2013) for LSAP1.

Our findings pertain to persistent changes in bond prices known in the literature as “stock effects.” These effects are distinct from the effects related to the ongoing implementation of asset purchases (“flow effects”), which could reflect improvements in liquidity conditions and market functioning, and are typically associated with periods of high financial stress. Available evidence in the literature suggests that APP flow effects are fairly contained and short-lived (see, for instance, De Santis and Holm-Hadulla 2020). Quantitatively, our estimated stock effects are broadly in line with other studies of the APP, being, for instance, slightly higher than those of Eser et al. (2019) and somewhat lower than those of De Santis (2020). Also, our estimates tend to be within the (admittedly wide)

distress, although they differ in breadth and scope from large-scale asset purchases. For instance, when assessing ECB bond purchases in stressed sovereign markets under the Securities Markets Programme (SMP) carried out in 2010–11, Eser and Schwaab (2016) find evidence of local supply effects in segmented markets. Similarly, when assessing a subsequent program of outright purchases of sovereign securities (Outright Monetary Transactions, OMTs), also announced under market distress but never activated, Altavilla, Giannone, and Lenza (2016) find yield effects concentrated in maturity brackets targeted by the program, with limited spillovers to nontargeted segments. When assessing the same two programs, Krishnamurthy, Nagel, and Vissing-Jorgensen (2018) also document relevant sovereign bond segmentation effects.

range of estimated effects of LSAPs. Various studies find that Federal Reserve asset purchases amounting to 10 percent of U.S. GDP are found to reduce the 10-year U.S. Treasury yield by between 37 and 165 basis points.³

Finally, our findings have important implications for the future design of asset purchase programs at times of high and low financial distress. At times of high financial distress, local supply channels might prevail, limiting the pass-through of the policy stimulus to nontargeted assets. Therefore, central banks are advised to broaden the spectrum of assets purchased to deliver more favorable financial conditions across market segments. At times of low financial distress, central banks relying on asset purchases to stabilize the economy and circumvent the lower bound on the policy rate can count on investors to facilitate the transmission of the QE stimulus beyond targeted segments.

In terms of methodology, a number of other papers have focused on financial market reactions to policy announcements, including Gagnon et al. (2011) and Krishnamurthy and Vissing-Jorgensen (2011) for the United States and Joyce et al. (2011) for the United Kingdom. In particular, our identification strategy shares similarities with D'Amico et al. (2012), Joyce and Tong (2012), and Cahill et al. (2013) in exploiting high-frequency responses of asset prices to news about the size and maturities of asset purchases. Regarding the regression analysis using security-level data, we extend the methodology developed by D'Amico and King (2013) and include empirical proxies for duration risk and credit risk channels, while similarly contemplating local supply channels. This extension turns out to be particularly relevant for assessing the effects of the APP considering the dominant contribution of duration risk and credit risk channels when compared with local supply channels. From this perspective, our paper also differs from Arrata and Nguyen (2017), who recently

³These estimates refer to various studies of LSAP1 and LSAP2, including Gagnon et al. (2011), Krishnamurthy and Vissing-Jorgensen (2011), Meaning and Zhu (2011), Swanson (2011), D'Amico et al. (2012), Cahill et al. (2013), D'Amico and King (2013), and Li and Wei (2013). LSAP1 is generally found to deliver the largest impact. At the same time, Cahill et al. (2013) show that the low response under LSAPs subsequent to LSAP1 often found in the literature mainly comes from event studies that do not control for market expectations which largely anticipated the announcement of the later LSAPs.

used security-level data to assess the impact of the ECB's purchase program on the French bond market, focusing on local supply channels. Also, our regression analysis based on security-level data on prices and purchased quantities differs from the analysis in Koijen et al. (2017), which instead focuses on changes in the holdings of securities by euro-area investors.

Our paper also differs from De Santis (2020), who assesses the stock effects of the APP using an index of Bloomberg news on the APP to account for possible anticipation effects. Eser et al. (2019) trace the impact of the APP by estimating a term structure model in which asset purchases are assumed to operate via the duration risk channel. Andrade et al. (2016) note that the asset price movements in response to the January 2015 announcement are consistent with versions of the portfolio rebalancing channel acting through the removal of duration risk and the relaxation of leverage constraints for financial intermediaries. Unlike those studies, our paper both tests for various QE transmission channels and estimates their relative importance. We show that this has relevance for understanding the propagation of asset purchases through broader financing conditions and the macroeconomy.

The remainder of the paper is organized as follows. Section 2 offers conceptual guidance on the APP transmission channels. Section 3 focuses on the impact of the APP by examining market reaction to policy announcements. Section 4 presents a regression analysis using data on purchased quantities. Section 5 elaborates on the interpretation of our findings, and section 6 provides conclusions.

2. QE Transmission Channels: An Illustrative Model

An illustrative model helps to provide the motivation for our empirical analysis and forms an intuitive basis for our results. The purpose of this section is to sketch out our extension of a term structure model with bond supply effects à la Vayanos and Vila (2009) to allow for bond credit risk premiums (for details, see the appendix).⁴ This extension is relevant for interpreting asset purchases in the

⁴Variations and extensions of the framework developed by Vayanos and Vila (2009) have been formalized by, among others, Hamilton and Wu (2012), Greenwood and Vayanos (2014), and King (2015, 2019).

euro area, in light of the different creditworthiness across member countries.

There are two types of agents: arbitrageurs and preferred-habitat investors. Arbitrageurs trade bonds across market segments and maximize a mean-variance objective function defined over their portfolio's return $R_{(t,t+1)}$

$$\begin{aligned} & \max_{\omega_t^{(n)}} \left[E_t R_{(t,t+1)} - \frac{1}{2} \sigma \text{Var}_t R_{(t,t+1)} \right] \\ R_{(t,t+1)} & \equiv \sum_{n=1} \omega_t^{(n)} [\exp(\bar{p}_{t+1}^{(n-1)} - \bar{p}_t^{(n)}) - 1], \end{aligned} \quad (1)$$

where σ is the risk-aversion coefficient, also proxying for limited risk-bearing capacity; $\omega_t^{(n)}$ is the share of the aggregate portfolio held in the n -period maturity zero-coupon bond. Bond returns result from purchasing an n -period bond at time t at price $\bar{p}_t^{(n)}$ and selling it at $t + 1$ with maturity of $n - 1$ at price $\bar{p}_{t+1}^{(n-1)}$. In our model, bonds are also subject to credit risk, whose intensity ψ_t is formalized as an affine function of risk factors $\psi_{t+1} = \gamma' X_{t+1}$. Preferred-habitat investors have instead clientele's demand defined as

$$\xi_t^{(n)} = \varphi(\bar{y}_t^{(n)} - \beta_t^{(n)}), \quad (2)$$

where $y_t^{(n)}$ is the yield on the n -period bond, given by $-p_t^{(n)}/n$, and $\beta_t^{(n)}$ captures demand factors. Equilibrium conditions in the bond market require that the demand from arbitrageurs, $\omega_t^{(n)}$, combined with the demand from preferred-habitat investors, $\xi_t^{(n)}$, equates to the supply of bonds $S_t^{(n)}$.

In this framework, it is possible to identify two (polar) types of equilibrium bond prices, which depend on arbitrageurs' risk aversion and are characterized by distinct QE transmission channels, all of which are part of the broader umbrella of portfolio balance channels. In the first case, which is characterized by heightened risk aversion, arbitrageurs are constrained in their ability to integrate market segments by their limited risk-bearing capacity. Equilibrium yields are then pinned down by equation (2) jointly with the bond supply. The first point to take away from the model is that, under heightened risk aversion, asset purchases push up bond prices by reducing the supply

of bonds available to preferred-habitat investors; at the same time, because of market segmentation, these effects are local to the those market segments targeted by the asset purchases. This describes the mechanism known in the literature as the “local supply channel.”⁵

In the second case, where risk aversion is less extreme and yet risk-bearing capacity is limited, arbitrageurs eliminate arbitrage opportunities by pricing risk consistently across market segments. As a result, the compensation per unit of risk, i.e., the market price of risk λ_t , is common to all bonds and given by

$$\lambda_t \equiv \sigma \sum_{n=1} (\omega_t^{(n)} \tilde{b}_{n-1}), \tag{3}$$

where \tilde{b}_{n-1} is the sensitivity of portfolio’s holdings to risk. In the absence of credit risk ($\gamma = 0$) and assuming that the short-term rate is the only risk factor, \tilde{b}_{n-1} takes the form of a nondecreasing concave function in the security’s maturity n ,

$$\tilde{b}_{n-1} \equiv \frac{1 - \exp(-\kappa n)}{\kappa},$$

where κ is endogenously determined. Holding a portfolio with longer-maturity bonds implies a higher market price of risk, and this affects all bonds. At the same time, individual securities with longer maturity entail greater exposure to interest rate risk and so command higher risk premiums, as reflected in the expected holding returns of bonds in excess of the short rate

$$E_t[R_{(t,t+1)}^{(n)} - \bar{r}_t] = \lambda_t \sigma_r^2 \frac{1 - \exp(-\kappa n)}{\kappa}, \tag{4}$$

where σ_r^2 is the conditional volatility of the short-term rate. When we turn on the credit risk intensity, equation (4) preserves the same functional form. The difference is that both λ_t and κ become in equilibrium a function of γ , meaning that the sensitivity of portfolio’s holdings to risk tends to increase, which is also the case for risk premiums. The second point to take away from the model is that in normal financial market conditions asset purchases reduce the overall risk that arbitrageurs must hold in equilibrium, hence affecting

⁵See also Cochrane (2008) and Vayanos and Vila (2009).

the entire term structure through the reduction of the market price of risk. At the same time, the decline in premiums is stronger for those securities more exposed to risk factors. These include securities with longer duration and thus more exposed to interest rate risk, a mechanism known as the “duration risk channel of QE.” The third point to take away from the model is that bond returns are higher for securities more exposed to credit risk, a mechanism we call the “credit risk channel of QE.” Both the duration risk and credit risk channels fall within the “broad channels” category for asset purchases. In the regression analysis, we draw explicitly on equations (3) and (4) to construct empirical proxies for duration risk and credit risk channels using security-level data.

The final point to take away from the model is that both the local supply channels (“narrow channels”) and the duration and credit risk channels (“broad channels”) pertain to the stock effects of asset purchases, i.e., the persistent changes in bond prices due to variations in the (risk-adjusted) stock of bonds that private investors must hold. This model and our empirical analysis do not assess flow effects generated by ongoing implementation of asset purchases, which could be related to the enhancement of liquidity conditions and unlocking of market functioning and is typically associated with periods of high financial distress.⁶ In the following, we estimate the stock effects of the APP and identify the relative strength of local supply channels and duration risk and credit risk channels.

3. Empirical Analysis: Effects and Transmission Channels Exploiting Policy Announcements

The APP was officially announced on January 22, 2015 in the form of purchases of investment-grade securities amounting to €60 billion per month intended to run until September 2016 and “*in any case*” until the Governing Council of the ECB saw inflation stabilizing at

⁶When investigating the flow effects of the APP, De Santis and Holm-Hadulla (2020) find that they are limited, short-lived, and concentrated in securities issued in higher-yield jurisdictions, with longer maturity and lower liquidity.

values consistent with its inflation aim.⁷ The targeted assets comprise public-sector securities issued primarily by member countries with residual maturities of up to 30 years and spanning various credit ratings with investment-grade status.⁸ The exact starting date for purchases of public-sector securities was communicated on March 5, 2015. A relevant aspect for our inference is that the APP was announced during a period of good market functioning and contained market distress as reflected in a number of indicators. For instance, the EURIBOR-OIS spread, which is often used as a measure of stress in the money market, was stable at around 20 basis points, markedly down from the peaks recorded after the collapse of Lehman Brothers (around 150 basis points) or at the height of the sovereign debt crisis in the autumn of 2011 (around 130 basis points).⁹ Likewise, measures of euro-area stock and bond market volatility were broadly around levels prevailing prior to the financial crisis, as were measures of market liquidity, such as the spreads between the yields of German government bonds and German agency bonds.¹⁰ Sovereign spreads of lower-rated euro-area countries had also receded substantially. For instance, at the start of 2015, Italian and Spanish government bond spreads (*vis-à-vis* German government bonds) stood at around 40 basis points at the one-year maturity, after having surged to above 500 basis points in the autumn of 2011.¹¹ Longer-term yield spreads had also narrowed considerably, standing slightly above 100 basis points at 10-year maturity at the start of 2015, after peaking at around 500 basis points during the sovereign debt crisis in 2011–12.

⁷Hence, upon the January 2015 announcement, intended purchases of private- and public-sector securities under the APP amounted to €1.14 trillion, roughly corresponding to 11 percent of euro-area annualized 2014:Q4 nominal GDP.

⁸Investment-grade status is one of the eligibility criteria underpinning the purchase of securities under the APP.

⁹The EURIBOR is the (average) rate at which euro-area banks offer to lend unsecured term funds to one another. The OIS is the euro overnight index swap rate.

¹⁰German government bonds (*bunds*) are highly liquid securities backed by the high-rated federal government. Bonds issued by the federal government-owned development bank, Kreditanstalt für Wiederaufbau (KfW), carry the same credit risk as *bunds* but are less liquid. At the start of 2015, the KfW-*bund* spread was below levels prevailing around the start of 2008.

¹¹This compression of sovereign spreads closely matches the narrowing of the respective credit default swap (CDS) spreads.

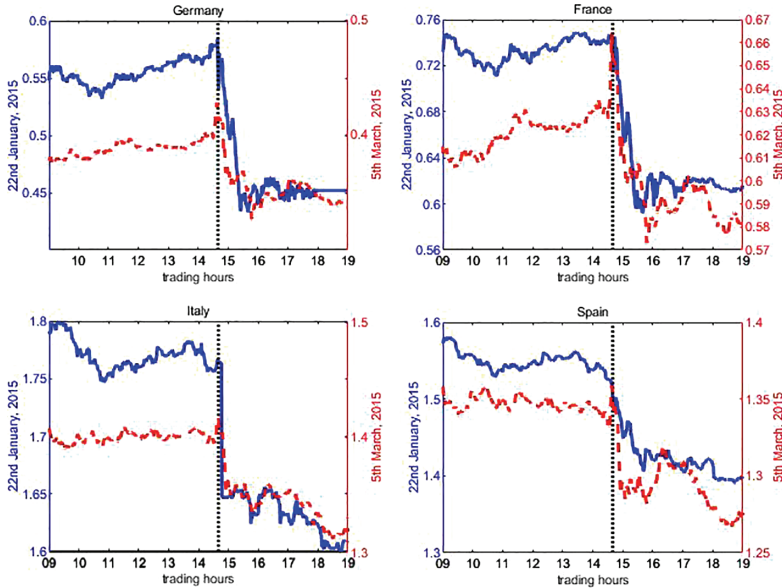
As described above, the contained financial distress prevailing at the time of the APP might have had a bearing on the QE transmission channels. Moreover, from a methodological perspective, being conducive of a prompt response of asset prices to news, low market distress makes event-study analyses a suitable approach for drawing inferences. As a result, we first focus on the market reaction to policy announcements, following a vast strand of literature on the financial market effects of large-scale asset purchases (for the United States, see, for instance, Gagnon et al. 2011 and Krishnamurthy and Vissing-Jorgensen 2011; and for the United Kingdom, see, for instance, Joyce et al. 2011).

3.1 Identification via Size and Maturity News

Being forward-looking, financial markets would be expected to respond to asset purchase programs upon announcement and prior to actual purchases taking place. Figure 1 displays the high-frequency intraday movements of sovereign yields for the four largest euro-area economies on the dates of the two official APP announcements, January 22 and March 5 (solid blue and dashed red lines, respectively¹²). The policy decisions regarding the APP were communicated during the press conferences that started at 14:30 after the respective Governing Council meetings, and after the release of monetary policy interest rate decisions at 13:45. The two APP announcements (denoted by the vertical dashed lines in figure 1) mark a significant step decline in 10-year sovereign bond yields on both event dates and across euro-area countries. This effect is more pronounced for the less creditworthy Italian and Spanish bonds, whose yields plummeted immediately after the policy announcements and continued to recede further in the course of the day. This market reaction was not due to (the anticipation of) stronger interventions by the ECB vis-à-vis riskier sovereigns. Indeed, as communicated at the start of the press conference on January 22, the ECB purchases of securities issued by euro-area governments were based on the shares of the respective central banks in the ECB's capital key, i.e., largely reflecting the size of the economies of member countries.

¹²For figures in color, see the online version of the paper at <http://www.ijcb.org>.

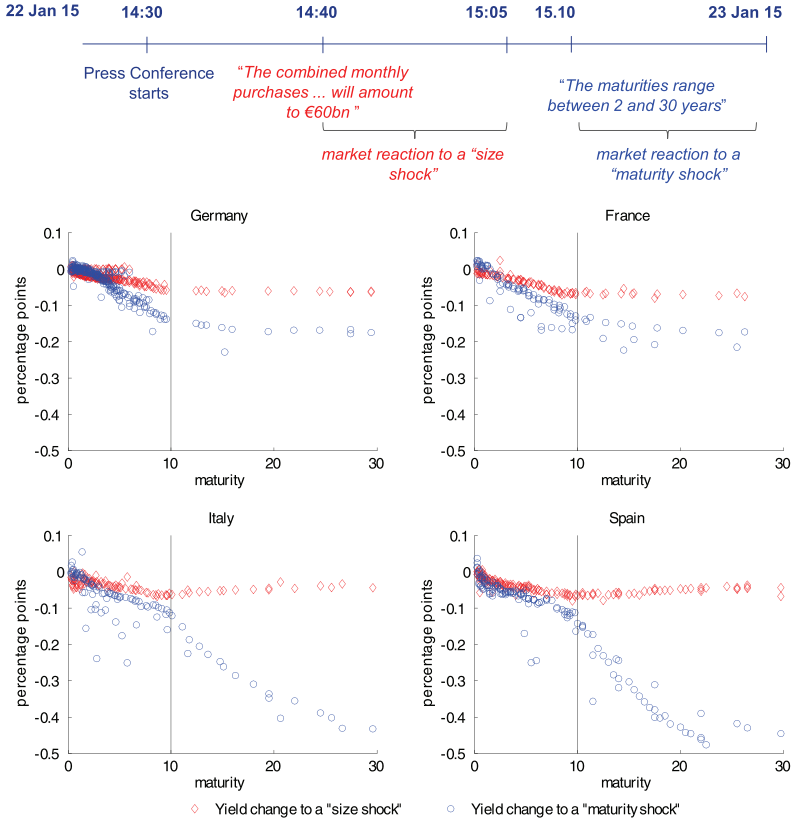
Figure 1. Intraday Movements in 10-Year Yields of Selected Euro-Area Sovereigns on the Two Initial APP Announcement Dates



Notes: The solid (blue) line represents movements on January 22, 2015 (LHS axis), and the dashed (red) line represents movements on March 5, 2015 (RHS axis). The start of the ECB press conference is identified with the vertical dotted lines.

To enhance the identification of the APP transmission channels, we exploit distinct announcements about the size and maturity distribution of purchases during the January 22 press conference. Specifically, the size of the program was communicated at 14:40, at the beginning of the press conference, when the ECB president announced that “the combined monthly purchases of public and private sector securities will amount to €60 billion. They are intended to be carried out until end-September 2016.” The range of maturities for the bond purchases was communicated at 15:10, during the question and answer session, when the president stated that “the maturities range between 2 and 30 years.” Figure 2 (top panel) displays the timeline of the announcements. The news content of these announcements depends on the extent to which they were

Figure 2. High-Frequency Reaction of Bond Yields of Selected Euro-Area Sovereigns around the Announcements of (i) the Size of the APP and (ii) the Maturities' Range of the Purchases during the January 22 Press Conference



Notes: Each diamond/circle represents the change in an individual bond yield at the ISIN level. The vertical solid line denotes the 10-year maturity.

anticipated by the market. As a proxy for market expectations, we rely on surveys of market participants. Specifically, survey-based information suggests that financial markets had anticipated the launch of the APP, with the median size of the program being around €550 billion and the median expected maturity range being up to

10 years (see table B.3 in the appendix).¹³ Figure 2 (bottom panel) depicts the high-frequency response of ISIN-level yields for the four largest euro-area countries to the announcements about (i) the size of the program (“size shock,” yield change shown by red diamonds) and (ii) and the maturity range (“maturity shock,” yield change shown by blue circles); the vertical line denotes the 10-year maturity, above which market participants were not expecting ECB purchases.

The conceptual framework described in the previous section entails distinct predictions regarding the way various QE transmission channels would operate in response to these shocks. First, local supply channels imply no movement in bond yields in maturity brackets above 10 years in response to the size shock, and prior to the maturity shock. This is because, as noted above, market participants did not anticipate asset purchases in those maturity brackets. Second, local supply channels predict an increase in yields at below 10-year maturity in response to the maturity shock. The reason is that the maturity shock is tantamount to an unexpected reallocation of the previously announced APP envelope from maturity brackets below 10 years to higher maturities. The resulting lower-than-expected purchases in brackets below the 10-year maturity would then lead to an upward adjustment of yields at those maturities. As shown in figure 2, the evidence is that German and French yields declined across the whole term structure following both the size and maturity shocks, and these effects rise with the term to maturity. The implication is that these yield responses are at odds with the local supply channel and are instead suggestive of, and consistent with, the duration risk channel. To investigate the emergence of the credit risk channel, we draw on the model’s prediction of higher bond returns for less creditworthy securities. As an empirical proxy of euro-area securities with higher credit risk, we depict in figure 2 the response of sovereign bond yields for Italy and Spain. According to the credit risk channel, following both the size and the maturity shocks, the compression in bond yields has been materially stronger

¹³The information about the expected size of the APP is extracted from surveys carried out by Bloomberg prior to the January 2015 announcement; the information about the expected maturity range is extracted from analyst reports published by, for instance, major investment banks such as J.P. Morgan, Goldman Sachs, Bank of America Merrill Lynch, and Nomura.

for Italy and Spain than for the more creditworthy Germany and France. As noted above, this is not related to stronger ECB interventions in less creditworthy jurisdictions. In fact, relative to the size of the outstanding debts, ECB purchases underpinning the APP were actually lower in more indebted countries.

Similar support for the duration risk and credit risk channels emerges when extending the analysis to other euro-area countries. We document these findings systematically using a regression exercise that exploits the high-frequency yield movements of euro-area sovereign bonds recorded between 14:00 and 16:00, hence covering the time windows around the size and maturity announcements. Specifically, we start with regressions of the form

$$\begin{aligned} \Delta y_{i,t} = & \alpha_i + \beta_1 T^{size\ shock} \times ISIN^{\tau > 10} \\ & + \beta_2 T^{maturity\ shock} \times ISIN^{\tau < 10} + \beta_3 T^{size\ shock} \times ISIN^{\tau < 10} \\ & + \beta_4 T^{maturity\ shock} \times ISIN^{\tau > 10} + \beta_5 T^{size\ shock} \times ISIN^{credit\ risk} \\ & + \beta_6 T^{maturity\ shock} \times ISIN^{credit\ risk} + \epsilon_{i,t}, \end{aligned} \quad (5)$$

where $\Delta y_{i,t}$ is the yield change of ISIN i at a five-minute interval; $T^{size\ shock}$ is a dummy that takes the value 1 in the time window between the announcement of the size shock at 14:40 and the announcement of the maturity shock at 15:10, and 0 otherwise; $T^{maturity\ shock}$ is a dummy that takes the value 1 after the announcement of the maturity shock, and 0 otherwise; $ISIN^{\tau < 10}$ identifies the ISINs with remaining maturity τ below 10 years; and $ISIN^{credit\ risk}$ is a dummy that takes the value 1 for ISINs associated with lower-rated euro-area sovereigns, comprising Spain, Italy, Portugal, Cyprus, and Greece, and 0 otherwise.¹⁴

By the same reasoning described above, the presence of local supply channels predicts that β_1 and β_2 should be non-negative. Their parameter estimates are reported in the first two rows of table 1, where the two columns refer, respectively, to specifications without

¹⁴The high-creditworthiness partition comprises sovereign securities issued by those member countries that, upon the launch of the APP, had been assigned a credit rating of A– or above by at least one of the three main rating agencies (Moody’s, Standard & Poor’s, and Fitch Ratings). Securities with a credit rating of A– or above are associated with issuers with a strong to extremely strong capacity to meet their financial commitments.

Table 1. Estimated High-Frequency Yield Responses to the APP Announcements on the Size and Maturities' Range of the Purchases

	Without Controlling for Credit Risk	Controlling for Credit Risk
$T^{size\ shock} \times ISIN^{\tau > 10}$	-1.355*** (0.05)	-1.355*** (0.05)
$T^{maturity\ shock} \times ISIN^{\tau < 10}$	-0.457*** (0.02)	-0.410*** (0.03)
$T^{size\ shock} \times ISIN^{\tau < 10}$	-0.690*** (0.02)	-0.690*** (0.02)
$T^{maturity\ shock} \times ISIN^{\tau > 10}$	-1.762*** (0.05)	-1.676*** (0.08)
$T^{size\ shock} \times ISIN^{credit\ risk}$		-0.05 (0.06)
$T^{maturity\ shock} \times ISIN^{credit\ risk}$		-0.127*** (0.05)
Fixed Effects	Yes	Yes
Adj. R ²	24%	25%
No. ISIN	603	603
No. Observations	11,960	11,960

Notes: t-statistics are in parentheses. ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

and with control for credit risk. We find that the estimates of β_1 and β_2 are both negative and of statistical significance. In response to the size shock, yields in maturity brackets above 10 years decline twice as much as those below that maturity. In response to the maturity shock, yields tend to decline across maturity brackets, and more prominently at longer maturities. The statistical and economic significance of these results rejects the prevalence of local supply channels and instead points to the importance of duration risk channels. In the second column, we add the interaction terms associated with the credit risk dummy $ISIN^{credit\ risk}$. The negative estimate of β_6 on the interaction term between the maturity dummy and the credit risk dummy means that securities more exposed to credit risk factors experience stronger yield declines, which is suggestive of credit risk channels at play.

There is an interesting contrast between our findings and the findings of, for instance, D'Amico et al. (2012), who similarly exploit

unanticipated announcements by the central bank to identify the transmission channels of asset purchases. Specifically, their focus rests on announcements by the Federal Open Market Committee (FOMC) about the maturity range of the reinvestment program in August 2010. They find that yields of securities included in the purchase range declined more than those on other securities, and conclude that local supply channels were prominent. Cahill et al. (2013) extend the analyses to additional FOMC announcements of Treasury purchase programs. They find strong supporting evidence for local supply channels under the first and second large-scale asset purchase programs (LSAP1 and LSAP2), which were communicated in 2009 and 2010, respectively. The results for the subsequent program, known as the Maturity Extension Program and announced in 2011, and its extension, communicated in 2012, are suggestive of a combination of local supply and duration risk channels.

The conceptual framework described in the previous section provides a possible way to rationalize these differences in the relevance of the transmission channels of asset purchases. The model's prediction is that, under heightened risk aversion of the type prevailing at the time of the programs launched in the aftermath of the global financial crisis, asset purchases push up bond prices and compress risk premiums by exerting demand pressures in targeted market segments. However, precisely because of market segmentation, asset purchases operate via local supply channels. Under less extreme financial distress, such as that characterizing the period of the APP, the compression in premiums is related to the overall quantity of risk absorbed by the central bank and the exposure of the securities to risk factors. Bond returns would be higher for securities with longer duration and higher credit risk.

3.2 Event-Study Evidence Accounting for Anticipation Effects

Quantifying the effects of policy decisions on the basis of changes around official announcements leads to unbiased estimates of their effects to the extent that these decisions were unanticipated by the market. This is a relevant concern for the APP because, as noted above, according to survey-based information, market participants had indeed anticipated the ECB asset purchases prior to the January

2015 announcement. Evidence of this is also found in the increasing number of articles published in international newspapers since September 2014 explicitly anticipating an ECB QE-type program.¹⁵

To account systematically for possible anticipation effects, we carry out an event-study exercise that considers a broader set of events than the two official announcements above. As reported in table B.1 in the appendix, this set of events includes selected speeches by the ECB president hinting at a forthcoming asset purchase program and press conferences following ECB Governing Council meetings, including the two official announcements referred to above.¹⁶ Because these events span a relatively wide time window, we explicitly control for the possible impact of concomitant macroeconomic news on asset prices. Formally, our estimates are obtained by regressing the daily changes in yields on the event dummies as well as on the surprise component of a wide set of macroeconomic releases¹⁷:

$$\Delta y_t = \sum_{j=1}^k \alpha_j D_{j,t} + \sum_{j=1}^k \beta_j D_{j,t-1} + \sum_{s=1}^m \gamma_s News_{s,t} + \varepsilon_t, \quad (6)$$

where Δy_t is the daily change in yield of a given asset, $D_{j,t}$ is dummy variable that takes the value 1 at the time of the policy event j and zero otherwise, k is the total number of events, and $News$ is

¹⁵For instance, on September 19, 2014, the *Financial Times* published the article entitled “Weak ECB Loans Demand Paves the Way for ‘QE’,” where “weak loans demand” refers to the lower than expected volumes in the second targeted longer-term refinancing operation (TLTRO II). On November 27, 2014, the *Financial Times* published the article entitled “US Data Disappoint as Possibility of European QE Comes into Focus,” and one day later it qualified the message with the article “Draghi Needs Support on QE in the Eurozone.” About one month later, on December 30, 2014, the *Economist* published the article “Euro-zone Quantitative Easing: Coming Soon?”

¹⁶As a robustness check, we compare this “narrative” approach to dating events with a more “agnostic” approach based on an index of intensity of news coverage on possible purchase programs in the euro area (see figure B.1 in appendix B). This index of news coverage is derived by using an extensive range of different news sources from the Dow Jones news database, Factiva. Overall, it is striking how the news index spikes around the identified event dates, and that is particularly the case for the six Governing Council meetings, which represent “local maxima” of the news index.

¹⁷The regression analysis follows the approach in Altavilla and Giannone (2017).

the surprise component of macro releases.¹⁸ Statistical significance is assessed by using heteroskedasticity-robust standard errors. The event dummy coefficients are jointly tested with an F -test under the zero-null hypothesis.

Based on regression equation (6), table 2 reports the estimated effects of the APP on euro-area sovereign yields, as well as on bond yields of the four largest member countries, at maturities of 5, 10, and 20 years, in terms of both one- and two-day window changes.¹⁹ Our preferred specification is based on the two-day changes, where we control for macroeconomic news (“controlled event study”). This specification strikes a balance between two dimensions. First, prior to the first official APP announcement, financial markets might have been slower in understanding, and responding to, the evolving policy communication from the ECB. This fact militates in favor of considering a wider time window for the analysis. At the same time, this choice makes it more likely that the policy signal can be contaminated by concomitant news; hence the need to control for macroeconomic surprises. In any case, to place these estimates in perspective, we also report the results based on a more standard approach that does not control for macroeconomic news (the “standard event study”).²⁰

Overall, the estimates suggest that the launch of the APP in January 2015 significantly lowered sovereign bond yields. The estimated effect on euro-area yields is around 60 basis points at the 10-year maturity. These estimates are slightly more conservative than those based on the standard event study. The cross-asset price reactions support the view that the APP worked primarily through broad

¹⁸Specifically, we consider macroeconomic news for the euro area, its largest economies, and the United States, collected from Bloomberg over the sample period ranging from the beginning of January 2014 to the end of March 2015 (see table B.2 in appendix B).

¹⁹The sovereign yield curve for the euro area is the yield curve estimated by the ECB using all euro-area central government bonds and is released on a daily basis. The daily releases are available at <http://www.ecb.europa.eu/stats/money/yc/html/index.en.html>.

²⁰A one-day window change is measured as the change in yield from the closing level on the day prior to the event to the closing level on the day of the event; a two-day window change is measured as the change in yield from the closing level on the day prior to the event to the closing level on the day after the event.

Table 2. Changes in Sovereign Bond Yields of Selected Euro-Area Economies around the APP Event Dates (basis points)

	5-Year Maturity					10-Year Maturity					20-Year Maturity				
	Euro Area	Germany	France	Italy	Spain	Euro Area	Germany	France	Italy	Spain	Euro Area	Germany	France	Italy	Spain
<i>Controlled Event Study</i>															
One-Day Change	-30*	0	-17	-62***	-67***	-30*	-22*	-36*	-77***	-84***	-22	-20	-33*	-72***	-79***
Two-Day Change	-42**	-7	-19*	-45**	-50***	-61*	-36**	-45**	-71***	-76***	-64**	-56**	-62**	-88***	-90***
<i>Standard Event Study</i>															
One-Day Change	-29*	-2	-24	-66***	-66***	-27*	-25*	-39*	-84***	-83***	-20*	-25*	-38**	-74***	-72***
Two-Day Change	-46**	-9	-31*	-49**	-54***	-63**	-39**	-53**	-83***	-78***	-66***	-60**	-73**	-94***	-89***

Notes: Based on the 18 event dates reported in table B.1, in appendix B.***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

channels. First, in line with the prediction of the duration risk channel, the compression in yields is more pronounced at longer maturities; for instance, the impact on 20-year yields is around 50 percent higher than the impact on 5-year yields. Second, the APP effects are stronger in countries with lower creditworthiness, which is consistent with the credit risk channel. Specifically, yield declines in lower-rated countries, such as Spain and Italy, are more pronounced than those in higher-rated countries such as Germany and France, and these declines tend to be more sizable at longer maturities. For instance, at 10-year maturity, Italian and Spanish yields decline between 10 and 15 basis points more than euro-area yields, and around 35 basis points more than German yields. To shed further light on this finding, table 3 reports the estimated effects on CDS spreads for Italy and Spain. The upshot is that the decline in sovereign spreads largely reflects a narrowing in CDS spreads, in particular for Italy, suggesting that the repricing of credit risk is an important driver of bond price movements. This evidence is consistent with our model's prediction, which implies that asset purchases induce a stronger decline in the compensation for risk on securities with higher credit risk.

Third, and in line with this prediction, we estimate relevant spillover effects on assets that were not targeted by the January 2015 APP. As reported in table 3, we find, for instance, that the program has compressed the spreads of BBB-rated bonds relative to risk-free rates by about 30 basis points for both euro-area financial and nonfinancial corporations.²¹

An alternative interpretation of the sizable APP impact on sovereign spreads could be that market participants viewed the APP as signaling the willingness of the ECB to scale up its policy support for high-yield member countries. This interpretation is, however, largely speculative because, as noted above, the APP interventions were carried out in proportion to the size of the economy of individual member countries, as maintained by the ECB since the launch of the APP. Besides, in isolation, this alternative interpretation might not

²¹As documented in table B.4, the ECB recalibrated the APP in March 2016, deciding to also include bonds issued by euro-area nonfinancial corporations in the list of eligible assets.

Table 3. Changes in Sovereign CDS Spreads and Corporate Bond Spreads around the APP Event Dates (basis points)

	Sovereign CDS Spreads (10y)		Corporate Bond Spreads	
	Italy	Spain	BBB Financial	BBB Nonfinancial
	<i>Controlled Event Study</i>			
One-Day Change	-42**	-34**	-18*	-16*
Two-Day Change	-74**	-52**	-31**	-29**
	<i>Standard Event Study</i>			
One-Day Change	-45**	-33**	-20*	-13*
Two-Day Change	-79**	-54**	-35**	-25**

Notes: Based on the 18 event dates reported in table B.1, in appendix B. The corporate bond spreads are vis-à-vis the mid-swap rate of corresponding maturity. ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

explain the spillover effects on nontargeted private-sector securities documented above.²²

4. Effects and Transmission Channels Exploiting Information from Actual Purchases

Since its launch, the APP has been recalibrated on various occasions in the form of extensions and expansions of the program. At the same time, it has been increasingly challenging to identify the APP effects on the basis of asset price reactions to these announcements. The reason is ultimately related to the state-contingent nature of the APP, whereby the ECB has communicated that asset purchases will “*in any case*” be conducted until a sustained adjustment in the path of inflation towards the price stability objective has been achieved. Therefore, having learnt over time how the ECB decisions depend on the outlook for inflation, market participants have gradually revised their expectations about the APP, in response to the stream of economic data releases, over and above the ECB’s official communications. This is reflected in the information in table 4, which displays ECB decisions on selected policy recalibrations and prior expectations formed by market participants and extracted from Bloomberg surveys. The vast majority of respondents correctly predicted the ECB’s policy recalibrations, in terms of both extension and expansion of the APP, as well as the timing of the policy announcements, suggesting a good understanding among market participants of the ECB communication and reaction function.

As a result of these considerations, in this section we employ a regression analysis by exploiting the cross-sectional variation in security-level data on prices and quantities. In practice, we run a cross-section regression of bond returns for individual securities on empirical proxies for local supply, duration risk, and credit risk channels, and control for a set of individual security characteristics. Price changes are computed between September 2014 and October 2016. The regression analysis focuses on public-sector securities, which

²²A mechanism that, like our framework, might help to explain the impact of the APP on credit spreads of both targeted and nontargeted assets is that market participants might have anticipated the improved macroeconomic conditions induced by the APP and revised the pricing of credit risk accordingly.

Table 4. ECB's Announcements on APP (Re)calibrations and Market Expectations

Date	ECB		Pre-announcement Market Expectations (Bloomberg Survey and Analyst Reports)	
	Decision	Size (Intended)	Step-Up in the APP (In % of Respondents)	Size (Median Response)
Jan. 22, 2015	€60 bn per month – public and private securities	€1140 bn	100%	€550 bn
Dec. 3, 2015	Six-month extension at €60 bn per month	€360 bn	100% of which, APP extension: 80%	€360 bn
Mar. 10, 2016	12-month expansion from €60 to €80 bn per month	€240 bn	82% of which, APP expansion: 72%	€180 bn

Notes: Bloomberg has carried out and published the surveys prior to the policy meetings, respectively, in December 2014, November 2015, and March 2016. The number of market participants varies between 50 and 60 across survey releases. Although the exact formulation of the questions tends to vary across releases, market participants have been consistently asked to report whether they expect a step-up in the monetary stimulus, by when, and in which form. The Bloomberg survey in November 2015 is complemented with information extracted from analyst reports to quantify the size of policy recalibration expected at the December 2015 policy meeting.

Table 5. Characteristics of APP Purchases of Public-Sector Securities for the Largest Euro-Area Countries as of October 2016

	Cumulative Net Purchases	Remaining Maturity (Average)	Yield to Maturity (Average)	Coupon Rate (Average)
Germany	€252 bn	8.8 Years	0.1%	2.4%
France	€199 bn	8.3 Years	0.2%	2.5%
Italy	€183 bn	9.4 Years	1.2%	2.8%
Spain	€126 bn	10.2 Years	1.4%	3.6%
Total	€1100 bn	8.8 Years	0.6%	2.6%

accounted for about 85 percent of the APP as at October 2016.²³ Considering changes since September 2014 is intended to account for the building up of market expectations about the program in anticipation of its official announcement, as documented above. The end of the review period in October 2016 allows major effects and transmission channels of the APP to be captured.²⁴

4.1 First Inspection of the ISIN-Level Data

As a first inspection of the ISIN-level data used in the estimation, table 5 illustrates selected characteristics of public-sector security holdings under the APP as October 2016, including the breakdown for the four largest euro-area economies. Overall, APP holdings of sovereign securities amounted to around €1.1 trillion, with an average maturity of 8.8 years, and were primarily concentrated in the four largest euro-area countries, reflecting the fact that asset purchases are distributed across euro-area jurisdictions according to the

²³The public-sector purchase program (PSPP) is one part of the APP. The other parts, which are concerned with the purchase of private-sector securities, are the asset-backed securities purchase program (ABSPP), the covered bond purchase programs (CBPPs), and the corporate-sector purchase program (CSPP). As noted above, bonds issued by euro-area nonfinancial corporations were included in the list of eligible assets in the March 2016 policy recalibration.

²⁴The analysis here does not cover subsequent policy recalibrations, including the decision made in September 2019 to relaunch net asset purchases. A comparison of the relative impacts of early and late recalibrations is left for future research.

ECB's capital key.²⁵ Yields to maturity were on average low, which is consistent with the contained financial distress and compressed risk premiums at the time. At the same time, euro-area yields conceal some cross-country heterogeneity, with yields in lower-rated countries, such as Spain and Italy, being higher than in higher-rated countries, such as Germany and France.

We formally evaluate the distance between the maturity structures of purchases and outstanding amounts using the Hellinger distance metric, as suggested by, for instance, Huther, Ihrig, and Klee (2017). This metric is given by $N_t \equiv 1 - (\sum_i \sqrt{s_{it}y_{it}})$, where s_{it} is the share of each individual security purchased by the ECB relative to its total purchases, and y_{it} is the corresponding share for outstanding securities. When s_{it} and y_{it} are equal, the Hellinger metric equals zero, signaling perfect neutrality. The Hellinger distance metric for the APP holdings evaluated in October 2016 is around 0.1, indicating a good match between the two distributions, which is consistent with the ECB's stated aim of achieving market neutrality in relation to its interventions.

4.2 *Empirical Proxies for Local Supply, Duration Risk, and Credit Risk Channels*

Regarding the construction of the empirical proxies for the local supply channels, we extend the approach of D'Amico and King (2013) to account for different creditworthiness across euro-area securities. Under a narrow interpretation of local supply channels, the price of a security n is influenced by the purchased amount of that particular security relative to its outstanding amount (referred to as "local supply – own purchases"). Under a slightly broader interpretation, the price of a security n may also be affected by purchases of similar securities ("local supply – purchases of substitutes"). We derive the two corresponding empirical proxies as follows. First, the set of euro-area public-sector securities is partitioned into two classes $I = 0, 1$, corresponding to high and low creditworthiness,

²⁵The ECB's capital key reflects the respective country's share in the total population and GDP of the European Union (EU). For the purpose of the PSPP, the capital key is rescaled to reflect only the EU countries that have adopted the euro and were eligible under the APP.

respectively, depending on whether their credit ratings were above or below A– at the time of the APP announcement.²⁶ Second, for each purchased security n , we define corresponding buckets of “substitutes” by partitioning the set of securities that belong to the same credit class as security n according to their “proximity” to n . Proximity is defined in terms of the difference between remaining maturities. In practice, for each security n , we consider a set of “near substitutes,” S_n , comprising all securities belonging to the same rating class as security n with remaining maturities within three years of the maturity of security n . Thus, we compute the following empirical proxies for local supply effects $h_{n,i} = H_{n,i}/O_{n,i}$, where i takes the value 0 for the partition comprising only the security n itself, and value 1 for the partition of its “near substitutes”; $H_{n,i}$ and $O_{n,i}$ refer, respectively, to the euro amounts of purchased and outstanding quantities for partition i associated with security n . For the partition of “near substitutes” (i.e., $i = 1$) we have $H_{n,1} \equiv \sum_{j \in S_n} H_j$ and $O_{n,1} \equiv \sum_{j \in S_n} O_j$, where the sum is over all securities j belonging to the set of “near substitutes” of security n , with H_j and O_j being, respectively, the purchased and outstanding amounts of security j .

Unlike D’Amico and King (2013), our regression analysis also includes empirical proxies for the duration risk and credit risk channels, which are derived as follows. The proxy for the duration risk channel builds on the security-level risk premiums defined in equation (4). These premiums are determined by the exposure of the individual security n to aggregate duration risk (ADR). Drawing from equation (3), the empirical proxy for the ADR is derived in terms of the amount of 10-year equivalents absorbed by the ECB through the APP and defined as follows:

$$ADR \equiv \sum_n \omega_n * \tilde{d}_n, \quad (7)$$

²⁶Similarly to the approach followed for the high-frequency regression, the partition of high-creditworthiness securities comprises sovereign securities issued by those member countries that, upon the launch of the APP, had been assigned a credit rating of A– or above by at least one of the three main rating agencies (Moody’s, Standard & Poor’s, and Fitch Ratings). The remaining securities eligible under the APP belong to the partition of low-creditworthiness securities, comprising securities issued by Italy, Spain, Portugal, and Cyprus. Sovereign securities issued by Greece were not eligible under the APP.

where $\omega_n \equiv H_n / \sum_n O_n$, with H_n and O_n being, respectively, the purchased and outstanding amounts of security n ; and \tilde{d}_n is the duration of the individual purchased security (d_n) relative to the duration of the 10-year benchmark security. We then make use of the *ADR* to derive the security-level proxy for the duration risk channel, dr_n , defined as a close empirical counterpart to equation (4)

$$dr_n \equiv ADR * \frac{(1 - \exp(-kd_n))}{k}, \tag{8}$$

where, as documented in the model description, k governs the degree of concavity in the sensitivity function of bond prices to risk factors.²⁷ For the purpose of our estimation, as noted in the next section, k is set so as to maximize the regression fit.

To capture the credit risk channel, we draw on the model’s prediction that the risk premium equation (4), and by extension its empirical counterpart (8), takes the same functional form for securities with and without credit risk. At the same time, the model predicts that securities with credit risk entail higher price sensitivity to central bank asset purchases in particular at longer maturities. We adopt a parsimonious way to capture this twofold dimension and consider the following empirical proxy for the credit risk channel:

$$cr_n \equiv I_n * dr_n, \tag{9}$$

where the indicator variable I_n takes the value 1 for securities belonging to the partition of lower credit rating (securities with rating below A-) and 0 otherwise. We then test empirically the additional price sensitivity of securities with higher credit risk by estimating the regression coefficient attached to the regressor cr_n .

4.3 Regression Results Using Security-Level Data

In summary, we run the following regression:

$$R_n = \sum_{i=0}^1 \alpha_i h_{n,i} + \beta dr_n + \theta cr_n + \sum_{i=1}^m \delta_i z_{n,i} + \varepsilon_n, \tag{10}$$

²⁷In the model, securities are zero-coupon bonds, and hence their duration (d_n) equals their maturity (n).

**Table 6. Estimated (stock) Effects and
Transmission Channels Using Security-Level
Data on Purchased Quantities**

	(1)	(2)	(3) Baseline
Local Supply – Own Purchases	0.0080 (0.3783)	0.0299 (1.5572)	0.0352* (1.9374)
Local Supply – Purchases of Substitutes	-0.1440 (-1.3899)	-0.0623 (-0.8090)	-0.0856 (-1.0595)
Duration Risk	0.3356*** (15.8803)	0.1934*** (8.8444)	0.2349*** (6.8488)
Credit Risk	0.0620*** (5.9797)	0.0565*** (7.6381)	0.0426*** (6.7613)
Remaining Maturity Squared	6.6074e-05*** (3.8957)	5.6911e-05*** (3.9955)	6.4655e-05*** (5.3235)
Log Initial Price		0.2500*** (11.5941)	0.1569*** (3.1032)
Coupon Rate			0.6049*** (2.9621)
Intercept	-0.0373*** (-4.2097)	-1.1804*** (-11.8171)	-0.7707*** (-3.3765)
Adj. R ²	88%	92%	94%
No. Observations	632	632	632

Notes: t-statistics are in parentheses. ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

where R_n is the gross return on security n and, as described above, h_n , dr_n , and cr_n are the empirical proxies for, respectively, the local supply, duration risk, and credit risk channels. The set of control variables z_n comprises security-level characteristics such as the coupon rate, the remaining maturity squared, and the (log of) initial prices. Regression equation (10) boils down to the specification proposed by D’Amico and King (2013) for $\beta = \theta = 0$ in the absence of the duration risk channel (dr_n) and the credit risk channel (cr_n).

Table 6 reports the estimated coefficients for three regression specifications which differ for the inclusion of alternative control variables. The baseline specification is the one reported in the third column and controls for all individual security’s characteristics, as motivated by their strong statistical significance. We set the parameter k underpinning the construction of the proxies for the duration risk and credit risk channels equal to 0.1 so as to maximize

the regression fit and then explore the sensitivity of our results to alternative values of k .²⁸

Starting from the local supply channel, the estimated coefficient of “*local supply – own purchases*” is positive and statistically significant. Quantitatively, the estimate of 0.0352 means that purchasing 10 percent of the outstanding amount of the security n leads to a 0.352 percent increase in the price of that security, which in turn implies a yield decline of about 4 basis points for a typical euro-area 10-year sovereign bond whose modified duration is around eight years. The estimated coefficient of “*local supply – purchases of substitutes*” is not statistically significant. This is the case across the various specifications in table 6, as well as in the robustness section below. However, the coefficient estimates on both the duration and credit risk regressors are positive, highly significant, and remain stable across various specifications. Quantitatively, the coefficient estimates imply that the duration and credit risk channels dominate the local supply channels. To illustrate this, we consider a standardized public-sector purchase program amounting to 10 percent of euro-area GDP in 2015 (i.e., around €1.0 trillion), hence slightly above the intended purchases of public-sector securities underpinning the January 2015 announcement. Our regression estimates imply an overall downward shift in GDP-weighted euro-area 10-year sovereign bond yields of around 65 basis points. This effect is comparable with that derived on the basis of the event-study analysis reported above. Local supply channels account for around 5 basis points. The estimated coefficient of the “duration risk” regressor implies a much higher contribution in the order of 50 basis points, with the compression in yields being generally more pronounced for securities with higher duration. The estimated credit risk coefficient implies an additional yield decline of around 10 basis points. We estimate an additional 15 basis points decline in 10-year yields for less creditworthy sovereigns again via the credit risk channel.

We examine the robustness of our estimates with respect to two specific assumptions underpinning our baseline specification. First, we contemplate alternative partitions for the derivation of the

²⁸As proved by McFadden (2001), this approach in selecting k is equivalent to direct estimation of the full set of parameters using nonlinear least squares; however, standard errors provided by least squares are biased.

“*purchases of substitutes*” regressor; specifically, we consider two substitute buckets comprising securities with remaining maturities within one and two years, respectively, of security n 's maturity. Second, we explore alternative values for the parameter k which governs the sensitivity of bond prices to risk factors in the duration risk and credit risk proxies.

The results presented in table 7 support the robustness of our main findings, as reflected in the stability of the estimates across the different specifications. The effects of “*own purchases*” remain statistically significant, albeit economically contained, while the effects of “*purchases of substitutes*” are not statistically significant, as in the baseline specification. The duration and credit risk channels are statistically significant and economically sizable. Our estimates are particularly robust to the alternative specifications for “*purchases of substitutes*,” supporting the view that the local supply channels are well identified relative to the other channels. When exploring alternative values for k , the regression fit, as captured by the R-squared, tends to deteriorate and coefficient estimates change somewhat. Higher values for k , i.e., a higher degree of concavity of the sensitivity function, are associated with larger coefficient estimates on the duration risk regressor, while the overall impact on longer-term yields is not tangibly different from the one underpinning our baseline specification.

5. Interpretation of Our Findings and Possible Implications

Direct comparability across studies is often challenging because of differences in the scope and breadth of various purchase programs, in the way pre-announcement market expectations are controlled for, and in the empirical methodologies employed. Bearing this in mind, the stock effects we find appear to be broadly in line with other studies on the APP. For instance, using a term structure model, Eser et al. (2019) find the same ballpark impact, estimating a compression of euro-area 10-year yields of around 50 basis points with an APP of roughly 10 percent of euro-area GDP in 2015. De Santis (2020), using an index of Bloomberg news, finds a reduction in euro-area 10-year sovereign bond yields of 72 basis points for an equivalent amount of asset purchases. Also, our estimated effects of

Table 7. Estimated (stock) Effects and Transmission Channels Using Security-Level Data on Purchased Quantities: Robustness

	Purchases of Near Substitutes		Sensitivity to Risk Factors	
	Within One Year	Within Two Years	$k = 0.05$	$k = 0.3$
Local Supply – Own Purchases	0.0357** (1.9604)	0.0352* (1.9334)	0.0360** (1.9628)	0.0345* (1.8683)
Local Supply – Purchases of Substitutes	-0.0658 (-1.5907)	-0.0717 (-1.2870)	-0.0261 (-0.3882)	-0.0823 (-0.6790)
Duration Risk	0.2361*** (8.4933)	0.2356*** (7.5016)	0.1859*** (7.1320)	0.4710*** (5.6397)
Credit Risk	0.0426*** (6.7790)	0.0426*** (6.7385)	0.0360*** (7.2899)	0.0831*** (5.6543)
Remaining Maturity Squared	6.3394e – 05*** (4.7767)	6.3928e – 05*** (4.9269)	-3.7865e – 06 (-0.2149)	1.7436e – 04*** (13.9866)
Log Initial Price	0.1543*** (3.1129)	0.1555*** (3.0835)	0.1669*** (3.4607)	0.1885*** (4.2385)
Coupon Rate	0.6076*** (3.0307)	0.6076*** (2.9933)	0.5872*** (2.9949)	0.5013*** (2.6094)
Intercept	-0.7616*** (-3.4608)	-0.7661*** (-3.4001)	-0.8131*** (-3.7410)	-0.9519*** (-4.8315)
Adj. R ²	94%	94%	93%	92%
No. Observations	632	632	632	632

Notes: t-statistics are in parentheses. ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

the APP tend to be within the (admittedly wide) range of estimated effects of the LSAPs. Various studies find that Federal Reserve asset purchases amounting to 10 percent of U.S. GDP reduce the 10-year U.S. Treasury yield by between 37 and 165 basis points.²⁹

A possible way to interpret our findings regarding the relative importance of the different transmission channels, also in relation to this literature, is through the lens of our illustrative model. According to the model's main predictions, during periods of high risk aversion, asset purchases might end up operating locally, with limited spillovers to nontargeted segments. Conversely, during periods of less-extreme financial stress, such as the one prevailing during the APP, investors are effective in diversifying the total amount of risk borne in their portfolios across market segments. The compression of premiums reflects the overall quantity of risk absorbed by the central bank and the exposure of the securities to risk factors. As a result, securities with longer duration and higher credit risk would experience the highest returns, in line with our findings.

A number of studies lend additional support to this possible interpretation. For instance, when assessing the Federal Reserve's LSAPs, Krishnamurthy and Vissing-Jorgensen (2011, 2013) find that large portfolio balance effects are generally confined to targeted assets under LSAP1, and namely at times of heightened financial distress. Similarly, D'Amico and King (2013) document significant local supply effects in the U.S. Treasury yield curve using security-level data during the course of LSAP1. For the United Kingdom, McLaren, Banerjee, and Latto (2014) also provide supporting evidence of local supply effects on U.K. government bond (gilt) yields, identified by exploiting Bank of England announcements about the maturity distribution of asset purchases carried out in the aftermath of the collapse of Lehman Brothers.³⁰ Similar messages emerge from studies of two earlier ECB programs announced in 2010 and 2012 at a time of market distress, although those programs differ in breadth and scope from the APP and large-scale asset purchases carried out

²⁹See Gagnon et al. (2011), Krishnamurthy and Vissing-Jorgensen (2011), Meaning and Zhu (2011), Swanson (2011), D'Amico et al. (2012), Cahill et al. (2013), D'Amico and King (2013), and Li and Wei (2013).

³⁰Similarly, for the United Kingdom, Breedon, Chadha, and Waters (2012) reach the conclusion that individual sovereign bond purchase operations had limited pass-through to other assets.

in the United States and the United Kingdom. For instance, Eser and Schwaab (2016) find evidence of local supply effects in segmented markets and liquidity market effects when assessing ECB bond purchases in stressed sovereign markets under the Securities Markets Programme (SMP) carried out in 2010–11. Similarly, when assessing a subsequent program of outright purchases of sovereign securities (Outright Monetary Transactions, OMTs) announced around the peak of the European debt crisis in the summer of 2012, but never activated, Altavilla, Giannone, and Lenza (2016) estimate announcement effects concentrated in maturity brackets targeted by the program, with limited spillovers to nontargeted maturity brackets and market segments. When assessing the financial market impact of these two programs, Krishnamurthy, Nagel, and Vissing-Jorgensen (2018) also document relevant sovereign bond segmentation effects. Conversely, the findings of Meaning and Zhu (2011), which are predicated on the same methodology developed by D’Amico and King (2013), suggest that the yield effects of LSAP2 work predominantly via duration risk channels, while local supply channels make a more limited contribution. Likewise, Li and Wei (2013) find significant effects of asset purchases on longer-term rates via duration channels on the basis of a term structure model augmented with supply factors and estimated over a pre-crisis period of good market functioning.

6. Conclusions

In this paper, we assess the bond yield effects and transmission channels of the ECB Asset Purchase Programme (APP). The APP provides helpful insights because the purchases took place under relatively low financial stress, particularly when compared with programs carried out in advanced economies in the immediate aftermath of the global financial crisis. Moreover, a distinct novelty of the APP when compared with large-scale asset purchases carried out in the United States and the United Kingdom is that it targets long-term sovereign securities spanning different degrees of creditworthiness. Our identification of the transmission channels is twofold. First, we draw on market reaction to distinct policy announcements, notably in the form of high-frequency asset price responses to news about the size and maturity distribution of asset purchases. Second, we

employ regression analysis that exploits cross-sectional variations in security-level data on prices and purchased quantities for sovereign securities. We find economically significant financial market effects working primarily via “broad channels”: bonds exposed to interest rate risk (duration risk channel) and credit risk (credit risk channel) experienced the highest returns. By contrast, local supply effects are estimated to play a limited role. Specifically, based on our regression estimates, we find that ECB asset purchases amounting to 10 percent of euro-area GDP in 2015 (i.e., around €1.0 trillion) compress GDP-weighted euro-area 10-year sovereign bond yields by around 65 basis points. In terms of relative strength of transmission channels, our estimates imply that the duration risk channel accounts for the bulk of the impact on euro-area 10-year yields, the credit risk channel for around 15 percent, and the local supply channel for around 7 percent. We estimate an additional decline of 15 basis points in 10-year yields for less creditworthy sovereigns via the credit risk channel. These findings have important implications for the future design of QE programs. At times of high financial distress, local supply channels might prevail, limiting the pass-through of the policy stimulus to nontargeted assets. Therefore, central banks are advised to broaden the spectrum of assets they purchase to deliver more favorable financial conditions across market segments. On the other hand, at times of low financial distress, central banks relying on asset purchases to stabilize the economy and circumvent the lower bound on the short-term interest rate can count on investors to facilitate the transmission of their QE stimulus beyond targeted segments.

Appendix A. A Reference Model of Bond Supply Effects

At the center of the model economy is the interaction between two types of agents: the arbitrageurs and the preferred-habitat investors. The arbitrageurs have limited risk-bearing capacity and a mean-variance objective function defined as

$$E_t R_{(t,t+1)}^P - \frac{1}{2} \sigma \text{Var}_t R_{(t,t+1)}^P, \quad (\text{A.1})$$

where σ is the risk-aversion coefficient and $R_{(t,t+1)}^P$ is the portfolio's return given by

$$R_{(t,t+1)}^P = \sum_{n=1}^N \omega_t^{(n)} R_{(t,t+1)}^{(n)} = \sum_{n=1}^N \omega_t^{(n)} [\exp(\bar{p}_{t+1}^{(n-1)} - \bar{p}_t^{(n)}) - 1], \tag{A.2}$$

where $\omega_t^{(n)}$ is the fraction of arbitrageurs' portfolio (relative to their net wealth W_t) held in n -period bonds, and $R_{(t,t+1)}^{(n)}$ is the one-period holding return of purchasing an n -period bond at time t at (log) price $\bar{p}_t^{(n)}$ and selling it at $t + 1$ with residual maturity of $n - 1$ at (log) price $\bar{p}_{t+1}^{(n-1)}$. We extend the model by Vayanos and Vila (2009) and contemplate the case that these zero-coupon bonds are subject to credit risk, meaning that

$$\bar{p}_{t+1}^{(0)} = \begin{cases} 1 & \text{with probability } \exp(-\psi_{t+1}) \\ 0 & \text{with probability } 1 - \exp(-\psi_{t+1}), \end{cases} \tag{A.3}$$

where the time- t credit risk intensity ψ_t is assumed to be affine in a set of macroeconomic factors,

$$\psi_{t+1} = \gamma' X_{t+1}.$$

In turn, macroeconomic factors follow a VAR process,

$$X_t = \mu + \Phi X_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma \Sigma').$$

To solve for the pricing equation, we conjecture that (log) bond prices are also affine functions in the set of macroeconomic factors

$$\bar{p}_t^{(n)} = -\bar{a}_n - \bar{b}'_n X_t \tag{A.4}$$

and the continuously compounded yield $y_t^{(n)}$ on n -period bond is given by $-\bar{p}_t^{(n)}/n$.

The first-order conditions of the arbitrageurs' optimal portfolio allocation are given by

$$\begin{aligned}
\frac{\partial L_t}{\partial \omega_t^{(n)}} &: (-\bar{a}_{n-1} - (\bar{b}'_{n-1} + \gamma')(\mu + \Phi X_t) \\
&+ \frac{1}{2}(\bar{b}'_{n-1} + \gamma')\Sigma\Sigma'(\bar{b}_{n-1} + \gamma) + \bar{a}_n + \bar{b}'_n X_t \\
&- (\bar{b}'_{n-1} + \gamma')\Sigma\Sigma'\sigma \sum_{n=1}^N (\omega_t^{(n)}(\bar{b}_{n-1} + \gamma)) - \varkappa_t = 0 \\
\frac{\partial L_t}{\partial \omega_t^{(1)}} &: -\gamma'(\mu + \Phi X_t) + \frac{1}{2}\gamma'\Sigma\Sigma'\gamma + (\bar{a}_1 + \bar{b}'_1 X_t) \\
&- \gamma'\Sigma\Sigma'\sigma \sum_{n=1}^N (\omega_t^{(n)}(\bar{b}_{n-1} + \gamma)) - \varkappa_t = 0 \tag{A.5}
\end{aligned}$$

and they can be expressed in the compact form of expected period returns of holding long bonds in excess of the short rate \bar{r}_t

$$R'_{(t,t+1)} - \bar{r}_t = \bar{b}'_{n-1}\Sigma\Sigma'\lambda_t, \tag{A.6}$$

where

$$\lambda_t \equiv \sigma \sum_{n=1}^N (\omega_t^{(n)} \tilde{b}_{n-1}). \tag{A.7}$$

Excess holding period returns can be decomposed into the quantity of risk $\bar{b}'_{n-1}\Sigma\Sigma'$, and the market price of risk λ_t . Absence of arbitrage requires that λ_t is the same for all bonds, and is endogenously determined by the degree of risk aversion σ , the arbitrageurs' bond holdings at various maturities $\omega_t^{(n)}$, and the associated sensitivity of bond prices to macroeconomic factors $\tilde{b}_{n-1} \equiv (\bar{b}_{n-1} + \gamma)$. The parameter governing the credit risk intensity γ affects \tilde{b}_{n-1} directly, as well as indirectly via \bar{b}_{n-1} , because the latter in equilibrium is also a function of γ as shown below in equation (A.14).

To gain further intuition, let us consider the case in which the short-term rate is the only macro factor and $\gamma = 0$. This case boils

down to Vayanos and Vila (2009), in which \tilde{b}_{n-1} take the compact form of an increasing concave function in the term to maturity n

$$\tilde{b}_{n-1} \equiv \frac{1 - \exp(-kn)}{k}, \tag{A.8}$$

where k is endogenously determined. The implication is that the holdings of longer-term bonds receive larger weights in the equation for λ_t , meaning that arbitrageurs demand a higher compensation to hold a portfolio that is more exposed to interest rate risk. *Ceteris paribus*, when bonds are subject to credit risk, the sensitivity of portfolios' holdings to risk factors tend to rise, arbitrageurs demand a higher compensation per unit of risk, and so risk premiums increase.

The demand for bonds by preferred-habitat investors is instead given by

$$\xi_t^{(n)} = \varphi(\bar{y}_t^{(n)} - \beta^{(n)}\beta_t), \tag{A.9}$$

where φ is positive. Without loss of generality, assuming constant supply of bonds $S^{(n)}$, the equilibrium condition in the bond market requires

$$\omega_t^{(n)} + \xi_t^{(n)} = S^{(n)}. \tag{A.10}$$

Asset purchases affect bond yields by changing the supply of bonds available to the private sector $S^{(n)}$. This effect might work via two distinct (polar) types of transmission channels, both part of the broadly defined portfolio balance channels, and whose emergence is related to the arbitrageurs' risk aversion. In one case, characterized by heightened risk aversion, arbitrageurs are constrained in the ability to integrate market segments. Equilibrium yields are then pinned down by the demand equation of preferred-habitat investors, jointly with the bond supply. The equilibrium term structure exhibits market segmentation, and so the effects of asset purchases on bond yields are local to those segments targeted by the purchases; this mechanism is known in the literature as "local supply channel." In a second case, when risk aversion is less extreme and yet risk-bearing capacity is limited, arbitrageurs price risk consistently across market segments, in the way described above. Asset purchases reduce the overall amount of risk that arbitrageurs must hold in equilibrium, hence affecting the entire term structure through the reduction of the

market price of risk. The decline in premiums is larger for those securities more exposed to risk factors. This pertains to securities with longer duration—i.e., more exposed to interest rate risk—a mechanism known in the literature as duration risk channel. It also applies to securities more exposed to credit risk, a mechanism which we call credit risk channel. So defined, duration risk and credit risk channels fall within the category of the so-called broad channels of asset purchases. In the latter case, equilibrium bond prices can therefore be derived as follows. Isolating $\omega_t^{(n)}$ in (A.10), and substituting it out in (A.9), we obtain the following expression for the market price of risk

$$\lambda_t \equiv \sigma \sum_{n=1}^N (S^{(n)} - \xi_t^{(n)}) (\bar{b}_{n-1} + \gamma), \quad (\text{A.11})$$

which can be recast in a more compact form as

$$\lambda_t = \lambda_0 \left(\overleftarrow{a}_N, \overleftarrow{b}_N; \sigma, \varphi, \overleftarrow{S}^{(N)}, \overleftarrow{\beta}^{(N)} \right) + \lambda_1 \left(\overleftarrow{b}_N; \sigma, \gamma, \varphi \right) X_t, \quad (\text{A.12})$$

where \overleftarrow{a}_N , \overleftarrow{b}_N , $\overleftarrow{S}^{(N)}$, and $\overleftarrow{\beta}^{(N)}$ collect, respectively, \bar{a}_i , \bar{b}_i , $S^{(i)}$, $\beta^{(i)}$ for all i from 1 to N .

To derive a solution for the pricing equation coefficients, we verify the conjectured solution (A.4) by using equations (A.5), (A.6), (A.11), and (A.9), where in the latter we substitute out for $\bar{y}_t^{(n)}$ using (A.4). In essence, this leads to a set of difference equations for the pricing coefficients that resemble the standard affine term structure pricing equations

$$\begin{aligned} \bar{a}_n &= a_1 + \bar{a}_{n-1} \\ &+ (\bar{b}'_{n-1} + \gamma') \left(\mu + \Sigma \Sigma' \lambda_0 \left(\overleftarrow{a}_N, \overleftarrow{b}_N; \sigma, \gamma, \varphi, \overleftarrow{S}^{(N)}, \overleftarrow{\beta}^{(N)} \right) \right) \\ &- \frac{1}{2} (\bar{b}'_{n-1} + \gamma') \Sigma \Sigma' (\bar{b}_{n-1} + \gamma) \end{aligned} \quad (\text{A.13})$$

$$\bar{b}'_n = b'_1 + (\bar{b}'_{n-1} + \gamma') \left(\Phi + \Sigma \Sigma' \lambda_1 \left(\overleftarrow{b}_N; \sigma, \gamma, \varphi \right) \right), \quad (\text{A.14})$$

where a_1 and b_1 are the pricing equation coefficients of the short-term risk-free rate r_t , the expression for λ_t is rearranged as $\lambda_t \equiv \lambda_0 + \lambda_1 X_t$, and \overleftarrow{a}_N , \overleftarrow{b}_N , $\overleftarrow{S}^{(N)}$, and $\overleftarrow{\beta}^{(N)}$ collect, respectively, \bar{a}_i , \bar{b}_i , $S^{(i)}$, $\beta^{(i)}$ for all i from 1 to N .

Appendix B. Additional Tables and Charts**Table B.1. Identified Event Dates**

Date	Event
September 04, 2014	ECB press conference
September 12, 2014	News conference following a meeting of euro-area finance ministers in Milan
September 24, 2014	President Draghi's interview with Europe 1, conducted on September 23, 2014 and aired on September 24, 2014
September 25, 2014	President Draghi's interview with Lithuanian business daily <i>Verslo Zinios</i>
October 2, 2014	ECB press conference
October 10, 2014	Statement at the Thirtieth meeting of the IMFC, Washington
October 24, 2014	An ECB spokesman reading from President Draghi's speaking points at a euro-area summit, Brussels
November 6, 2014	ECB press conference
November 17, 2014	Introductory remarks by President Draghi at the EP's Economic and Monetary Affairs Committee
November 21, 2014	President Draghi's speech at the Frankfurt European Banking Congress, Frankfurt am Main
November 27, 2014	Introductory remarks by President Draghi at the Finnish parliament and speech at the University of Helsinki
December 4, 2014	ECB press conference
January 2, 2015	President Draghi's interview with <i>Handelsblatt</i> , published on January 2, 2015
January 8, 2015	Letter to Mr. Luke Ming Flanagan (member of the European Parliament), published on January 8, 2015
January 14, 2015	President Draghi's interview with <i>Die Zeit</i> , published on January 15, 2015
January 22, 2015	ECB press conference
March 5, 2015	ECB press conference
March 9, 2015	Start of public-sector security purchases

**Table B.2. Data Releases of Macroeconomic Variable
Used in the Event Study**

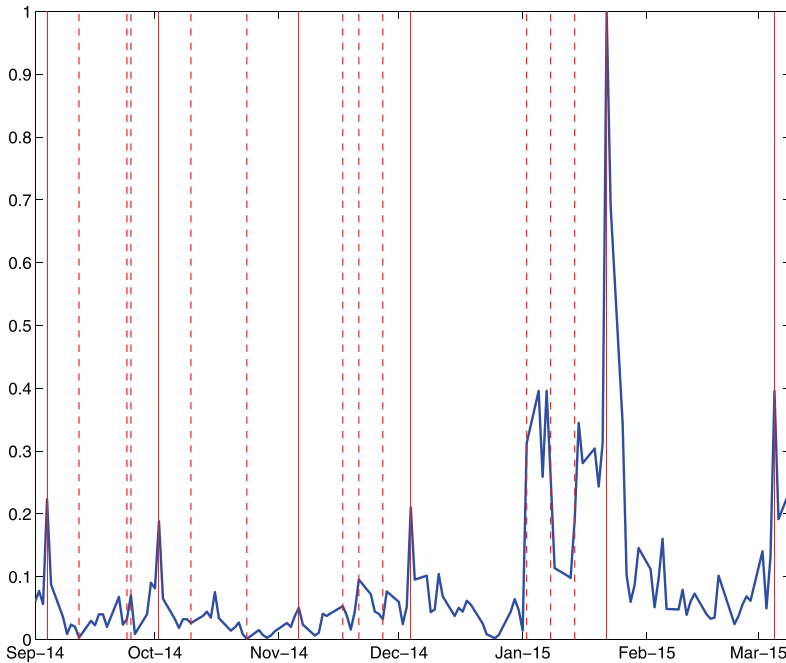
Country	Variable
Euro Area	Consumer Confidence
Euro Area	CPI MoM
Euro Area	Economic Confidence
Euro Area	GDP SA QoQ
Euro Area	Industrial Production SA MoM
Euro Area	Markit Eurozone Manufacturing PMI
France	Consumer Confidence
France	CPI YoY
France	GDP QoQ
France	Industrial Production MoM
France	Markit Manufacturing PMI
Germany	CPI MoM
Germany	GDP SA QoQ
Germany	IFO Business Climate
Germany	Industrial Production SA MoM
Germany	Markit/BME Manufacturing PMI
Germany	Unemployment Rate
Germany	ZEW Survey Expectations
Italy	Business Confidence
Italy	CPI EU Harmonized YoY
Italy	GDP WDA QoQ
Italy	Industrial Production MoM
Italy	Markit/ADACI Manufacturing PMI
Spain	GDP QoQ
Spain	Markit Manufacturing PMI
Spain	Retail Sales YoY
Spain	Unemployment Rate
Spain	CPI EU Harmonized YoY
United States	Chicago Purchasing Manager
United States	Consumer Confidence Index
United States	CPI MoM
United States	FOMC Rate Decision (Upper Bound)
United States	GDP Annualized QoQ
United States	GDP Price Index
United States	Housing Starts
United States	Initial Jobless Claims
United States	ISM Manufacturing
United States	U. of Mich. Sentiment
United States	Unemployment Rate
United States	Change in Nonfarm Payrolls

Table B.3. APP Official Announcement on January 22, 2015, and Market Expectations

ECB Announcement		Pre-announcement Market Expectations (Bloomberg Survey and Analyst Reports)			
Statement	Size (Intended)	Maturities	Probability of APP (% of Respondents)	Size (Median)	Maturities (Median)
€60 bn per Month – Public and Private Securities	€1140 bn	2–30 years	100%	€550 bn	Up to 10 years
<p>Notes: The Bloomberg survey carried out in December 2014, prior to the January 2015 announcement, is employed to derive an estimate about the probability of the APP and expected size of the program. Specifically, the probability of the APP is the percent of survey respondents reporting an expansion of ECB's stimulus in the form of asset purchases; the size of the APP is calculated as the increase in the size of ECB's balance sheet by end-2016 pursued via asset purchases, as reported by survey respondents. Analyst reports are employed to extract the maturity range of the APP expected by market participants.</p>					

**Table B.4. ECB's Announcements on APP
(Re)calibrations**

Data of Announcement	Amount of Purchases	Announcement Details
January 22, 2015	€1140 bn	Combined monthly purchases of public- and private-sector securities amounting to €60 bn, intended to run until end of September 2016, or beyond, if necessary. The maturities range between 2 and 30 years. The purchases of securities issued by euro-area governments and agencies are based on the national central banks' shares in the ECB's capital key.
December 3, 2015	€360 bn	Extension of the APP, with monthly purchases of €60 bn intended to run until the end of March 2017, or beyond, if necessary. Inclusion of debt instruments issued by regional and local governments located in the euro area in the list of eligible assets.
March 10, 2016	€240 bn	Expansion of the monthly purchases from €60 bn to €80 bn, intended to run until the end of March 2017, or beyond, if necessary. Inclusion of investment-grade euro-denominated bonds issued by euro-area nonbank corporations in the list of eligible assets.

Figure B.1. Index of News Coverage on the APP

Notes: The blue line is the index of news computed from Factiva. The vertical (red) solid lines represent the date of the ECB’s Governing Council meetings, i.e., September 4, 2014; October 2, 2014; November 6, 2014; December 4, 2014; January 22, 2015; and March 5, 2015. The vertical (red) dashed lines represent the non-Governing Council events. The index of news coverage use an extensive range of different news sources from the Dow Jones’ news database, Factiva. Specifically, for each calendar day starting from September 1, 2014, we search for a number of keyword variables connected to the announcement and the implementation of the APP. The query is set so that for an article to be included in our sample it should simultaneously contains at least one word coming from two different sets. The first set is “ECB,” “European Central Bank,” and “Draghi.” The second set is “QE,” “quantitative easing,” “asset purchase,” and “APP.” To avoid possible contamination of the results from the quantitative easing programs of other central banks, we exclude the article if it does contain one of the following words: “Federal Reserve,” “Bank of Japan,” “Bank of England,” “BoJ,” “BoE,” “Fed,” “Japan,” “US,” “U.S.,” and “England.” We limited the search to English-language news sources. The total volume of news articles connected to our query over the period spanned from September 2014 to March 2015 is about 20,000, mostly coming from “publication” and “web news.” Blogs, boards, pictures, and multimedia only have a very limited coverage in the selected sample.

References

- Altavilla, C., and D. Giannone. 2017. "The Effectiveness of Non-Standard Monetary Policy Measures: Evidence from Survey Data." *Journal of Applied Econometrics* 32 (5): 952–64.
- Altavilla, C., D. Giannone, and M. Lenza. 2016. "The Financial and Macroeconomic Effects of the OMT Announcements." *International Journal of Central Banking* 12 (3, September): 29–57.
- Andrade P., J. Breckenfelder, F. De Fiore, P. Karadi, O. Tristani. 2016. "The ECB's Asset Purchase Programme: An Early Assessment." ECB Working Paper No. 1956.
- Arrata, W., and B. Nguyen. 2017. "Price Impact of Bond Supply Shocks: Evidence from the Eurosystem's Asset Purchase Program." Working Paper No. 623, Banque de France.
- Breedon, F., J. S. Chadha, and A. Waters. 2012. "The Financial Market Impact of UK Quantitative Easing." *Oxford Review of Economic Policy* 28 (4): 702–28.
- Cahill, M. E., S. D'Amico, C. Li, and J. S. Sears. 2013. "Duration Risk versus Local Supply Channel in Treasury Yields: Evidence from the Federal Reserve's Asset Purchase Announcements." Finance and Economics Discussion Series No. 2013–35, Board of Governors of the Federal Reserve System.
- Christensen, J. H. E., and G. D. Rudebusch. 2012. "The Response of Interest Rates to US and UK Quantitative Easing." *Economic Journal* 122 (564): F385–F414.
- Cochrane, J. 2008. "Comments on 'Bond Supply and Excess Bond Returns' by Robin Greenwood and Dimitri Vayanos." Working Paper, University of Chicago.
- D'Amico, S., W. English, D. López Salido, and E. Nelson. 2012. "The Federal Reserve's Large-Scale Asset Purchase Programs: Rationale and Effects." *Economic Journal* 122 (564): F415–F446.
- D'Amico, S., and T. B. King. 2013. "Flow and Stock Effects of Large-Scale Treasury Purchases: Evidence on the Importance of Local Supply." *Journal of Financial Economics* 108 (2): 425–48.
- De Santis, R. A. 2020. "Impact of the Asset Purchase Programme on Euro Area Government Bond Yields Using Market News." *Economic Modelling* 86 (March): 192–209.

- De Santis, R. A., and F. Holm-Hadulla. 2020. “Flow Effects of Central Bank Asset Purchases on Sovereign Bond Yields: Evidence from a Natural Experiment.” *Journal of Money, Credit and Banking* 52 (6): 1467–91.
- Eser, F., W. Lemke, K. Nyholm, S. Radde, and A. L. Vladu. 2019. “Tracing the Impact of the ECB’s Asset Purchase Programme on the Yield Curve.” ECB Working Paper No. 2293.
- Eser, F., and B. Schwaab. 2016. “Evaluating the Impact of Unconventional Monetary Policy Measures: Empirical Evidence from the ECB’s Securities Markets Programme.” *Journal of Financial Economics* 119 (1): 147–67.
- Gagnon, J., M. Raskin, J. Remache, and B. Sack. 2011. “The Financial Market Effects of the Federal Reserve’s Large-Scale Asset Purchases.” *International Journal of Central Banking* 7 (1, March): 3–43.
- Gilchrist, S., D. López-Salido, and E. Zakrajšek. 2015. “Monetary Policy and Real Borrowing Costs at the Zero Lower Bound.” *American Economic Journal: Macroeconomics* 7 (1): 77–109.
- Greenwood, R., and D. Vayanos. 2014. “Bond Supply and Excess Bond Returns.” *Review of Financial Studies* 27 (3): 663–713.
- Hamilton, J. D., and J. C. Wu. 2012. “The Effectiveness of Alternative Monetary Policy Tools in a Zero Lower Bound Environment.” *Journal of Money, Credit and Banking* 44 (s1): 3–46.
- Huther, J., J. Ihrig, and E. Klee. 2017. “The Federal Reserve’s Portfolio and its Effect on Interest Rates.” Finance and Economics Discussion Series No. 2017–075, Board of Governors of the Federal Reserve System.
- Joyce, M. A. S., A. Lasaosa, I. Stevens, and M. Tong. 2011. “The Financial Market Impact of Quantitative Easing in the United Kingdom.” *International Journal of Central Banking* 7 (3, September): 113–61.
- Joyce, M. A. S., and M. Tong. 2012. “QE and the Gilt Market: A Disaggregated Analysis.” *Economic Journal* 122 (564): F348–F384.
- King, T. B. 2015. “A Portfolio-Balance Approach to the Nominal Term Structure.” Working Paper No. 2013–18, Federal Reserve Bank of Chicago.
- . 2019. “Expectation and Duration at the Effective Lower Bound.” *Journal of Financial Economics* 134 (3): 736–60.

- Koijen, R. S., F. Koulischer, B. Nguyen, and M. Yogo. 2017. "Euro-Area Quantitative Easing and Portfolio Rebalancing." *American Economic Review* 107 (5): 621–27.
- Krishnamurthy, A., S. Nagel, and A. Vissing-Jorgensen. 2018. "ECB Policies Involving Government Bond Purchases: Impact and Channels." *Review of Finance* 22 (1): 1–44.
- Krishnamurthy, A., and A. Vissing-Jorgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity* (Fall): 215–65.
- . 2013. "The Ins and Outs of LSAPs." Paper presented at the Federal Reserve Bank of Kansas City Symposium on Global Dimensions of Unconventional Monetary Policy, Jackson Hole, Wyoming, August 21–24.
- Li, C., and M. Wei. 2013. "Term Structure Modeling with Supply Factors and the Federal Reserve's Large-Scale Asset Purchase Programs." *International Journal of Central Banking* 9 (1, March): 3–39.
- McFadden, D. 2001. "On Selecting Regressors To Maximize Their Significance." In *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. C. Hsiao, K. Morimune, and J. L. Powell, 259–80. New York: Cambridge University Press.
- McLaren, N., R. N. Banerjee, and D. Latto. 2014. "Using Changes in Auction Maturity Sectors to Help Identify the Impact of QE on Gilt Yields." *Economic Journal* 124 (576): 453–79.
- Meaning, J., and F. Zhu. 2011. "The Impact of Recent Central Bank Asset Purchase Programmes." *BIS Quarterly Review* (December): 73–83.
- Rogers, J. H., C. Scotti, and J. H. Wright. 2018. "Unconventional Monetary Policy and International Risk Premia." *Journal of Money, Credit and Banking* 50 (8): 1827–50.
- Swanson, E. T. 2011. "Let's Twist Again: A High-Frequency Event-Study Analysis of Operation Twist and Its Implications for QE2." *Brookings Papers on Economic Activity* (Spring): 151–207.
- Vayanos, D., and J.-L. Vila. 2009. "A Preferred-Habitat Model of the Term Structure of Interest Rates." CEPR Discussion Paper No. 7547.

The Single Resolution Fund and the Credit Default Swap: What Is the Coasian Fair Price of Their Insurance Services?*

Anna Naszodi

Honorary Member of the Centre for Economic and Regional Studies

This paper develops an option-based model to analyze the relationship between two insurances, both providing protection against bank failures. One of these insurances is offered to European banks by the Single Resolution Fund on a compulsory basis in return for their contributions to the Fund, while the other is by the CDS market. The model provides a theoretical framework for testing whether the contributions of banks are fair in the Coasian sense relative to the CDS spreads.

JEL Codes: G28, G13.

1. Introduction

Since January 1, 2016, the Single Resolution Board (SRB), together with the National Resolution Authorities in the member states, is responsible for the resolution of credit institutions and investment firms (henceforth banks) domiciled in the European Union. It takes eight years to build up the Single Resolution Fund (SRF) backing SRB from the contributions of banks and to gradually phase out the national resolutions in the banking union. The contributions

*Dedicated to the memory of my father, Laszlo Naszodi, who once dealt with taxing property while being employed by Mesa County (CO) USA and addressed questions similar to those raised in this paper. The author is grateful for comments from Harrison Hong and an anonymous reviewer. The views expressed in this paper are those of the author and do not necessarily reflect the official views of any policy institution. Author contact: Centre for Economic and Regional Studies (KRTK), Budapest, Tóth Kálmán utca 4, H-1097, Hungary. E-mail: anna.naszodi@gmail.com.

are determined by the Commission Delegated Regulation (EU) 2015/63.¹ Henceforth, we refer to it as the Regulation (with a capital R). In return for the contributions, banks can benefit from the resolution service during the contribution period ending in 2023 and beyond when needed.²

In addition to the service provided by the SRB, insurance against bank failure can also be bought by investors from the market on an optional basis. The most common insurance provided by the market is the *credit default swap* (CDS), which gives the right to the buyer of the protection to swap the bond issued by the bank with its face value in case the issuer bank defaults on its repayment obligation.

As a first step, this paper investigates how much the bondholders of a covered bank benefit from two insurance schemes, i.e., the compulsory one provided by the SRF and the optional one offered by a CDS contract. Specifically, this paper derives the *functional relationship between the values* generated by the two insurances for the bondholders and finds this function to be highly nonlinear. As a second step, the derived functional relationship between the values is used to impose a *normative criterion against the fees* charged for the insurance services. As a third step, the paper proposes and implements a *test on whether the normative criterion* is met in practice.

Since the values are not observable, a theoretical single-bank model is built to derive those as a function of some bank-specific variables, such as the market price of the bank's total assets and

¹The Commission Delegated Regulation (EU) 2015/63 offers different methods for calculating the contributions of different financial institutions. This paper focuses on the risk-adjusted method applicable by big and/or risky banks. The corresponding formulas are presented in the appendix. The motivation for narrowing down the analysis to this method is that the contributions of the big and/or risky banks in 2016 make up 96 percent of the total ex ante contributions to the Fund, although these banks represent only 20 percent of all the institutions under the jurisdiction of the SRB. See <https://www.srb.europa.eu/en/content/2016-ex-ante-contributions>.

²More precisely, until 2023 the SRF, together with the national resolution funds, will be used for any bank resolution. However, the national resolutions are gradually phased out and the SRB will be solely responsible for bank resolutions after 2023. For the sake of simplicity, this paper neither models the intermediate period nor distinguishes between the national resolution funds and the supranational resolution fund by assuming that all the resolution service is provided by the SRB.

its volatility, the leverage, and the maturity of the bank's liabilities. The model proposed in this paper describes the insurances as *options* using a *Merton-type model*. The advantage of this model is that its simplicity allows us to concentrate on the regulatory specificities.

Our model is similar in spirit to the model developed by Necula and Radu (2012) for valuing the liabilities of a recapitalization fund. The common features of these two models are that both rely on the Merton (1974) model, where the underlying asset of the options is the market price of total assets, while the value of the compulsory insurance is a nonlinear function of it. The distinctive feature of the model in this paper is that the strike price and other characteristics of the option capturing the value of the compulsory insurance are chosen in this study so as to reflect the following regulatory specificities: (i) the SRB can intervene only after a bail-in has already taken place, (ii) there is a limit to the funding the SRB is authorized to provide to each bank, (iii) this funding is used for covering losses and not for recapitalizing the bank, and (iv) resolution can happen even without an explicit default.³

Another set of differences between this paper and the paper by Necula and Radu (2012) relates to the estimation of some parameters key to pricing the service provided by the SRB or the liabilities of the recapitalization fund. They calibrate the parameters of the market price of total assets of some banks and the corresponding volatilities to the monthly stock prices of the examined banks and the historical equity volatilities by using the method proposed by Ronn and Verma (1986). By contrast, this paper estimates the above parameters not only from stock prices but also from the actuarial spread calculated and published by the Credit Research Initiative (CRI) from a broad set of variables including CDS data.⁴

³Among these four differences between this paper and the paper by Necula and Radu (2012), the first two can be considered to be solely semantical. First, Necula and Radu (2012) model the intervention point by a threshold parameter (with no reference to the bail-in rule). However, their threshold parameter corresponds to a parameter determined by the bail-in rule in this paper. Second, the ceiling on the premium of the recapitalization service in their model is due to the co-existence of a deposits guarantee fund and not to the regulatory limit on the intervention by the SRB as it is in this paper.

⁴There is a growing literature on testing credit risk models using the information from the CDS market. Huang, Shi, and Zhou (2020) give an overview of

The choice of the applied method in this paper is motivated by the literature: Hull, Nelken, and White (2005) compare two approaches for implementing Merton's model. Of the considered approaches, one is the same as used by Necula and Radu (2012), whereas the other one is closer to the approach applied in this paper.⁵ Hull, Nelken, and White (2005) find that the latter approach usually performs better when the basis of comparison is the goodness of fit of the implied credit spreads on the CDS spreads.

Once one is equipped with the option-based model, one can *compare the model-implied values* of the insurance services provided by the SRF and a CDS contract. Suppose that the outcome of the comparison is that one of the insurances is twice as valuable as the other according to the theory. Provided that the fee charged for this insurance is double the fee charged for the other, that would mean that the fees (or prices, or premiums, or taxes, or levies) are in *parity* with the theoretical values. Given that the service of the SRF is not market based, and the fee charged for it is not determined by the logic of the market, but by law, no mechanism guarantees the parity condition to hold in reality.⁶

Why does the parity condition qualify to be a normative criterion against the Regulation determining the fee for the compulsory insurance? To answer this question, it is important to make the following remarks. First, theoretically, the debtholders of banks could voluntarily establish a resolution fund and could divide the related cost among themselves following the logic of Coasian bargaining.⁷

the literature and conduct a specification analysis of various structural credit risk models, including Merton's model, using the term structure of CDS spreads and equity volatility from high-frequency return data.

⁵Hull, Nelken, and White (2005) use the implied volatilities of options on the company's equity, while this paper exploits the information in the actuarial spread.

⁶Another peculiarity of the contributions collected by the SRB on top of the fact that they should not necessarily meet any equilibrium conditions is that these are paid from the profit of the banks, while the primary beneficiaries of the service of bank resolutions are not the owners of the banks but the debtholders, as recapitalization of banks by the SRB involves writing down shareholders' value to zero. Investigating empirically whether the cost of insurance is passed over to the debtholders remains for future research.

⁷As is pointed out by Tirole (2010, p. 3), a necessary assumption for the feasibility of a spontaneously developed vehicle for extending liability to a third party, such as a privately established resolution fund, is that this third party has

Second, the parity condition ensures that the fee charged by the resolution authority is equal to the value generated by the resolution service for the debtholders of each bank under the assumption that the CDS market is efficient. For the above two reasons, the parity condition offers a possible cake-cutting that provides at least as much utility for the players as the opt out from the Coasian game,⁸ i.e., the resulting allocation of the cost is in the core of the game.

Why should the parity condition not necessarily be a normative criterion against the Regulation? First, the core of the Coasian game is not necessarily uni-element, but it can contain vectors of contributions other than the one fulfilling the parity condition. This is not surprising, as the creation of a public good typically enhances the “cake.” In our specific case of the SRF, the cake is enhanced due to the positive externality of reducing the risk of contagion among banks, and both the debtholders and the shareholders of each bank benefit from the fact that other banks are also covered by the compulsory insurance. Second, besides the Coasian approach, its natural alternative, the Pigovian approach, also offers a solution for internalizing externalities.⁹

The model in this paper disregards some of the externalities mentioned in the first point and assumes the total value created by the SRF to be equal to the direct benefits generated exclusively for the debtholders. Regarding the second point, the model is built on the assumption that the Regulation is on the ground of Coasian fair pricing.

sufficiently deep pockets to cover even the large damages occurring during a bank crisis. In other words, the market-based solution can work only if the resolution fund cannot be “judgment proof,” in legal terms.

⁸See Coase (1960).

⁹There is disagreement in the academic literature on whether the Coasian or the Pigovian approach should be followed. For instance, Goodhart and Schoenmaker (2009) explore possible ex ante mechanisms for fiscal burden sharing in a banking crisis in Europe by expanding the model by Freixas (2003). Their mechanisms rely on the logic of the Coasian approach (although not declared explicitly), as those countries are assumed to shoulder a larger part of the burden that benefit more from the public good of financial stability. The option-based model by Necula and Radu (2012) also offers a method to determine the Coasian fair contributions. In contrast, Brunnermeier et al. (2009) and Schoenmaker (2010) advocate the Pigovian tax.

The rest of the paper is structured as follows. Section 2 presents a simple analysis. Section 3 introduces the theoretical model. Section 4 derives some policy-relevant implications, proposes and implements a test on whether the contributions are fair in the Coasian sense, and presents an example for calculating the contribution of a hypothetical bank. Finally, section 5 concludes.

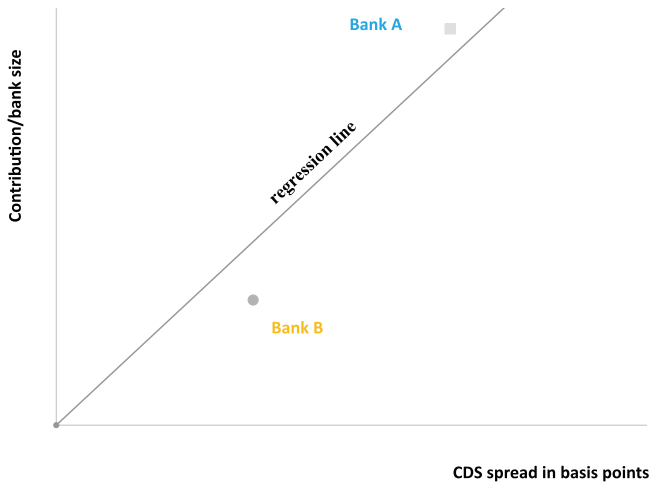
2. A Simple Analysis and Its Limitations

Suppose that the managers of a hypothetical bank A find the contribution payable by their bank to the SRF to be unfairly high and they propose a change in the parameters of the Regulation. Their argument is as follows. First, banks with zero market-perceived risk should not pay any contribution. Henceforth, we refer to this criterion as the “*zero-risk criterion*.” Second, the contribution (relative to the size of the bank) should be proportional to the CDS spread, as both the CDS and the SRF provide insurance against the same event that is the failure of the bank. Henceforth, we refer to this criterion as the “*proportionality criterion*.” As a consequence of these criteria, the relationship between the CDS spreads and the contributions (normed by the bank size) should be linear with a zero intercept. Finally, by running a linear regression on bank-level data, the managers find that bank A is overcharged, while bank B is undercharged by the SRF relative to the market-provided insurance since the former is above the regression line, while the latter is below. This is illustrated in figure 1.

Is this argument correct? Should the “zero-risk criterion” and the “proportionality criterion” be met? Should banks with higher market-perceived risk (with higher CDS) contribute more in accordance with an intuitive criterion that we call the “*monotonicity criterion*”? Should the regulator consider changing the parameters of the Regulation if any of the above three criteria is violated? Should the regulator think that there is no need to revise the Regulation if the slope in the linear regression is positive and the goodness of fit is perfect or reasonably good?

As is shown in this paper, the answers to the above questions are: no, no, no, no, and no, respectively. Although the argument of the bank managers is intuitive, it is wrong and their simple theory on fair pricing and their empirical analysis is misleading. What

Figure 1. The Regression Line Fit on Data of Two Hypothetical Banks



makes the intuition fail is that important differences between the two insurance services are overlooked. Still, their simple analysis shows clearly that *one needs to rely on some kind of theory* to be able to judge whether the contributions are fair. In the following, a theoretical model is developed that is more suitable to determine what the relationship should be between the fees.

3. Benchmark Option-Based Model

This section develops a single-bank Merton's-type model¹⁰ in order to derive the values of two insurance services: one is offered by the CDS market, while the other is provided by the SRB. To do that, we impose the simplifying assumptions that the bank has only one type of debt, which is a zero-coupon bond. In this setup, a bank failure can happen only at the maturity of the bonds.

Under these assumptions, the classical Merton model suggests that the equity of the bank is a European-style call option on the total assets of the bank with the strike price being the face value of

¹⁰See Merton (1970, 1974).

the debt. Similarly, going long on bonds is equivalent to holding the following portfolio: (i) going short a European-style put option on the assets of the bank with the strike price being also the face value of the bonds, and (ii) going long the present value of the strike price. Formally,

$$Equity_t = Call_t^{European}(A_t, K, T-t, \sigma, r, q), \quad \forall t \leq T, \quad (1)$$

$$Bonds_t = Ke^{-r(T-t)} - Put_t^{European}(A_t, K, T-t, \sigma, r, q), \quad \forall t \leq T, \quad (2)$$

where A_t denotes the market price of total assets at time t and K is the face value of the debt, i.e., the principal amount that needs to be repaid to the debtholders at maturity T . The risk-free rate and the yield of return on the underlying asset are denoted by r and q , respectively. Finally, σ is a vector of parameters describing the process of the market price of total assets. In general, σ contains those parameters that describe the deterministic drift, and the probability distribution of the random term of the process. For instance, if the process is determined by one of the simplest models, the Cox-Ross-Rubenstein (CRR) binomial model,¹¹ then σ is unielement and contains only the volatility of the underlying asset.

Now, let us see how one can model the values of the insurances by options. Suppose that one buys the bonds together with the insurance provided by the CDS and the bank defaults on the bonds at maturity. Then, the buyer and the seller of the CDS swap the defaulted bonds and money in the amount of the face value. More precisely, the buyer gives the defaulted bonds to the seller of the protection and in return receives the face value of the bonds. In the case of no default, the bank pays the face value of the bonds to the bondholder at maturity. In either case, the owner of the portfolio of the bonds and the CDS gets the face value of the bonds:

$$Bonds_T + CDS_T^{value} = K. \quad (3)$$

¹¹See Cox, Ross, and Rubenstein (1979).

For the sake of simplicity, let us assume that the risk-free rate is zero:

$$r = 0. \quad (4)$$

Under this assumption, the value of the risk-free portfolio consisting of the bonds and the CDS is equal to the face value of the bonds, K , even for $t < T$:

$$Bonds_t + CDS_t^{value} = K, \quad \forall t \leq T. \quad (5)$$

By combining equations (2), (4), and (5), we obtain that the value of the CDS is equivalent to the price of a European-style put option:

$$CDS_t^{value} = Put_t^{European}(A_t, K, T - t, \sigma, r, q), \quad \forall t \leq T. \quad (6)$$

Next, let us see how the insurance provided by the SRF can be modeled. Similar to the CDS, it can also be described as a put option, but with some specific characteristics reflecting the differences between the conditions of payoffs of the two insurance schemes. There are three important differences that our model captures. The first is due to the bail-in rule, i.e., the SRB can intervene only after a bail-in of 8 percent of liabilities has already taken place. This shifts the strike price of the option describing the value of the service provided by the SRF relative to the strike price describing the CDS by 8 percent of the total liabilities.

Second, there is a limit to the funding the SRB is authorized to provide. The funding cannot exceed 5 percent of the total liabilities including own funds, and it is used only for covering losses and not for recapitalizing the banks. This is captured in the model by putting a cap on the value of the SRF.¹²

Third, regarding the style of the put option that best describes the service provided by the SRF, we can say that it is an *American*

¹²In reality, there is no obligation for the SRB to intervene automatically after 8 percent of the liabilities are bailed-in and the SRB can spend less than the ceiling of 5 percent of the liabilities on a secured bank. The model in this paper disregards the possibility of any discretionary decisionmaking from the side of the SRB. By that, the derived theoretical value of the compulsory insurance overestimates the corresponding value in practice.

one, as resolution can happen anytime (even before an explicit default):

$$SRF_t^{value} = \min[Put_t^{American}(A_t, K - 0.08L, T - t, \sigma, r, q), 0.05L], \quad \forall t \leq T, \quad (7)$$

where SRF_t^{value} denotes the time- t value of the insurance provided by the SRF, while the book value of total liabilities of the bank is denoted by L .

Modeling how the above three specificities affect either the benefits that the SRF provides to the bondholders or the contingent liabilities of the SRF (which is just the mirror image of the benefits) brings us closer to understanding the Regulation.¹³

As an alternative to the option-based approach, one could build a model from scratch, i.e., by using stochastic calculus to derive how the values of the insurances depend on the process of the market price of total assets. The main motivation for choosing the option-based model instead is that our general knowledge on option pricing provides us shortcuts to some results.¹⁴

3.1 Option Valuation

This section elaborates on how one can value the options describing the insurances provided by the CDS and the SRF. First, let us make it explicit how the values of the options depend on the market price

¹³Due to the obvious limitation of the model, it does not account for some further specificities of the Regulation. For instance, one could argue that the style of the option capturing the service of the SRF is exotic, as once the bank is resolved, its compulsory insurance does not expire but instead gets renewed automatically. This kind of renewability was typical to the practice of the National Resolution Authorities during the recent financial crisis. As is noted by Gros and De Groen (2015), many banks needed capital support more than once during the crisis because the initial losses were not accurately estimated or the resolution required more money. Not modeling this kind of renewability makes the derived theoretical value of the compulsory insurance underestimate the corresponding value in practice.

¹⁴Naszodi (2010) develops another option-based model with the same motivation. That model describes the process of an exchange rate managed in a target zone with the help of two options. There the shortcut offered by the option pricing literature is used to derive how the target zone exchange rate depends on the latent exchange rate, i.e., the exchange rate that would prevail under a free float.

of total assets of the bank *at the maturity of the bonds* ($t = T$). By substituting the formula for the intrinsic value of the options into equations (6) and (7),¹⁵ we obtain

$$CDS_T^{value} = \max(K - A_T, 0) \quad (8)$$

$$SRF_T^{value} = \min[\max(K - 0.08L - A_T, 0), 0.05L]. \quad (9)$$

Second, once an assumption is made on the process of the underlying asset, the price of the options can be derived even for $t < T$. However, Occam's razor prevents us from making any assumption on the process before section 4.6.

4. Implications of the Option-Based Model

This section derives eight implications of the option-based model. The first three are *technical implications* about the *values of the insurances*, while the next five are written partially at a *nontechnical* level and cover *normative implications about the fees*:

- (i) Even when the option describing the service of the SRF is out of the money, i.e., when the price of the underlying asset of the corresponding put option exceeds the strike price, $A_t > K - 0.08L$, its value is positive before expiration ($CDS_t^{value} > 0$ for $t < T$).
- (ii) The functional relationship between the value of the compulsory insurance (SRF_t^{value}) and the value of the optional insurance (CDS_t^{value}) is nonlinear.
- (iii) The leverage of the bank determines the exact shape of the above nonlinear function.
- (iv) Even the banks that seem very safe should contribute to the SRF under the Coasian approach, as their benefit from the service provided by the Fund is strictly positive.
- (v) A simple and seemingly tempting analysis approximating the empirical unconditional linear relationship between the contributions to the SRF and the CDS spreads is limitedly informative about whether the service of the SRF is fairly priced in the Coasian sense in practice.

¹⁵See Hull (2012, p. 201) for the definition of the intrinsic value.

- (vi) A modified version of the analysis in (v) above is a better candidate for the same test.
- (vii) If one finds that the fee charged by the SRB and the price paid by the protection buyer of the CDS are not in parity with the values of these insurances and there is political will for putting the Regulation on the ground of Coasian fair pricing,¹⁶ then one possibility for that is to calibrate some of its parameters to the observed CDS spreads by using the option-based model sketched in this paper. In such an exercise, the parameters should be calibrated jointly due to some interdependencies among them. Once the parameters are calibrated, any change affecting only one single parameter could make the pricing deviate from the fair one.
- (viii) The vintage of the data can affect whether calibrating the parameters to the CDS spreads is feasible.

These implications, with the exceptions of implications (i) and (iv), can be obtained by examining the theoretical values of the insurances at maturity ($t = T$). As these values do not depend on the assumed process of the underlying asset (see equations (8) and (9)), the implications are robust to the process.

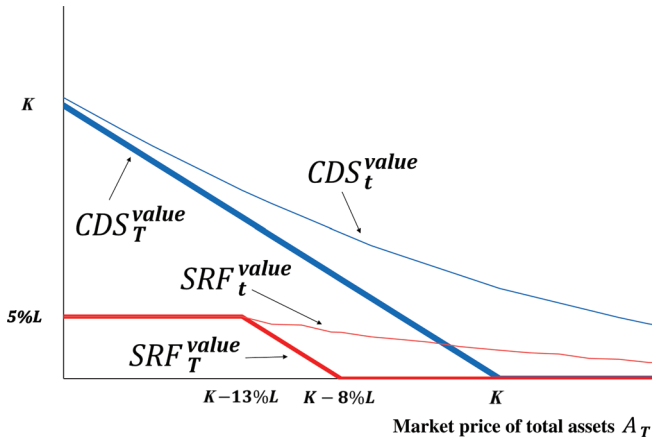
In addition, it is intuitive to say, although not proven here, that the above implications are also robust to whether an assumption of the Merton-type model is relaxed or not. Specifically, even if the secured bank has a more realistic liability structure than consisting of only a zero-coupon bond, all of the eight qualitative implications hold true.

4.1 The Technical Implications of the Option-Based Model

Figure 2 illustrates how the values of the insurances depend on the market price of total assets both at the maturity of the bonds and before. What one can learn from this figure, besides the apparent presence of nonlinearity, is that even when the market price of total assets is high, the value of the insurance provided by the SRF is positive for $t < T$. This phenomenon is a consequence of the style of the

¹⁶As is discussed in the introduction of this paper, the Coasian fair pricing is not the only candidate for being the normative criterion against the Regulation.

Figure 2. The Theoretical Values of the Two Insurances as Functions of the Market Price of Total Assets Both at the Maturity of the Bonds ($t = T$) and Before ($t < T$)



Notes: Both the value of the CDS and the value of the insurance provided by the SRF are in terms of money. The process of the market price of total assets is assumed to be described by the CRR binomial model. Its parameters, namely the volatility of the underlying asset, the time to maturity, the risk-free rate, the yield on returns, and the time steps are set to 20 percent, 1 year, 0, 0, and 30, respectively. Lack of smoothness of the curves representing the values of the insurances before maturity is due to the imprecision of the applied numerical method.

corresponding option.¹⁷ Given that zero does not belong to the set of values of the function assigning the value of an American-style option to the price of the underlying asset for $t < T$, implication (i) is proven.

By inverting function (8) mapping the market price of total assets to the value of the CDS and substituting it to equation (9), we obtain how the value of the insurance provided by the SRF depends on the value of the CDS at the maturity of the bonds ($t = T$):

$$SRF_T^{value} = \min [\max (CDS_T^{value} - 0.08L, 0), 0.05L]. \quad (10)$$

Since the resulting function in equation (10) is nonlinear, implication (ii) is proven.

¹⁷See Hull (2012, p. 215).

To make the theory closer to the empirics, we scale both the value of the CDS and the value of the service of the SRF by dividing both by the face value of bonds (K). In addition, the obtained quantity for the CDS is multiplied by 10^4 . As a result of these transformations, the theoretical value of the CDS is expressed as its price is quoted in practice, i.e., not in terms of money, but as a spread expressed in basis points (bps). Similarly, the value of the insurance provided by the SRF can also be measured as a spread:

$$CDS_t^{value, spread \text{ in bps}} = 10^4 \frac{CDS_t^{value}}{K}, \quad \forall t \leq T \quad (11)$$

$$SRF_t^{value, spread} = \frac{SRF_t^{value}}{K}. \quad \forall t \leq T. \quad (12)$$

By substituting equations (11) and (12) into equation (10), we obtain

$$SRF_T^{value, spread} = \min \left[\max \left(10^{-4} CDS_T^{value, spread \text{ in bps}} - 0.08 \frac{L}{K}, 0 \right), 0.05 \frac{L}{K} \right]. \quad (13)$$

Equation (13) shows that the nonlinearity is preserved by the functional relationship between the values of the insurances after being scaled. In addition, it shows that a specific measure of the *leverage* (i.e., the ratio of the total liabilities to the face value of bonds $\frac{L}{K}$) is an important determinant of the exact shape of this piecewise linear function because it determines where the kinks are. This proves implication (iii).

4.2 *The Normative Criteria and Some Implications of the Model with Direct Policy Relevance*

This section defines formally three concepts: the efficiency of the CDS market, the parity condition, and fair pricing in the Coasian sense. Then, these definitions are used for proving implications (iv) and (v).

4.2.1 Normative Criteria against the Contributions

In order to facilitate the definition of the normative criteria and the efficiency of the CDS market, the assumption of having only one bank in the model is relaxed. Henceforth, it is assumed to have N banks. The yield on returns of the assets, the volatility of total assets, and the leverage are allowed to vary across banks.

The *CDS market is efficient*, if the observed price (in other words, the fee charged for the market-based insurance) is equal to the corresponding theoretical value of the insurance service for each bank:

$$CDS_{t,i}^{fee} = CDS_{t,i}^{value}, \quad \forall t \leq T, \quad \forall i \in \{1, \dots, N\}, \quad (14)$$

where $CDS_{t,i}^{fee}$ is the overall price of the CDS in terms of money providing protection against the default of bank i on its bonds.

The *normative criteria against the fees*, i.e., the *parity condition*, is formalized as

$$\frac{SRF_{t,i}^{fee}}{CDS_{t,i}^{fee}} = \frac{SRF_{t,i}^{value}}{CDS_{t,i}^{value}}, \quad \forall t \leq T, \quad \forall i \in \{1, \dots, N\}, \quad (15)$$

where $SRF_{t,i}^{fee}$ is the contribution in terms of money that bank i pays to the Fund for the availability of the resolution service until the maturity of its debt.

Obviously, the *parity condition* can be written also for the spreads:

$$\frac{SRF_{t,i}^{fee, spread}}{CDS_{t,i}^{fee, spread \text{ in bps}}} = \frac{SRF_{t,i}^{value, spread}}{CDS_{t,i}^{value, spread \text{ in bps}}}, \quad \forall t \leq T, \quad \forall i \in \{1, \dots, N\}, \quad (16)$$

where $SRF_{t,i}^{fee, spread} = \frac{SRF_{t,i}^{fee}}{K_i}$ and $CDS_{t,i}^{fee, spread \text{ in bps}} = 10^4 \frac{CDS_{t,i}^{fee}}{K_i}$.

Under (14) and (15), the fee paid by each bank to the SRF is equal to the value generated by the SRF for the bondholders of the given bank:

$$SRF_{t,i}^{fee} = SRF_{t,i}^{value}, \quad \forall t \leq T, \quad \forall i \in \{1, \dots, N\}. \quad (17)$$

If the externalities increasing the size of the cake are assumed away by $\sum_{i=1}^N SRF_{t,i}^{value} = \sum_{i=1}^N SRF_{t,i}^{fee}$,¹⁸ then the above condition is not stricter than the condition of $SRF_{t,i}^{fee} \leq SRF_{t,i}^{value}$, guaranteeing the distribution of costs to be in the core of the Coasian game. Therefore, we will refer to equation (17) as the criterion for *fair pricing in the Coasian sense*.

4.2.2 The “Zero-Risk Criterion”

Implication (iv) suggests that even the least risky banks should contribute to the SRF under the Coasian approach. This implication is an immediate consequence of implication (i): if the value of the service provided by the compulsory insurance is positive ($SRF_{t,i}^{value} > 0$), then so should be the fee ($SRF_{t,i}^{fee} > 0$) under equation (17). In other words, the “zero-risk criterion” (proposed by the managers of bank A) and the Coasian fair pricing criterion are mutually exclusive.

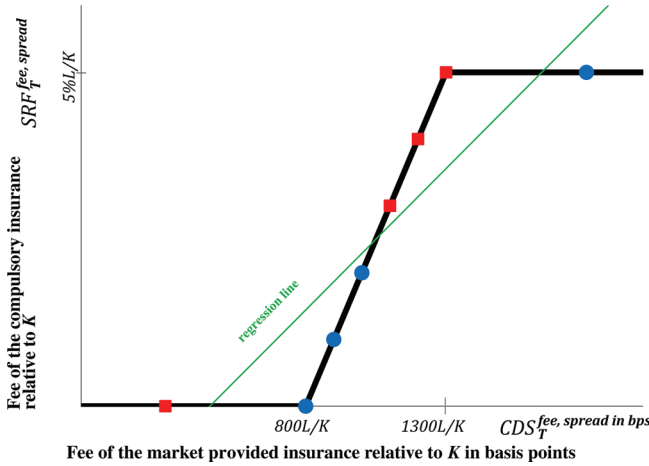
4.2.3 Limitations of Testing the Normative Criteria with the Linear Regression

Let us turn to implication (v) and investigate why, how, and when a two-variable linear regression (proposed by the managers of bank A) can mislead us on whether a bank is overcharged or undercharged by the SRB. First, let us derive how the Coasian fair price for the service provided by the SRF depends on the observed CDS price at the maturity of the bonds ($t = T$). From equations (13), (14), and (17), we obtain

$$SRF_{T,i}^{fee, spread} = \min \left[\max \left(10^{-4} CDS_{T,i}^{fee, spread \text{ in bps}} - 0.08 \frac{L_i}{K_i}, 0 \right), 0.05 \frac{L_i}{K_i} \right]. \quad (18)$$

¹⁸Gros and De Groen (2015) calculate how much funding would have been needed from the SRF during the last banking crisis and find the total amount of about €72 billion, which is more than the target size of the SRF (€55 billion) determined by the Regulation but less than the amount the SRF could draw on, if the ex post levies are also taken into account. Their calculation can be thought of as a joint test on $\sum_{i=1}^N SRF_{t,i}^{value} = \sum_{i=1}^N SRF_{t,i}^{fee}$, since the contingent liabilities of the SRF represent the mirror image of the benefits.

Figure 3. Illustration of the Linear Regression on Data with Nonlinear Relationship

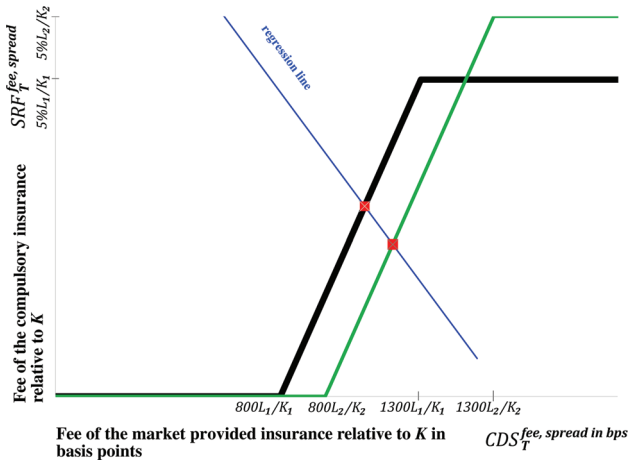


Notes: Each box and each circle correspond to a hypothetical bank. Boxes are above the regression line, while circles are below it. The banks are assumed to distribute the cost of establishing the SRF fairly among themselves, and they are identical in many of their relevant characteristics (leverage $\frac{L_i}{K_i} = \frac{L}{K}$, maturity date of their zero-coupon bonds $T_i = T$ for $\forall i \in \{1, \dots, N\}$); however, their default risks are perceived to be different by the market ($CDS_i^{fee, spread \text{ in bps}} \neq CDS_j^{fee, spread \text{ in bps}}$ for $\forall i, \forall j \in \{1, \dots, N; i \neq j\}$). Evidently, these assumptions guarantee that at the maturity of the bonds ($t = T$) all the bank-level observations (the pair of CDS spreads and the price of the service provided by the SRF) are on the same piecewise linear function representing equation (18) under no variation in the leverage.

Now, let us highlight, by two examples, what the limitations of the two-variable linear regression are at analyzing whether a “cake-cutting” is fair. In both of the examples, the banks in the hypothetical samples distribute the cost of establishing the SRF fairly among themselves. In other words, the bank-level data fulfill equation (18) under the efficiency of the CDS market.

In our first example, the banks operate with the same leverage and there is a nonlinear relationship between their CDS spreads and contributions, as is depicted by figure 3. Now, running a linear regression and investigating the residuals would falsely suggest that banks with moderate market-perceived risk (lower CDS) tend

Figure 4. Illustration of the Two-Variable Linear Regression with Uncontrolled Heterogeneity



Notes: The boxes correspond to two hypothetical banks with different leverage ratios $\frac{L_1}{K_1} \neq \frac{L_2}{K_2}$. Their CDS spreads are not the same either $CDS_1^{fee} \neq CDS_2^{fee}$, while $800 \frac{L_1}{K_1} < CDS_1^{fee, spread \text{ in bps}} < 1300 \frac{L_1}{K_1}$ and $800 \frac{L_2}{K_2} < CDS_2^{fee, spread \text{ in bps}} < 1300 \frac{L_2}{K_2}$. Their zero-coupon bonds have the same maturity date $T = T_1 = T_2$. These banks are assumed to distribute the cost of establishing the SRF fairly between themselves. Evidently, these assumptions guarantee that at the maturity of the bonds ($t = T$) the bank-level observations (the pair of CDS spreads and the price of the service provided by the SRF) are in the middle part of a piecewise linear function representing equation (18). However, each is on a different one due to the difference in their leverage ratios. The thick piecewise linear functions represent equation (18) when the leverage is $\frac{L_1}{K_1}$, while the thin one represents the same equation when the leverage is $\frac{L_2}{K_2}$.

to be undercharged (as these, although not all of them, are typically below the regression line) relative to the banks with high market-perceived risk (as most of the banks with high CDS, although not all, are above the regression line).

In our second example, there are only two banks operating with different leverage ratios. The bank-level observations (the pair of $CDS_{T,i}^{fee, spread \text{ in bps}}$ and $SRF_{T,i}^{fee, spread}$) are depicted by figure 4. This figure illustrates that the simple analysis with linear regression is not adequate in this setup either, due to the omitted-variable bias.

Specifically, if not controlling for the leverage, then the slope of the regression line can even be negative. In other words, the unconditional version of the “monotonicity criterion” does not qualify to be a normative criterion.¹⁹

4.3 *The Extended Option-Based Model and an Alternative Test*

This section first extends the option-based model in order to account for the interaction between the compulsory insurance and the market-based insurance. Then, it proposes a test for the normative criterion of Coasian fair pricing using the extended option-based model. In principle, this test has the potential to address both the misspecification error (i.e., due to working with a linear model in the empirics, while the right model is nonlinear) and the omitted-variable problem (not controlling for the leverage of the banks) discussed already in section 4.2.3.

4.3.1 *The Extended Option-Based Model*

The extended option-based model accounts for the fact that once a credible resolution fund is established, the insurance bought from the market offers only an additional service on top of the compulsory one. If every single euro injected by the resolution fund to the bank decreases the burden on the seller of the CDS,²⁰ then the value of the CDS is

$$CDS_{t,i}^{value, with SRF} = CDS_{t,i}^{value, without SRF} - SRF_{t,i}^{value}, \quad (19)$$

¹⁹Equation (18) suggests that in contrast to the unconditional version of the “monotonicity criterion,” the conditional version of it does qualify to be a normative criterion against the contributions. The conditional version of the “monotonicity criterion” can be defined as follows: among banks that are identical in almost all relevant characteristics (leverage, maturity of the outstanding debt, volatility of the market price of total assets, etc.), those that have higher CDS spreads should contribute more to the Fund.

²⁰The implicit assumption here is that the SRF is used only for covering losses and not for recapitalizing banks. This is realistic if recapitalization is not costly: every euro injected into the capital stock of a failing bank by the SRB pays back once the SRB sells its shares in the resolved bank.

where $CDS_{t,i}^{value, with\ SRF}$ denotes the value of the insurance offered by the market when the SRF already operates. $CDS_{t,i}^{value, without\ SRF}$ would be the value of the optional insurance in the absence of the compulsory one, i.e., when a resolution fund is not even foreseen to start operating until the CDS contract expires. Since the benchmark option-based model disregards the interaction between the insurances by construction, $CDS_{t,i}^{value, without\ SRF}$ in the extended model is identical to the $CDS_{t,i}^{value}$ in the benchmark model.

4.3.2 Testing whether the Contributions Are Fair in the Coasian Sense

This section proposes and performs an empirical test on whether the contributions are fair in the Coasian sense. Formally, the hypothesis to be tested is $\alpha_0 = 0$ and $\alpha_1 = 1$ in the following equation:

$$SRF_{t,i}^{fee} = \alpha_0 + \alpha_1 SRF_{t,i}^{value} \quad \forall t \leq T, \quad \forall i \in \{1, \dots, N\}. \quad (20)$$

Under H_0 , equation (20) is equivalent to equation (17), which is the formal criterion for Coasian fair pricing. The alternative hypothesis is that the cake-cutting is too generous either with the more risky banks ($\alpha_0 > 0$ and $\alpha_1 < 1$) or with the less risky banks ($\alpha_0 < 0$ and $\alpha_1 > 1$) at the expense of the other banks.

Since $SRF_{t,i}^{value}$ is not observable, it is impossible to estimate α_0 and α_1 directly. However, the following approach can be used to circumvent this problem. As a first step, one needs to estimate the CDS spreads of the European banks under the counterfactual that the SRF was not set up by using the hedonic pricing method. We denote the counterfactual CDS spread by $CDS_{t_1,i}^{fee, spread\ without\ SRF}$ for bank i at time t_1 corresponding to a year that is after the SRF was established. It can be approximated by

$$\widehat{CDS}_{t_1,i}^{fee, spread\ without\ SRF} = CDS_{t_0,i}^{fee, spread} + \hat{\delta}_i, \quad (21)$$

where $CDS_{t_0,i}^{fee, spread}$ denotes the observed CDS spread of bank i at time t_0 , the year when the SRF was not even anticipated to be set up. And $\hat{\delta}_i$ is an estimate on the potential change in the CDS spread of bank i , between t_0 and t_1 which is due to all of the factors except

the investigated regulatory change. Inter alia, it captures the effect of the changing risk appetite of the investors between t_0 and t_1 . One option for identifying δ_i is to estimate it from the CDS spreads of banks in a country outside the jurisdiction of the SRB.

If this country is the United Kingdom, then $\hat{\delta}_i = CDS_{t_1,j}^{fee, spread} - CDS_{t_0,j}^{fee, spread}$, where bank j is a U.K. bank which has similar characteristics to bank i and its observed CDS spread is denoted by $CDS_{t_0,j}^{fee, spread}$ and $CDS_{t_1,j}^{fee, spread}$ at times t_0 and t_1 , respectively.

As a second step, the regressions corresponding to equations (22) and (23) need to be run using the estimates on $CDS_{t_1,i}^{fee, spread \text{ without SRF}}$ obtained in the first step:

$$SRF_{t_1,i}^{fee, spread} = \beta_0 + \beta_1 \left(\widehat{CDS}_{t_1,i}^{fee, spread \text{ without SRF}} - 0.08 \frac{L_i}{K_i} \right) + \varepsilon_i, \quad (22)$$

$$SRF_{t_1,i}^{fee, spread} = \gamma_0 + \gamma_1 \left(CDS_{t_1,i}^{fee, spread} - 0.08 \frac{L_i}{K_i} \right) + \omega_i. \quad (23)$$

As is apparent from equations (22) and (23), the leverage of the banks is controlled for by the term $\frac{L_i}{K_i}$, while the misspecification problem may be handled to some extent by restricting the sample to those banks with censoring neither in the dependent variable nor in the independent variable.

It is easy to see that testing H_0 is equivalent to testing \widetilde{H}_0 : $\beta_0 = \frac{\gamma_0}{1+\gamma_1}$ and $\beta_1 = \frac{\gamma_1}{1+\gamma_1}$ under equations (14), (19), and (21).²¹ The proposed test is a *joint test* on whether the contributions are fair in the Coasian sense; the CDSs are priced efficiently; the counterfactual is constructed properly; the size of the cake is unaffected by establishing the SRF, i.e., $\sum_{i=1}^N SRF_{t,i}^{value} = \sum_{i=1}^N SRF_{t,i}^{fee}$; the fund used for covering losses as a fraction of the total liabilities does not vary across banks, i.e., being 5 percent for each secured banks; and the market prices the CDS as an additional insurance after the SRF is established. Therefore, were \widetilde{H}_0 rejected, it would not imply

²¹If heterogeneity among banks is coming only from the differences in their leverages, then the conditional version of the “monotonicity criterion” (defined in footnote 20) is equivalent to $\gamma_1 > 0$. It is easy to see that $\gamma_1 > 0$, together with $\gamma_1 > \beta_1 > 0$, implies $\alpha_1 > 0$, which is a necessary, but not a sufficient, condition for fair pricing in the Coasian sense.

automatically the rejection of H_0 , since it could also be due to the violation of any of the latter five criteria.

Let us illustrate the application of the test on a small sample of banks. For this analysis, we use CDS data of five big U.K. banks (Barclays Bank PLC, HSBC Bank PLC, Lloyds Bank PLC, Royal Bank of Scotland Pl, Standard Chartered Bank), while the EU banks included in the test are those for which not only the CDS data are available but also the contributions paid in 2016. The latter criterion restricts the sample: by browsing the publicly available annual reports of the large banks with CDS quotes, I could collect these data only for five banks.²²

The relevant characteristics of the EU banks studied are summarized by table 1. It shows that none of the banks in the sample had such a high CDS spread that would allow us to work with a sub-sample that has no censoring in the independent variable. Although low CDS spreads are favorable from the point of view of financial stability, they do not provide an optimal setting for the test. Still, the test can be implemented. By doing so, one cannot reject that the contributions are fair in the Coasian sense (see the Wald test in table 2).

Let us close this section by discussing several potential sources of type I errors and type II errors of the test. One source of error is the omission of some important variables, such as the maturity of bonds (T_i) and the parameters describing the process of the market price of total assets (σ_i). Another source of error is this: the way in which the leverage is controlled for in the test is adequate for $t = T$, but might not be perfectly adequate for $t < T$. In addition, the strike price of an option representing the senior CDS of a bank is typically different from the one used in the test, which is the total liabilities reduced by the book value of equity. Finally, the proposed test is limitedly informative about whether the contributions paid by each individual bank are fair, since the test is based on some aggregate statistics. These caveats will be partially addressed in section 4.6.

²²The SRB publishes data on the contributions only at an aggregated level for confidentiality reasons.

Table 1. Contributions Paid by Five Large Banks to the SRF and Other Inputs to the Test

	Levy** Paid in 2016 (€Millions)	Levy Spread Paid in 2016 (bps)	Five-Year Senior CDS Dec. 31, 2015 (bps)	Five-Year Senior CDS Dec. 31, 2012 (bps)	Total Liabilities Dec. 31, 2015 (€Billions)	Liability Excl. Book Value of Equity Dec. 31, 2015	Counterfact. Five-Year Senior CDS*** Dec. 31, 2015 (bps)
BBVA *	137.0	3.80	133.6	285.3	397	360	227.3
BNP Paribas	508.0	2.68	72.2	137.0	1,994	1,894	79.0
Deutsche Bank	280.4	1.80	98.1	95.2	1,629	1,562	37.2
ING	176.0	2.21	54.7	109.2	839	797	51.2
Intesa Sanpaolo	578.0	9.21	95.4	282.0	676	628	224.0

Notes: *BBVA: Banco Bilbao Vizcaya Argentaria. **The bank levies are sourced from the annual reports of the banks. For some banks, the reported levy covers not only the contributions to the SRF but also the irrevocable payment commitments to the deposit guarantee schemes. ***The counterfactual five-year CDS spread at the end of 2015 is calculated from the five-year CDS spread observed at the end of 2012 adjusted by the change in the weighted average CDS spread of five U.K. banks between the end of 2012 and the end of 2015. For details about the CDS data, see notes below figure 6.

Table 2. Illustrating the Steps of the Test Proposed on a Small Sample

	Equation (22)				Equation (23)		Restricted Model	
	Coefficient	Std. Error	γ_0	γ_1	Coefficient	Std. Error	Coefficient	Std. Error
β_0	28*	10	γ_0		-9	67	$\beta_0 - \gamma_0 / (1 + \gamma_1)$	68.07
β_1	0.033*	0.013	γ_1		-0.018	0.089	$\beta_1 - \gamma_1 / (1 + \gamma_1)$	0.09
R^2			R^2				Wald Test	0.68
No. Obs.	0.68	5	No. Obs.		0.01	5	p-value	0.71

Notes: * denotes significance at 10 percent. The small sample size is due to data limitation.

4.4 *Implications of the Model on the Calibration of the Parameters*

This section discusses implications (vii) and (viii) on how some parameters of the Regulation can be calibrated to preserve or achieve Coasian fair pricing.

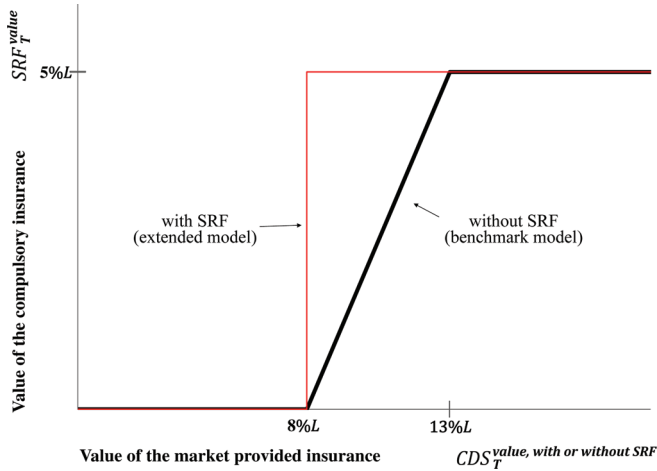
Suppose that the contributions meet the normative criterion of Coasian fair pricing and someone proposes to change one single parameter in the Regulation. For the sake of the thought experiment, suppose that this parameter is the one that determines the maximum extent of intervention by the SRB, which is set to 5 percent of the total liabilities. Increasing this ceiling affects the contingent liabilities of the SRB, making it necessary to adjust the target size of the Fund in order to maintain the credibility of the SRB.

Similarly, such a modification in the ceiling increases both the value of the compulsory insurance generated for the bondholders (SRF_t^{value}) and its maximum, which is $5\%L$ before the hypothetical regulatory change. If the regulator wishes to preserve the Coasian fair pricing captured by equation (17), then the maximum of the fee charged for the compulsory insurance should be modified as well. As is shown in the appendix, the fee charged by the SRB in terms of spread ($SRF_t^{fee, spread}$) is proportional to the so-called rescaled final composite indicator defined by the Regulation.²³ Therefore, a change in the parameter determining the maximum extent of intervention by the SRB should be accompanied by adjusting the cap parameter of the rescaled final composite indicator.²⁴ The above example illustrates that the parameters should be calibrated jointly, as is suggested by implication (vii).

²³See the appendix for the definition of the rescaled final composite indicator. In addition, its equation (A.5) presents how the SRF spread relates to the rescaled final composite indicator.

²⁴The Regulation determines the maximum of the rescaled final composite indicator to be 1.5. See equation (A.2) in the appendix. According to equation (A.2), the minimum of the rescaled final composite indicator is 0.8. It is important to note that figure 3 suggests falsely the minimum to be 0 merely due to some simplifying assumptions of the model in this paper. First, the option-based model disregards that the small and moderately risky banks calculate their contributions with a method different from the “risk-adjusted contribution” method. Second, figure 3 depicts the relationship between the insurances at the maturity of the bonds and not before.

Figure 5. The Value of the Insurance Provided by the SRF as a Function of the Value of the CDS in the Benchmark Model and in the Extended Model at $t = T$



Notes: The function corresponding to the benchmark model is the one in equation (10), while the function corresponding to the extended model can be obtained as follows. First, we substitute equation (8) and (9) into (19) and get $CDS_T^{value, with SRF} = \max(K - A_T, 0) - \min[\max(K - 0.08L - A_T, 0), 0.05L]$. Second, by inverting the above function and substituting it into equation (9), we obtain how the value of the insurance provided by the SRF depends on the value of the CDS at $t = T$ in the extended option-based model.

Finally, let us turn to implication (viii). We use the extended option-based model introduced in section 4.3.1 to prove that the vintage of the data affects the feasibility of calibrating the contributions to the CDS spreads. We assume that the market applies the benchmark model in the pre-SRF era, while it uses the extended model for pricing the CDS after the SRF is already set up. Figure 5 shows that at the maturity of the bonds ($t = T$) the value of the insurance provided by the SRF reacts much more to changes in the value of the CDS in the extended model than in the benchmark model.

The same holds for the fees under Coasian fair pricing and the efficiency of the CDS market. In addition, it is intuitive to say that setting up a resolution fund makes the functional relationship between the fees steeper not only at the maturity of the

bonds ($t = T$) but also before ($t < T$). Obviously, the steeper this function is, the less robust the calibration is. In other words, data from the pre-SRF era, when some banks were considered to be too big to be rescued, can facilitate the calibration of their contributions to CDS spreads. However, once a resolution fund (national or supranational) is expected to cover at least a portion of the losses of some debtholders, the calibrated contributions become sensitive to changes in the CDS spreads and also to their observation errors.

4.5 Structural Break in the CDS Spreads

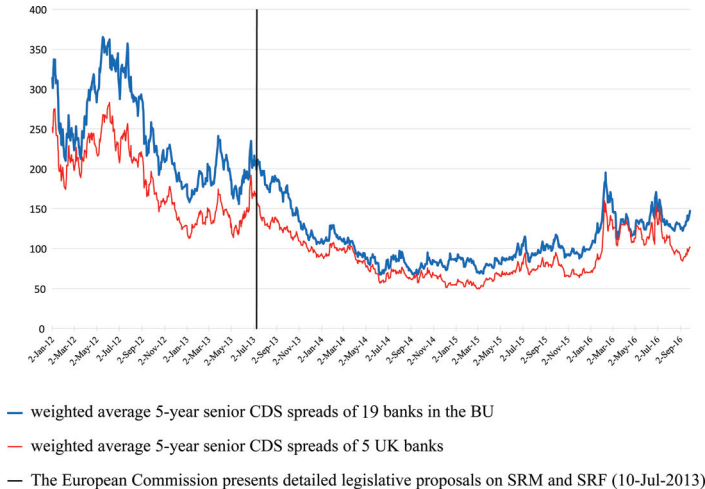
This section identifies a structural break in the CDS spreads of the European banks. The break is indicative of the time when the pre-SRF era has ended.²⁵ By looking at the time series of the CDS spread indicator of some European banks and that of some U.K. banks depicted by figure 6, we can see that the CDS market started to price in the expected change in the regulation around July 10, 2013, when the European Commission presented detailed legislative proposals on the SRM and the SRF.²⁶ Before that date the aggregate CDS spreads seem to have had parallel trends in the United Kingdom and in the Banking Union. After July 2013, the difference between the spreads started to shrink.

Based on this simple analysis, we conclude that if one would like to test the contributions of the banks to the SRF against the criterion of fair pricing in the Coasian sense by using CDS data, then the CDS spreads from the period preceding July 2013 are preferable to be used for this purpose. In subsection 4.3.2, we followed this practice when constructing the CDS spread under the counterfactual by using CDS data from the end of 2012.

²⁵Naszodi and Katay (2020) provide a more detailed analysis of the time series of the CDS spreads of some European banks, with the purpose of quantifying to what extent the SRM has enhanced financial stability in the Banking Union.

²⁶See <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52013PC0520&qid=1499762308645&from=EN>.

Figure 6. An Important Milestone Towards the Single Resolution Mechanism and the Weighted Average CDS Spreads of Some Large Banks in the Banking Union and in the United Kingdom between January 2, 2012 and September 27, 2016



Source: Naszodi and Katay (2020), who used CDS data from CMA Datavision for the period preceding the year 2014, and CDS data from Bloomberg for the period afterwards.

Notes: The weights represent the relative size of the bank in the Banking Union or in the United Kingdom in terms of total assets. The 19 banks in the Banking Union are AXA Bank Europe SA/NV, Banca Monte dei Paschi di Siena SpA, Banca Nazionale del Lavoro SpA, Banco Bilbao Vizcaya Argentaria SA, Banco Comercial Português SA, Banco Popolare, Banco Santander SA, BNP Paribas SA, BNP Paribas Fortis SA/NV, Commerzbank AG, Crédit Agricole S.A., Deutsche Bank AG, ING Bank NV, Intesa Sanpaolo, Mediobanca SpA, Novo Banco SA, Rabobank Nederland, Société Générale, and UniCredit SpA. The five U.K. banks are Barclays Bank PLC, HSBC Bank PLC, Lloyds Bank PLC, Royal Bank of Scotland Plc, and Standard Chartered Bank.

4.6 Calculating the Coasian Fair Contribution of a Hypothetical Bank from its Actuarial Spread

To illustrate further how the benchmark model can be used in practice, this section presents the calculation of the Coasian fair contribution of a hypothetical bank. In this example, the *leverage of the hypothetical bank* and the *maturity of its bonds* are taken into

account, and an assumption is made on the *process of the market price of its total assets*. The focus is on one bank since the concept of fairness is applicable to *individual entities* rather than to groups. Furthermore, its contribution is calculated from the market capitalization of the bank and its annualized *actuarial spread (AS)* estimated and published by the Credit Research Initiative (CRI).

Using the actuarial spreads from the CRI has some advantages relative to the market-based CDS spreads. First, the actuarial spread captures the solvency risk in relation to the failure of the bank. At the same time, it is free from various premiums, i.e., the premiums compensating for the illiquidity of the CDS market and the bond market, and the premiums capturing the market power of protection sellers and their solvency risk. Second, it is plausible that the contribution of a secured bank determined under the Coasian perspective is linked to the solvency risk of the given bank, but it is not plausible to be linked to the premiums listed above. Third, the actuarial spread does not account for any kind of public intervention explicitly.

Let us turn now to the numerical example. In this example, the process of the bank's market price of total assets is assumed to be described by the CRR binomial model.²⁷ Its two parameters, the market price of total assets (A_t) and its volatility (σ), are chosen so as to fulfill equations (24) and (25).

²⁷This simple process facilitates the valuation of even exotic options by a numerical method (by backward induction) under the no-arbitrage condition. Our approach for calculating the contributions can be further refined along the works by Merton (1976), Duan, Sun, and Wang (2012), Duan and Fulop (2013), Duan (2014), and Du, Elkamhi, and Ericsson (2019). These papers model the default probability not only with the leverage of the bank and the volatility of its market price of total assets but also with at least six factors neglected by this paper. These are the jumps in the stock price, correlation among default probabilities of different banks, the defaults over multiple horizons, the changes in the interest rate, the risk aversion of investors, and stochastic asset volatility. Since the above models have richer structures than the one in this paper, those are likely to perform better when the basis of comparison is the in-sample fit, but it is not necessarily the case for the out-of-sample fit. For instance, Hull, Nelken, and White (2005, p. 22) find that the more complex Merton (1976) "model has statistically significant explanatory power, but in all cases the Merton (1974) model provides significantly better predictions of default probabilities and credit spreads at the 1% level."

$$AS_t^{fee, \text{ spread in bps}} = \frac{10^4}{K(T-t)} Put_t^{European} (A_t, K_d = \text{€1,374 billion}, \\ T-t = 0.5 \text{ yr}, \sigma, r = 0, q = 0.4\%), \quad (24)$$

$$Equity_t = Call_t^{European} (A_t, K_e = \text{€1,561 billion}, \\ T-t = 0.5 \text{ yr}, \sigma, r = 0, q = 0.4\%), \quad (25)$$

where the choice of the values for the strike prices (K_d, K_e), the time until maturity ($T-t$), and the yield on the underlying asset (q) are motivated by the corresponding characteristics of Deutsche Bank AG at the end of 2015.²⁸ Similarly, the left-hand side of equation (24) is set equal to the actuarial one-year spread (in bps) of Deutsche Bank AG on December 12, 2015,²⁹ i.e., $AS_t^{fee, \text{ annualized spread in bps}} = 29.86$, while the left-hand side of equation (25) is set equal to its market capitalization $Equity_t = \text{€31.07 billion}$.

As a first step, we solve the above system of equations and obtain $A_t = \text{€1,540 billion}$ and $\sigma = 9.63$ percent. As a second step, the above parameters are used to calculate the Coasian fair price for the service provided by the SRF to the hypothetical bank. This calculation involves the pricing of an American-style put option with the numerical binomial option valuation method along the lines of equations (7) and (16). We obtain the Coasian fair price for the annual coverage $\frac{SRF_t^{fee}}{T-t}$ to be around €33 million.

²⁸All the bank-specific information used in this exercise is publicly available and comes from the Deutsche Bank's Annual Report 2015. The yield of the underlying asset (q) is set equal to the return on assets (ROA). The strike price K_d is set to be equal to the book value of total liabilities reduced by the sum of liabilities that are less senior than the long-term debt guaranteed by the CDS. These liabilities include the total equity, other liabilities, other financial liabilities, and trust preferred securities, while K_e equals the total liability reduced by the sum of the total equity. The time until maturity ($T-t$) is set to be the weighted average of the midpoints of each maturity intervals reported. (The midpoint is replaced by five years in the case of the maturity category of "over five years.") When calculating the duration of the liabilities, covered deposits (that are rarely withdrawn at their expiration date) are not differentiated from other types of liabilities, as the zero-coupon bonds are assumed to be the only type of external liabilities in the model.

²⁹Source of data: National University of Singapore, Risk Management Institute, CRI database. Available at <https://nuscri.org> (accessed on February 20, 2019).

How would some of the omitted factors modify the above figure? As is shown next, our method offers a lower bound for the fair contribution for three reasons. First, the calculation above does not account for the fact that coverage is potentially provided not only during the eight years of contribution period but also beyond. In order to count with it, the above figure should be scaled up by approximately $\frac{8+h}{8}$, where h denotes the assumed number of additional years for which the resolution service is offered without any further contributions for the given bank.

Second, as is discussed in section 4.3.1, it does matter whether the interaction between the insurances are taken into account or not. Although the actuarial spread calculated and published by the CRI does not account for any kind of public intervention explicitly, it might not be perfectly immune to the changes in the bank regulations. For instance, it might factor in the enhanced financial stability due to the established SRB via being calculated from higher recovery rates. In the latter case, one might underestimate the fair contributions with the method presented above.

Third, while certain premiums present in the CDS spreads might be considered inadequate to be built in the contributions payable to an ex ante resolution fund, some others might qualify to be charged for despite the fact that the actuarial spread does not capture them. The types of premiums falling into the latter category are those that compensate for the risk of illiquidity of the secured bank,³⁰ and the systemic risk generated by the secured bank that affects the same bank.³¹ If the latter two premiums are negligible in magnitude, then it is adequate to calibrate the contributions to the actuarial spread capturing only the solvency risk. However, if these premiums are

³⁰If banks become more resilient against liquidity shocks by relying on more stable funding and decreasing their maturity mismatch, then the bondholders can be compensated with lower risk premium. How this affects the liquidity premium charged on the CDS market is not evident.

³¹The CDS spread of a given bank captures that slice of the systemic risk that is due to shocks generated by the bank in question and affects the very same bank via the feedback from the other banks to the initiator of the shock. By decomposing the CDS spreads of banks, Keiler and Eder (2013) find the relative weight of the systemic risk component to be around 10 percent. As a matter of fact, the third risk pillar (called “importance of an institution to the stability of the financial system or economy”) used in the Regulation for determining the risk profile of the banks is assigned also the weight of 10 percent.

large, then the CDS spread might serve to be a better reference. What period the CDS spread is sampled from for such an exercise might not be neutral according to the theoretical finding of section 4.3.1.

5. Conclusions

This paper presented a Merton-type model describing both the value of the compulsory insurance provided by the Single Resolution Fund and the value of the optional insurance provided by the CDS market as put options. This model offers a framework for testing whether the contributions of banks to the Fund satisfy a normative criterion. The normative criterion analyzed in this paper is the Coasian fair pricing under which the bondholders of each bank benefit from the existence of a resolution fund at least as much as their banks contribute to the Fund.

The option-based model and the concept of fair pricing can help the regulator decide what proposals on the changes in the Regulation determining the contributions are worth considering. If there is political will for changing some parameters in the Regulation, then one possibility is to calibrate the parameters either to the CDS spreads or the actuarial spreads, for instance, by using a concept of fair pricing and a model similar to the one sketched in this paper.

This paper presented some advantages and some potential limitations of such a calibration. It highlighted that in such an exercise, the parameters should be estimated jointly. When those are calibrated to the CDS spreads, then it is advised to use data from the pre-SRF era, prior to July 2013. Once the parameters are calibrated, any change affecting only one single parameter could make the pricing deviate from the fair one. Whether these implications are robust to the assumed normative theory forming the basis of the Regulation is the subject of future research.

Appendix

This appendix presents some formulas of the *risk-adjusted method* in the Regulation. Then, it derives how the *SRF spread* relates to the *rescaled final composite indicator* defined by the Regulation.

The risk-adjusted method is applicable to big and/or risky banks to calculate their contributions to the SRF. The formulas for computing the annual contributions defined by annex 1, step 6, paragraphs 1 and 2 of the Regulation are

$$c_i = Target \frac{\frac{B_i}{\sum_j B_j} \tilde{R}_i}{\sum_j \left(\frac{B_j}{\sum_m B_m} \tilde{R}_j \right)}, \quad (\text{A.1})$$

$$\tilde{R}_i = (1.5 - 0.8) \frac{FCI_i - \min FCI_j}{\max FCI_j - \min FCI_j} + 0.8, \quad (\text{A.2})$$

where i , j , and m index financial institutions. The annual contribution in terms of money payable by bank i is denoted by c_i . The rescaled final composite indicator of bank i is \tilde{R}_i . *Target* is the annual target level of the total contributions collected from those banks that calculate their individual contributions with the risk-adjusted method (and not with any of its alternatives, i.e., the partial risk-adjusted method or simply contributing by a lump sum). B_i is the amount of liabilities (excluding own funds) less covered deposits of institution i . Finally, FCI_i is the *final composite indicator* to be calculated from a number of components. See the Regulation for further details.

What is the correspondence between these notations in the Regulation and the notations in this paper? First, if liabilities consist only of zero-coupon bonds and own funds, then

$$B_i = K_i. \quad (\text{A.3})$$

Second, both c_i and SRF_i^{fee} denote certain kinds of contributions payable by bank i . Their functional relationship can be obtained after transforming both to annuities:

$$c_i \frac{8}{8+h} = \frac{SRF_i^{fee}}{T_i - t}, \quad (\text{A.4})$$

where the annual contribution c_i (used in the Regulation) should be paid only during the eight-year transition period which is assumed to be followed by h years of contribution holidays. We can think of this period of h years as first having $h - 1$ years of tranquility that

is followed by 1 year of severe bank crisis consuming all the Fund. For instance, if severe bank crises are believed to take place every 70 years and the 8-year transition period is free of crises, then h should be 62.

Finally, using the above formulas in the appendix and the terminology in the Regulation, one can give a new interpretation to the annualized SRF spread:

$$\begin{aligned} \frac{SRF_{t,i}^{fee, spread}}{T_i - t} &= \frac{SRF_{t,i}^{fee}}{K_i} \frac{1}{T_i - t} = \frac{c_i}{K_i} \frac{8}{8+h} \\ &= \frac{8}{8+h} Target \frac{\frac{1}{\sum_{j=1}^N K_j}}{\sum_{j=1}^N \left(\frac{K_j}{\sum_{m=1}^N K_m} \tilde{R}_{t,j} \right)} \tilde{R}_{t,i} = M_t \tilde{R}_{t,i}, \quad (A.5) \end{aligned}$$

where $\tilde{R}_{t,i}$ is the time- and bank-specific *rescaled final composite indicator* defined by equation (A.2), while the *multiplier* M_t is the

same across all banks: $M_t = \frac{8}{8+h} Target \frac{\frac{1}{\sum_{j=1}^N K_j}}{\sum_{j=1}^N \left(\frac{K_j}{\sum_{m=1}^N K_m} \tilde{R}_{t,j} \right)}$.

References

- Brunnermeier, M., A. Crockett, C. Goodhart, A. Persaud, and H. Shin. 2009. "The Fundamental Principles of Financial Regulation." Geneva Reports on the World Economy 11.
- Coase, R. H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3 (October): 1–44.
- Cox, J. C., S. A. Ross, and M. Rubinstein. 1979. "Option Pricing: A Simplified Approach." *Journal of Financial Economics* 7 (3): 229–63.
- Du, D., R. Elkamhi, and J. Ericsson. 2019. "Time-Varying Asset Volatility and the Credit Spread Puzzle." *Journal of Finance* 74 (4): 1841–85.
- Duan, J.-C. 2014. "Actuarial Par Spread and Empirical Pricing of CDS by Decomposition." *Global Credit Review* 4 (1): 51–65.
- Duan, J.-C., and A. Fulop. 2013. "Multiperiod Corporate Default Prediction with the Partially-Conditioned Forward Intensity." RMI Working Paper.

- Duan, J.-C., J. Sun, and T. Wang. 2012. "Multiperiod Corporate Default Prediction—A Forward Intensity Approach." *Journal of Econometrics* 170 (1): 191–209.
- Freixas, X. 2003. "Crisis Management in Europe." In *Financial Supervision in Europe*, ed. J. J. M. Kremers, D. Schoenmaker, and P. Wierds, 102–19. Cheltenham: Edward Elgar.
- Goodhart, C., and D. Schoenmaker. 2009. "Fiscal Burden Sharing in Cross-Border Banking Crises." *International Journal of Central Banking* 5 (1, March): 141–65.
- Gros, D., and W. P. De Groen. 2015. "Estimating the Bridge Financing Needs of the Single Resolution Fund: How Expensive Is It to Resolve a Bank?" CEPS Special Report No. 122, Centre for European Policy Studies.
- Huang, J.-Z., Z. Shi, and H. Zhou. 2020. "Specification Analysis of Structural Credit Risk Models." *Review of Finance* 24 (1): 45–98.
- Hull, J. C. 2012. *Options, Futures, and Other Derivatives*. 8th Edition. Upper Saddle River, NJ: Pearson Prentice Hall.
- Hull, J. C., I. Nelken, and A. White. 2005. "Merton's Model, Credit Risk and Volatility Skews." *Journal of Credit Risk* 1 (1): 3–27.
- Keiler, S., and A. Eder. 2013. "CDS Spreads and Systemic Risk: A Spatial Econometric Approach." Discussion Paper No. 01/2013, Deutsche Bundesbank.
- Merton, R. C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29 (2): 449–70.
- . 1970. "A Dynamic General Equilibrium Model of the Asset Market and Its Application to the Pricing of the Capital Structure of the Firm." Working Paper No. 497-70, Sloan School of Management, Massachusetts Institute of Technology (December).
- . 1976. "Option Pricing when Underlying Stock Returns Are Discontinuous." *Journal of Financial Economics* 3 (1–2): 125–44.
- Naszodi, A. 2010. "The Options-based Model of the Target Zone and Its Application to the French Franc." *Journal of Derivatives and Hedge Funds* 16 (3): 212–18.
- Naszodi, A., and G. Katay. 2020. "Has the Single Resolution Mechanism Enhanced Financial Stability in Europe? Evidence from CDS Data." Unpublished Manuscript.
- Necula, C., and A.-N. Radu. 2012. "Quantifying the Recapitalization Fund Premium Using Option Pricing Techniques." *Economics Letters* 114 (3): 249–51.

- Ronn, E. I., and A. K. Verma. 1986. "Pricing Risk-adjusted Deposit Insurance: An Option-based Model." *Journal of Finance* 41 (4): 871–95.
- Schoenmaker, D. 2010. "Do We Need a Separate Resolution Fund?" VoxEU, January 14. <http://voxeu.org/article/against-separate-resolution-fund>.
- Tirole, J. 2010. "From Pigou to Extended Liability: On the Optimal Taxation of Externalities Under Imperfect Financial Markets." *Review of Economic Studies* 77 (2): 697–729.

Monetary Policy Transmission via Loan Contract Terms in the United States*

Esteban Argudo
Vassar College

I study monetary transmission via changes in contract terms for C&I loans. I find that nonprice terms tighten and price terms relax following a surprise monetary contraction, consistent with a decrease in loan supply. Adjustments in nonprice terms (maximum line size, covenants, and collateral requirements) are responsible for a statistically significant decrease in GDP of about 0.3 percentage point following a monetary surprise. I also document a lag between the response in bond market credit indicators and the loan contract terms. I interpret this finding as evidence of an important interaction between these two markets.

JEL Codes: E43, E44, E51, E52.

1. Introduction

The study of monetary policy is based on the premise that central banks can influence economic activity. One possible transmission mechanism involves the economy's credit conditions. Adjustments in monetary policy tools lead to changes in credit conditions, which in turn have consequences for aggregate borrowing, consumption, investment, and output. Why do adjustments in the monetary policy tools affect the credit conditions? Which credit conditions are relevant?

*This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. I thank the editor and the anonymous referees for their constructive feedback. I am also grateful to Bulent Guler, Juan Carlos Hatchondo, Eric Leeper, Amanda Michaud, and the participants at the 2017 Indiana University Macroeconomics Brownbag Seminar and Macroeconomics Workshop, the 2018 Midwest Economics Association Conference, and the LACEA-LAMES 2018 Conference for their helpful comments and suggestions. Author e-mail: eargudo@vassar.edu.

This paper focuses on the second question. I investigate the transmission of monetary policy shocks via changes in price and nonprice contract terms for commercial and industrial (C&I) loans. My motivation is simple; the literature often assumes that changes in credit conditions due to monetary surprises are primarily captured by changes in the risk-free and spread components of interest rates. Although this assumption might be justified when considering bond contracts, it does not obviously follow for loan contracts given that they are intrinsically higher-dimensional objects. For instance, C&I loan contracts often include collateral requirements, covenants, and a maximum line size. Therefore, it is entirely plausible that adjustments in these nonprice terms are relevant for monetary transmission.

One might argue that thinking about transmission via these nonprice terms is unnecessary: after all, most theoretical models conclude that monetary transmission via credit conditions can be thought of “as if” it was captured solely by price-based mechanisms. However, it is important to remember that several of the micro-founded theories that result in price-based transmission mechanisms reflect, at their core, adjustments in nonprice credit terms. For example, the idea of monetary policy transmission via spreads over risk-free rates relies on the seminal contributions of Kiyotaki and Moore (1997) and Bernanke, Gertler, and Gilchrist (1999). Both of these papers use collateral requirements as the main modeling device to capture the effect of credit market frictions.¹ Thus, even if we might think of monetary transmission “as if” it was entirely captured by price-based mechanisms, it is important to provide empirical evidence that supports the underlying assumptions of our theoretical models.

Studying transmission via nonprice loan contract terms is challenging; it is much easier to obtain data on interest rates than on nonprice terms. I overcome this issue using data from the Senior Loan Officer Opinion Survey (SLOOS), which contains information about loan demand and adjustments in several different loan

¹In their introduction, Bernanke, Gertler, and Gilchrist (1999) state that one of the reasons for incorporating credit market effects into their model is the empirical finding (from the household consumption literature) about the importance of credit limits on borrowing.

contract terms. The SLOOS asks a subset of U.S. banks if they have faced stronger than usual loan demand, if they tightened their requirements for approving loan applications, and which specific loan contract terms they adjusted on those loans they were willing to approve.² There are three types of contract terms for C&I loans captured by the SLOOS: (i) price terms (cost of credit line and interest rate spread), (ii) *extensive* margin nonprice terms (covenants and collateral requirements), and (iii) *intensive* margin nonprice terms (maximum line size). I validate the SLOOS data using several measures of lending volume from the Survey of Terms of Business Lending (E.2). I show that, after accounting for a common component, the SLOOS price and nonprice terms do indeed reflect adjustments in the margins to which they allude. In particular, the standards and nonprice terms reflect adjustments *other* than changes in interest rates.

My empirical setup is based on quarterly vector autoregressions (VARs) that include the one-year U.S. Treasury yield, real gross domestic product (GDP), the consumer price index, the excess bond premium (as a control for the overall credit conditions), the SLOOS demand for C&I loans, and (one-by-one) the SLOOS C&I loan contract terms. I use the external instrument approach proposed by Gertler and Karadi (2015) to identify monetary policy shocks. This approach allows me to recover the vector \mathbf{s}^p that collects the contemporaneous change in each VAR variable following a monetary policy shock at time t . Given that the identification procedure does not impose any a priori restrictions on the interaction between the different VAR variables, s_x^p captures the contemporaneous monetary policy transmission via variable x . The nature of the VAR implies s_x^p can cause a change in any of the other VAR variables y at any future date $\tau \geq t$. That is, I can compute the transmission of a monetary policy shock to variable y at time τ via variable x . This allows me to document and quantify the

²The subset of banks is carefully selected in accordance with the purpose of the survey, which is “to provide qualitative and limited quantitative information on bank credit availability and loan demand, as well as on evolving developments and lending practices in the U.S. loan markets.” A complete description of the reporting panel can be found in the Supporting Statement for the Senior Loan Officer Opinion Survey on Bank Lending Practices.

contribution of the price and nonprice contract terms to monetary policy transmission.

My results suggest that a surprise monetary contraction leads to a decrease in loan supply. Following a monetary contraction, the nonprice terms *tighten* while the loan interest rate spreads *relax*. The decrease in loan spreads can be rationalized by noting that loan rates do not adjust as fast as (government) bond market rates. Nonprice terms are responsible for a statistically significant decrease in GDP of about 0.3 percentage point following a monetary surprise. Although the contribution of adjustments in collateral requirements accounts for most of the decrease in GDP, changes in nonprice terms are not individually, but *collectively*, relevant for monetary transmission.

My study also sheds light on the interaction between the loan and bond markets for monetary transmission. I find that the adjustments in loan contract terms happen immediately upon the monetary surprise, while the increase in the excess bond premium happens with a lag. The lagged increase in the excess bond premium suggests that firms turn to the (corporate) bond market to raise funds after they are unable to get funds from banks due to the tightened lending conditions. Note that this interaction between the loan and bond markets is absent in the standard mechanisms à la Kiyotaki and Moore (1997) and à la Bernanke, Gertler, and Gilchrist (1999) commonly used for modeling financial frictions.

My results are robust to different subsamples, number of lags, and proxies for overall credit conditions. The results actually become quantitatively and statistically more significant when only the pre-crisis period is considered (i.e., there is a larger effect of nonprice terms on GDP). Decreasing the number of lags from four to two or dropping the loan demand controls helps attenuate overfitting concerns. However, in both cases adjustments in lending standards lose some statistical significance and changes in collateral requirements become relatively more important within the nonprice terms. The former might be a mechanical consequence of just having fewer regressors. The latter suggests that adjustments in collateral requirements are more persistent than adjustments in the other margins and are more strongly associated with changes in loan demand. Including other credit spreads instead of the excess bond premium does alter the response of GDP and other macroeconomic variables to a

monetary surprise.³ However, it doesn't alter the results about the contribution of the nonprice loan terms to the monetary transmission mechanism.

My results provide empirical support for modeling financial frictions à la Kiyotaki and Moore (1997) and à la Bernanke, Gertler, and Gilchrist (1999). Furthermore, they also uncover two avenues that could be useful to resolve the critique that these types of financial frictions, although qualitatively attractive, are quantitatively unimportant. The first one is considering other nonprice margins of adjustments in addition to collateral requirements. The second one is explicitly modeling the interaction between the corporate bond and loan markets.

The rest of the paper is organized as follows. Section 2 discusses the related literature. Section 3 presents the data I use for my empirical study and validates the SLOOS data. Section 4 presents the empirical framework. Section 5 discusses the main results of the paper and performs several robustness checks. Section 6 concludes. Online appendixes (available at <http://www.ijcb.org>) contain details about the methodology and robustness checks.

2. Related Literature

There is an extensive list of empirical studies that focus on the effect of monetary policy on credit conditions. My work is most closely related to those that analyze the impact of monetary policy on banks' lending standards (willingness to give loans). Some recent examples include Maddaloni and Peydró (2011, 2013), Jiménez et al. (2012), and Ciccarelli, Maddaloni, and Peydró (2015). The first two studies analyze the impact of short- and long-term rates on lending standards within the context of securitization, bank supervision, and macroprudential policy. The third one investigates if short-term rates have a different impact on the probability of loans being granted depending on the strength of a bank's balance sheet. The last one isolates the effect of monetary surprises on loan supply

³This is not surprising given that there are several other studies that document that the excess bond premium contains relevant additional information not reflected by most other credit indicators. Thus excluding it from the VAR specifications can result in omitted-variables bias.

and tries to identify the underlying factors leading to such changes (i.e., separately identifying the bank lending, balance sheet, and cost of credit channels). Clearly, the goal of all of these studies is to assess the extent to which monetary surprises lead to changes in loan supply and determine which factors might be responsible for such changes. Unlike my study, however, none of them focus on *which* credit conditions (loan contract terms) adjust.

There are several other studies with similar identification strategies. For instance, Kuttner (2001), Faust, Swanson, and Wright (2004), Gürkaynak, Sack, and Swanson (2005), Hamilton (2009), Barakchian and Crowe (2013), and (of course) Gertler and Karadi (2015) all use some variant of a high-frequency identification procedure (and a few of these studies use it within the context of VARs.) However, these papers either focus on introducing a new identification scheme or on applying an existing one to study monetary transmission via *prices* (bond and asset prices, interest rates, term premiums, credit spreads). Additionally, my study expands the scope of the methodology proposed by Gertler and Karadi (2015) by showing that it can be used to quantify the contribution of different variables to monetary transmission.

There are also a few studies that resemble mine in that they use the SLOOS (or similar survey data). For instance, Lown and Morgan (2006) use the SLOOS data to analyze the predictive power of lending standards for U.S. GDP. Bassett et al. (2014) construct a new credit supply indicator using the SLOOS data and then study the effect of credit supply shocks on output, borrowing, bond credit spreads, and monetary policy. The goal of these studies is to validate the SLOOS data and show that it contains useful information about the U.S. credit conditions, rather than to assess the impact and transmission of monetary surprises via credit conditions.

3. Data

3.1 Data Description

I use quarterly macroeconomic and credit data from 1990:Q1 to 2016:Q3. The macroeconomic data include real GDP (Y), the one-year U.S. Treasury yield (1YR), and the consumer price index (P). All three macroeconomic variables are obtained from the Federal

Reserve Economic Database (FRED). Real GDP and the consumer price index are logged.⁴

I use the excess bond premium (EBP) from Gilchrist and Zakrajšek (2012) as a proxy for the overall credit conditions in the economy. Their original EBP monthly series extends only through 2012:M6. Simon Gilchrist has an updated EBP monthly series which extends through 2016:M8.⁵ I construct my EBP series by taking the quarterly average of the updated EBP.

I focus only on commercial and industrial loans.⁶ The data come from the SLOOS and the Survey of Terms of Business Lending (E.2), which are provided by the Federal Reserve Board (FRB). I use the E.2 release to obtain data on interest rates and different measures of lending volume, mostly for the external validation of the SLOOS variables.

I use the SLOOS variables to proxy for changes in loan demand and changes in price and nonprice terms of loan contracts. The SLOOS data reported at any given quarter pertain to the demand, lending terms, and standards for the *previous* quarter. In other words, one must lag the SLOOS data by one quarter to align it with the other macroeconomic and credit data series. The SLOOS question pertaining to loan demand asks banks if they have seen a change in loan demand after accounting for normal seasonal variation. The SLOOS loan contract data consist of two different types of questions: those that ask about changes in lending “standards” and those that ask about changes in lending “terms.” The survey questions related to the “standards” ask if banks tightened their requirements for approving loan applications, while the questions related to the “terms” ask about the specific contract conditions that banks adjusted on those loans they were willing to approve.

⁴The sample period is selected purely for reasons of data availability. The SLOOS credit conditions are available starting in 1990:Q1, the external instruments are available only through 2016:Q4, and the excess bond premium is available only through 2016:Q3. The main results of the paper are robust to using the detrended (HP-filtered) version of these variables.

⁵See <https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/updates-the-recession-risk-and-the-excess-bond-premium-20161006.html>.

⁶An earlier version of the paper included the households’ credit card market in the analysis. However, I decided not to include it in the present version given that I found that adjustments in credit card terms are irrelevant for monetary transmission.

Table 1. SLOOS Data

Commercial and Industrial (C&I) Loans		
Variable Name	Type	Availability Period
Demand (C&I)	N/A	1991:Q3–2016:Q3
Standards (C&I)	Nonprice	1990:Q1–2016:Q3
Spread (C&I)	Price	1990:Q1–2016:Q3
Cost of Line (C&I)	Price	1990:Q2–2016:Q3
Loan Covenants (C&I)	Nonprice	1990:Q1–2016:Q3
Maximum Line Size (C&I)	Nonprice	1990:Q1–2016:Q3
Collateral Requirement (C&I)	Nonprice	1990:Q1–2016:Q3
Source: SLOOS.		

Although the distinction between the two types of questions is conceptually clear, it becomes less evident in practice. For instance, one of the questions related to the “terms” asks banks if they tightened their collateral requirements. One might argue that collateral requirements are part of the “standards” banks use for approving loan applications. Therefore, I include variables that reflect both types of questions.

Table 1 lists the SLOOS variables that are relevant for my study. The demand variable is constructed as the net percent of U.S. domestic banks that reported a stronger loan demand. Each of the loan contract variables is constructed as the net percent of U.S. domestic banks that “tightened” the specified margin (standards, spreads, covenants, collateral requirements, etc.) within a given quarter. As it can be seen from the table, I can distinguish between adjustment in price and nonprice loan contract terms using these variables.⁷

3.2 External Validation of the SLOOS Data

Several studies have established the validity of the SLOOS demand and lending “standards.” Lown, Morgan, and Rohatgi (2000) find

⁷There are other SLOOS variables pertaining to C&I lending that I don’t use in my study due to their limited sample size. Among those variables are the risk premium (available from 1998:Q3 onwards) and the maximum maturity (available from 2005:Q2 onwards).

Table 2. Contemporaneous Correlation between SLOOS Variables

	Standards	Spread	Cost of Line	Covenants	Line Size	Collateral
Standards	1.00	0.90	0.92	0.95	0.92	0.92
Spread	-0.66	1.00	0.98	0.93	0.93	0.89
Cost of Line	-0.36	0.13	1.00	0.95	0.92	0.92
Covenants	0.29	-0.58	-0.50	1.00	0.94	0.96
Line Size	0.19	-0.47	-0.43	0.19	1.00	0.92
Collateral	0.24	-0.57	-0.38	0.47	0.22	1.00

Source: Author’s computation using data from the SLOOS.
Notes: Elements over the main diagonal correspond to the correlation between the raw variables. Elements under the main diagonal correspond to the correlation when the common component is removed.

that the tightening of C&I “standards” is strongly negatively correlated with aggregate commercial loan growth and with various measures of economic activity. Lown and Morgan (2006) find that the C&I “standards” dominate loan rates in explaining variation in business loans and output. Bassett et al. (2014) construct a new credit supply indicator using the lending “standards” for C&I and consumer loans (adjusted for macroeconomic and bank-specific factors) and show that this indicator can substantially explain changes in output. Ciccarelli, Maddaloni, and Peydró (2015) use the SLOOS demand and lending “standards” for C&I loans to identify different channels of monetary policy transmission. However, not much has been said regarding the SLOOS “terms.” The purpose of this section is to show that the SLOOS “terms” do indeed convey relevant information about changes in credit conditions.

Table 2 presents the correlation coefficient between different SLOOS variables related to C&I lending. The elements above the main diagonal refer to the correlation between the raw SLOOS variables. The elements below the main diagonal (in bold) correspond to correlation between the transformed SLOOS variables (I discuss such transformation shortly). All C&I lending terms are strongly positively correlated with the lending “standards.” In light of the results from the aforementioned literature, this suggests that the SLOOS “terms” are valid indicators of the state of credit conditions for C&I loans.

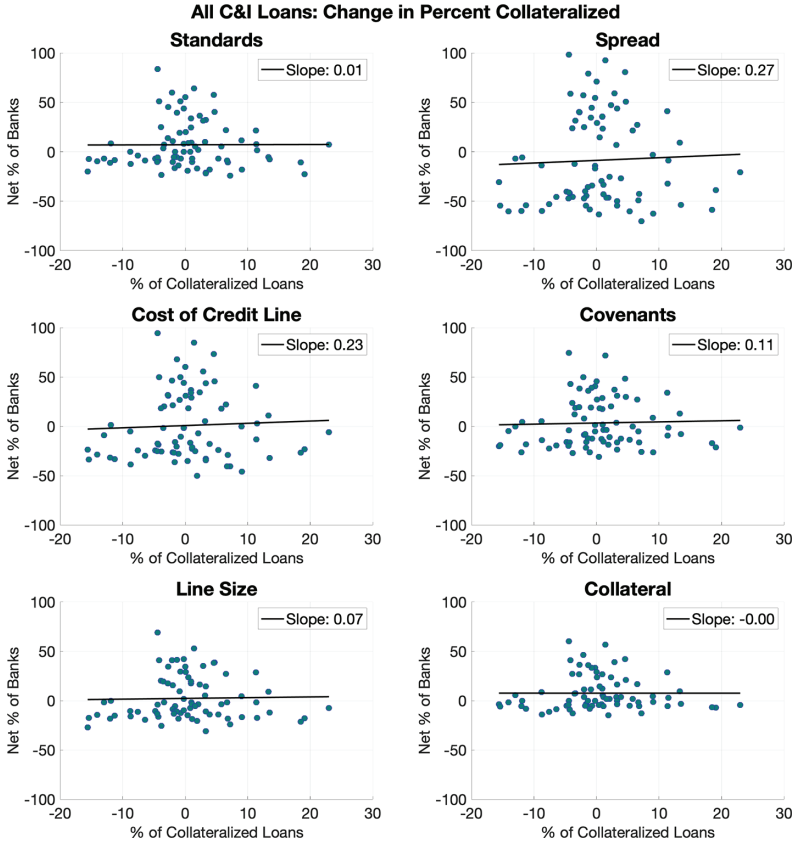
The C&I lending “terms” are also strongly correlated with each other, evidence of an underlying common factor(s). This is not surprising considering that different banks might adjust different contract terms (some may raise rates, some may increase collateral requirements, some may decrease the maximum line size). It is also possible that the same bank might choose to adjust rates for some subset of contracts, covenants for another, collateral requirements for another, etc. Therefore, it is important to account for the common factor(s) in order to isolate the effect of changes in each different loan “term.” I remove the common factor using the principal component decomposition of the “terms” and lending “standards.” Given this decomposition, I regress each of the “terms” and “standards” on the corresponding main principal component and use the residual as my transformed SLOOS variable. As I show next, this transformation effectively isolates adjustments in the corresponding loan contract “term.”

The magnitude of the correlations between the transformed C&I variables (bold elements below the main diagonal in table 2) becomes much smaller, a consequence of removing the common factor(s). This correlation also offers some insights about the relationship between the price and nonprice loan terms. All of the price terms (spread and cost of line) and nonprice terms (covenants, line size, and collateral) are positively correlated within each category but negatively correlated across categories. The correlation between price and nonprice terms becomes positive when the nonprice terms lead the price terms by about four quarters. This is consistent with basic economic theory; after a change in nonprice terms (shift in loan supply) price terms (slowly) adjust to reach the new equilibrium. Finally, the transformed “standards” are positively correlated with the nonprice terms and negatively correlated with the price terms. In other words, the “standards” reflect mainly nonprice factors.⁸

Within the nonprice terms, the correlation between covenants and collateral requirements is much stronger than either of their

⁸That “standards” reflect nonprice terms is not only intuitive but also consistent with previous studies. For instance, Lown and Morgan (2006, p. 1577) explicitly state that they use lending “standards” as a proxy for the full vector of nonprice lending terms.

Figure 1. Relationship between Different C&I Credit Variables and the Percent (value) of C&I Loans Secured by Collateral.



Source: SLOOS and E.2.

Note: The slope of the best fit line is included for each case.

correlations with the maximum line size. Again, this is not surprising given that covenants and collateral requirements both capture the willingness of banks to approve additional loans (*extensive margin*). On the other hand, the maximum line size mostly reflects credit conditions within the existing loans (*intensive margin*).

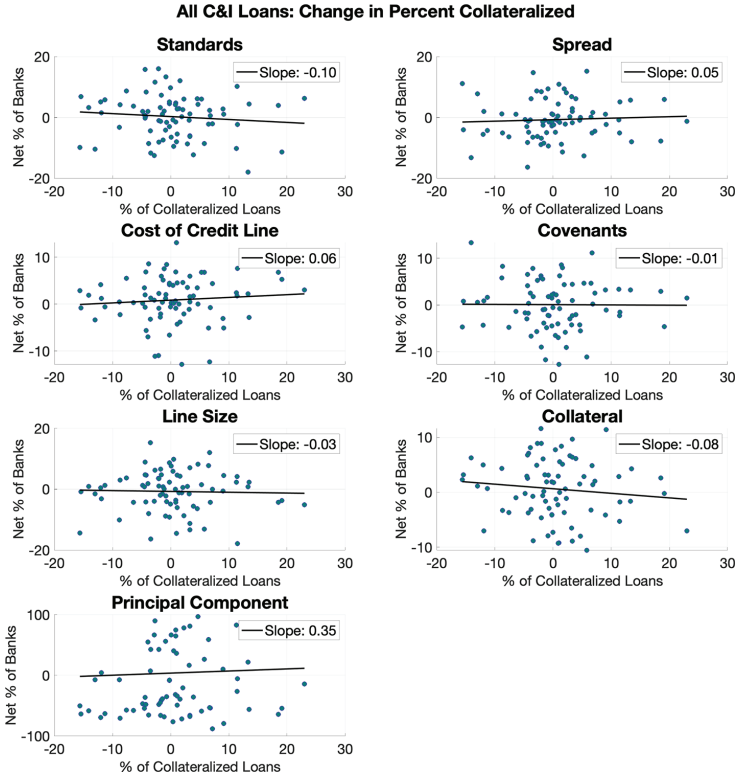
Figure 1 presents scatter plots that illustrate the relationship between the raw SLOOS C&I variables and the change in percent

(value) of C&I loans secured by collateral from the E.2. The best fit line and its corresponding slope are included in each graph. All price and nonprice terms are positively correlated with the change in the percent of loans secured by collateral, but the correlation is largest for the price terms (spread and cost of line). At first it might seem counterintuitive that a tightening in credit conditions leads to an increase in the percent of loans secured by collateral. However, economic theory suggests that a tightening in credit conditions would lead to a decrease in both total loans *and* collateralized loans. A larger decrease in total loans would explain the positive correlation. For the terms that more directly affect collateralized loans rather than noncollateralized loans (such as a tightening in collateral requirements or covenants), one would expect to see a smaller positive correlation, which is indeed the case.

However, collateral requirements should affect *only* collateralized loans; a tightening in collateral requirements should lead to a *decrease* in the percent of collateralized loans. That is, the correlation should be negative. Why is it not? Because the SLOOS variables (including the tightening of collateral requirements) are all contaminated by the common factor. Figure 2 is the equivalent of figure 1 but for the transformed SLOOS variables. A tightening in collateral requirements is indeed associated with a decrease in collateralized loans after accounting for the common factor. Importantly, the correlation is still positive for the price terms (spread and cost of line) given that these terms affect collateralized and noncollateralized loans. The results also reaffirm that the nonprice factors are strongly related to one another; tightening in the line size, covenant requirements, or lending standards are all associated with a decrease in collateralized loans. Finally, I have also included the common component in figure 2 to show that it is positively correlated with changes in collateralized loans. Again, this suggests the common component does indeed capture factors that affect all loans (I will argue shortly that it mostly captures changes in interest rates).

Table 3 summarizes the normalized covariance (i.e., slope of the best fit line) between different lending measures from the E.2 and the raw (panel A) and transformed (panel B) SLOOS variables for C&I loans. The raw SLOOS variables are all positively correlated with changes in interest rates and negatively correlated with changes in lending volume. Again, this observation reaffirms that the SLOOS

Figure 2. Relationship between Different C&I Variables (with the common component removed) and the Percent (value) of C&I Loans Secured by Collateral



Source: SLOOS and E.2.

Note: The slope of the best fit line is included for each case.

“terms” do capture changes in credit conditions.⁹ Note that the spread and cost of line size terms have the highest correlation with the interest rate, which is another reason why I refer to them as the price terms.

⁹The SLOOS variables are also negatively correlated with average measures from the E.2 (such as the average loan size and average maturity). However, the magnitude of the correlation is about an order of magnitude smaller than for the variables presented here. The reason for such small correlation is that averages reflect a ratio of intensive to extensive margins, and both are affected by changes of credit standards.

Table 3. Covariance between Lending Measures and SLOOS Variables

		A. Raw SLOOS						
Change In:	Standards	Spread	Cost of Line	Covenants	Line Size	Collateral		
Interest Rate	23.23	42.04	32.87	25.98	22.66	17.23		
Percent Collateralized	0.01	0.27	0.23	0.11	0.07	0.00		
Total Value (\$ Billions)	-0.20	-0.39	-0.32	-0.25	-0.26	-0.16		
		B. Transformed SLOOS						
Change In:	Standards	Spread	Cost of Line	Covenants	Line Size	Collateral	Common	
Interest Rate	1.49	-2.02	-1.51	2.08	2.17	0.79	70.08	
Percent Collateralized	-0.10	0.05	0.06	-0.01	-0.03	-0.08	0.35	
Total Value (\$ Billions)	0.01	0.03	0.01	-0.02	-0.07	0.00	-0.67	

Notes: The values presented in the table correspond to the normalized covariance: $cov(x, y)/var(x)$. y (columns) refer to the data series from the Senior Loan Officer Opinion Survey (SLOOS). x (rows) refer to the data series from the Survey of Terms Business Lending (E.2).

The correlation of the SLOOS variables with changes in the interest rates becomes much smaller once the common component is removed. In other words, the common component reflects most of the overall lending conditions captured by C&I lending rates. Interestingly, the correlation for the price terms becomes *negative*, while it remains positive for the nonprice terms. Although this might seem counterintuitive, it is actually consistent with standard economic theory. The results suggest that a tightening in the economy's credit conditions is manifested in the loan market by an increase in the lending rate accompanied by a tightening in the nonprice terms (standards, covenants, line size, and collateral). However, (aggregate) loan rates might not adjust as fast as government bond rates (think, for instance, of fixed-rate loan contracts), which implies a decrease in the C&I loan spread (and other price terms).

One might be concerned about the positive correlation between the nonprice terms and the changes in interest rate. After all, the goal of the transformation is to ensure that the SLOOS variables isolate changes in the different margins of adjustments *not captured* by changes in interest rates (hence the nonprice tag). Nonetheless, one must keep in mind that all of the transformed nonprice variables are negatively correlated with changes in the percent of collateralized loans. If the transformed nonprice terms were mainly driven by changes in the interest rate, this correlation would be positive (as discussed earlier). In fact, the correlation between the *actual* changes in the interest rate and the percent of collateralized loans is about 0.12. That is, the transformed standards and nonprice terms do indeed reflect margins of adjustments *other* than changes in interest rates (i.e., changes in the maximum line size, covenants, and collateral requirements).

The correlation of the SLOOS variables with the change in total C&I lending also becomes an order of magnitude smaller for all variables. However, it remains negative only for the three nonprice margins of adjustment (maximum line size, covenants, and collateral requirements). Furthermore, this negative correlation is significantly larger (in magnitude) for the maximum line size. By definition, the change in total value of loans is *conditional* on only the approved loans. Therefore, one would expect the intensive margins of adjustments (such as tightening of the line size) to matter the most.

The results show that this is indeed the case. Note also that correlation between total lending volume and the common component is also negative (and is the largest in magnitude), which provides further evidence that this term does indeed capture overall changes in the lending rate.

The previous discussion shows that, after accounting for a common component, the different SLOOS “terms” do indeed reflect adjustments in the margin to which they allude. The common component reflects mostly changes in the lending rate (overall credit conditions). The lending “standards” capture adjustments in the nonprice terms, which are also individually captured by the maximum line size, covenants, and collateral requirements. Finally, the spread and the cost of line variables (the price terms) account for changes in the loan interest rate relative to other (i.e., bond) rates.

4. Methodology

4.1 *Econometric Framework*

I use VARs that include the one-year U.S. Treasury yield, real GDP, and the consumer price index as the three macroeconomic variables, the excess bond premium as an indicator of overall credit conditions, the SLOOS loan demand, and (one-by-one) the SLOOS loan contract terms. Each VAR includes four lags.¹⁰

I use the external instrument methodology proposed by Gertler and Karadi (2015) to identify the monetary policy shock. I use this identification procedure for three reasons. First, it allows me to include credit variables in the VAR specification without imposing a priori restrictions on the interaction between them and the monetary policy indicator.¹¹ Second, depending on the choice of policy

¹⁰To alleviate overfitting concerns, I do robustness checks removing the SLOOS demand control and including only two lags instead of four (see section 5.2).

¹¹For instance, the identification scheme of Christiano, Eichenbaum, and Evans (1996) assumes that the federal funds rate responds to all the variables in the VAR within a period but not vice versa. For aggregate macroeconomic variables, such as prices or real output measures, this assumption might be justified if the frequency of the data is not too low (monthly or quarterly). For financial and credit variables, such an assumption is less likely to hold (even for monthly or quarterly data).

indicator and external instrument, the identified policy surprise can be informative not only about the current policy stance but also about the expected future policy stance. This is precisely why I use the one-year rate as the monetary policy indicator rather than the federal funds rate. The use of the one-year rate as the policy indicator does not imply that the Federal Reserve conducts policy by directly manipulating this rate. As the general consensus dictates, I presume that the Federal Reserve conducts policy by setting a target federal funds rate (i.e., the policy instrument). However, any movements in the federal funds rate affect the one-year rate per the standard term structure argument. In this sense, the one-year rate is an indicator of the monetary policy stance. The advantage of using this mid-term rate is that it captures movements in the expected future path of the policy instrument in addition to current movements.¹² Finally, this identification approach allows me to quantify the contribution of different channels to the monetary policy transmission mechanism.

I use the surprise in the three-month-ahead (FF_3) federal funds futures as my external instrument for the identification procedure. There are two advantages of using the surprise in the three months ahead over the surprise in current-month federal funds futures (FF_0). First, FF_3 reflects expectations of short rate movements further into the future.¹³ Second, the original FF_0 and FF_3 monthly series from Gertler and Karadi (2015) extend only through 2012:M6. Jarociński and Karaki (2020) have an updated version that extends through 2016:12, but they provide it only for FF_3 . Thus I construct my FF_3 series by taking the quarterly average of the updated monthly FF_3 series.

¹²Using the one-year rate as the policy indicator does also alleviate some of the concerns about the zero lower bound. Refer to Gertler and Karadi (2015) for a detailed discussion of this and other benefits of using mid-term rates as policy indicators over the federal funds rate. Kuttner (2001), Bernanke, Reinhart, and Sack (2004), and Swanson and Williams (2014) provide evidence that mid-term rates instrumented by surprises in futures contracts better capture the persistent effect of monetary policy news.

¹³For a more detailed discussion about the validity of futures rates surprises as external instruments for monetary policy shocks, refer to Kuttner (2001), Piazzesi and Swanson (2008), Hamilton (2009), Gertler and Karadi (2015), and Ramey (2016).

4.2 Monetary Policy Transmission

Let \mathbf{Z}_t denote the vector of variables included in the VAR, ϵ_t the vector of fundamental shocks, and $\epsilon_t^p \in \epsilon_t$ the fundamental monetary policy shock. The vector of reduced-form shocks can then be expressed as $\mathbf{u}_t = \mathbf{s}^p \epsilon_t^p + \tilde{\mathbf{S}}\epsilon_t$, where \mathbf{s}^p captures the impact of the monetary policy shock in each of the reduced-form errors. The advantage of the external instrument high-frequency indicators (HFI) procedure is that it allows one to identify the vector \mathbf{s}^p .¹⁴

Once \mathbf{s}^p is identified and the reduced-form VAR is estimated, one can easily assess and quantify the contribution of different variables to the transmission of monetary policy shocks. For any horizon $t \geq \tau$, a given sequence of monetary policy shocks $\{\epsilon_j^p\}_{j=\tau}^t$ leads to changes in \mathbf{Z}_t that are the result of the propagation of \mathbf{s}^p . In other words,

$$\mathbf{Z}_t = \mathbf{B}(L) \mathbf{Z}_{t-1} + \mathbf{s}^p \epsilon_t^p, \quad t \geq \tau \text{ and } \mathbf{Z}_{\tau-1} \text{ given}, \quad (1)$$

where all nonmonetary fundamental shocks have been set to zero. For $t = \tau$ and conditional on the system being unperturbed (i.e., $\mathbf{Z}_{\tau-1} = 0$), a monetary policy shock of one standard deviation implies that $\mathbf{Z}_\tau = \mathbf{s}^p$. In other words, $s_j^p \in \mathbf{s}^p$ is an indicator of the *contemporaneous transmission* of the monetary policy shock via variable $z_j \in \mathbf{Z}$.

For $t > \tau$, the *transmission* of monetary policy shocks via variable $z_j \in \mathbf{Z}$ depends on both $s_j^p \in \mathbf{s}^p$ and the reduced-form VAR coefficients $\mathbf{B}(L)$. Equation (1) can be used to obtain the impulse response functions (IRFs) after a one-time monetary policy shock at date τ (i.e., $\epsilon_\tau^p = 1$ and $\epsilon_t^p = 0$, $\forall t > \tau$). The importance of variable $z_j \in \mathbf{Z}$ for the transmission of monetary policy shocks can then be evaluated by comparing two sets of IRFs. The first set is just obtained using the estimated coefficients $\mathbf{B}(L)$ and the contemporaneous transmission vector \mathbf{s}^p . The second set is obtained by counterfactually setting $s_j^p = 0$ while keeping the VAR coefficients $\mathbf{B}(L)$ and all other elements of the vector \mathbf{s}^p at their estimated values. If variable $z_j \in \mathbf{Z}$ (for instance, one of the SLOOS lending terms)

¹⁴For details on the identification procedure, refer to online appendix A.

is relevant for monetary policy transmission, then the IRFs corresponding to the counterfactual experiment should be significantly different than their counterparts.¹⁵

Another way to evaluate the contribution of variable $z_j \in \mathbf{Z}$ to the transmission of monetary policy shocks is via the forecast error variance decomposition. For $q \in \{0, 1, 2, \dots\}$, let Ψ_q denote the matrix of coefficients corresponding to the moving-average representation of the reduced-form VAR.¹⁶ As usual, $\psi_q^{i,j}$ refers to the ($i^{\text{th}}, j^{\text{th}}$) element of Ψ_q . For any horizon $h = \tau - t$ define

$$\phi_{i,j}(h) \equiv \sum_{q=0}^{h-1} (\psi_q^{i,j} s_j^p)^2, \tag{2}$$

which measures the forecast error variance of variable $z_i \in \mathbf{Z}$ at horizon h due to changes (caused by contemporaneous monetary policy shocks) in variable $z_j \in \mathbf{Z}$.¹⁷ In other words, it measures the transmission of monetary policy shocks to variable z_i via variable z_j at horizon h . Note that $\phi_i(h) \equiv \sum_{z_j \in \mathbf{Z}} \phi_{i,j}(h)$ then measures the total variation in z_i due to monetary policy shocks. Suppose z_i is real GDP and z_j is one of the SLOOS lending terms. The ratio $\phi_{i,j}(h) / \phi_i(h)$ provides an idea of the contribution of changes in the SLOOS lending terms to the transmission of monetary policy to real GDP.

Finally, a third way to assess the contribution of variable $z_j \in \mathbf{Z}$ to the transmission of monetary policy shocks is using historical decomposition. The finite approximation of the moving-average representation of equation (1) can be written as

$$\tilde{\mathbf{Z}}_t = \sum_{q=0}^{t-1} \Psi_q \gamma \text{IV}_{t-q}, \tag{3}$$

where IV_t is the external instrument used for the identification of the monetary policy shock. The vector of coefficients γ is estimated as a byproduct of the two-stage least square implementation of the

¹⁵Instead of just shutting off the *contemporaneous* transmission via variable $z_t^j \in \mathbf{Z}_t$ (i.e., $s_j^p = 0$), one could also shut off the transmission via z_t^j for all periods (i.e., $z_t^j = 0, \forall t \geq \tau$).

¹⁶Note that one can easily obtain Ψ_q for $q \in \{0, 1, 2, \dots\}$ given the estimated coefficients of the reduced-form VAR.

¹⁷The derivation of equation (2) is presented in online appendix B.

identification procedure.¹⁸ Intuitively, if IV_t is a valid external instrument for the monetary policy shock (i.e., relevant and exogenous), then equation (3) follows given that $s^p \epsilon_t^p \propto \gamma IV_t$ and $\tilde{\mathbf{S}} \epsilon_t \perp \gamma IV_t$. The contribution of variable z_j to the transmission of monetary policy shocks can then be isolated by setting (counterfactually) $\gamma_k = 0, \forall \gamma_k \in \gamma, k \neq j$ in equation (3). Denote this counterfactual by $\tilde{\mathbf{Z}}_{c_j,t}$. Suppose $z_i \in \mathbf{Z}$ is real GDP. Then $\tilde{z}_{c_j,t}^i \in \tilde{\mathbf{Z}}_{c_j,t}$ refers to the element corresponding to real GDP in the finite approximation counterfactual. For each period t , the ratio $|\tilde{z}_{c_j,t}^i| / |\sum_{z_j \in \mathbf{Z}} \tilde{z}_{c_j,t}^i|$ captures the contribution of changes in variable $z_j \in \mathbf{Z}$ to the historical fluctuations in real GDP caused by monetary policy shocks.

5. Results

This section presents and validates the main results of the paper: a surprise monetary contraction leads to a decrease in loan supply, there is an important interaction between the loan and bond markets for monetary transmission, and adjustments in the nonprice loan terms (such as maximum line size, covenants, and collateral requirements) are relevant for monetary policy transmission.¹⁹

5.1 Discussion of Main Results

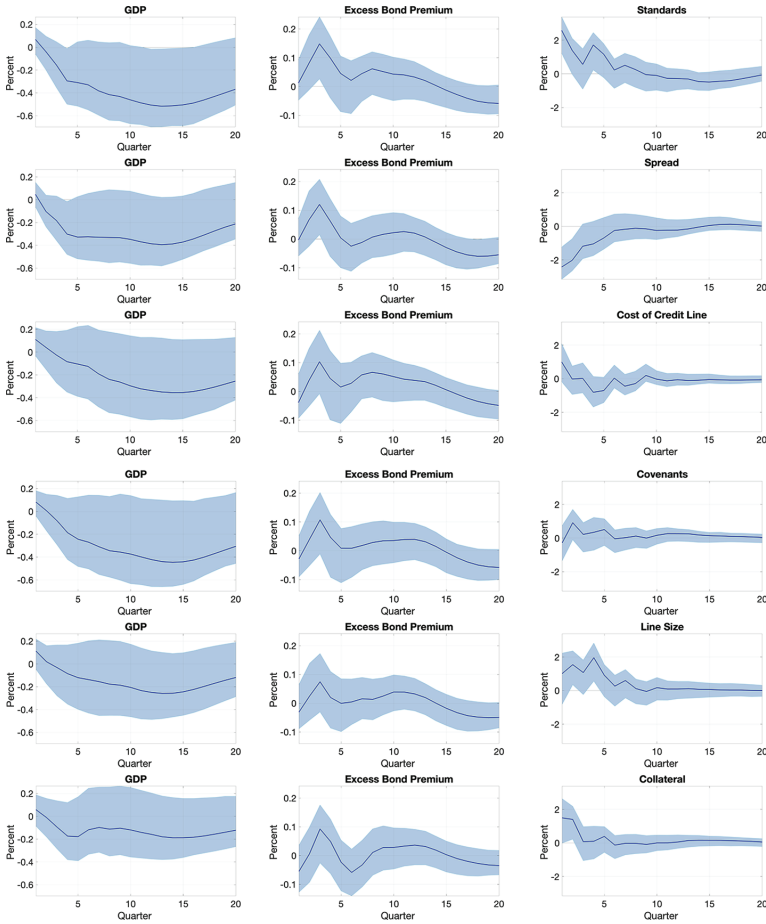
Figure 3 presents the impulse response functions for real GDP (left pane), the excess bond premium (middle pane), and the SLOOS net percent of banks tightening the specified C&I loan contract term (right pane) after a surprise monetary contraction.²⁰ The responses are robust across the different specifications and are consistent with conventional theory. The one-year rate increases by about 30 basis points upon impact and then reverts back to trend after roughly

¹⁸The derivation of equation (3) is presented in online appendix C.

¹⁹I obtain the confidence intervals for all the IRFs presented in this section using a wild bootstrap; see Gonçalves and Kilian (2004). The regression to obtain the SLOOS transformed variables as well as both stages of the identification procedure are included in the bootstrapping procedure, hence effectively addressing the “generated regression” problem.

²⁰The IRFs for the one-year government bond rate and the CPI can be found in online appendix D.

Figure 3. Effect of a Surprise Monetary Tightening on Credit Conditions for Corporate Bonds and C&I Loans



Notes: The IRFs correspond to one standard deviation of the monetary policy shock. The shaded area represents the 90 percent confidence interval.

six quarters. This increase is statistically significant across all specifications. The CPI does not experience any statistically significant change.²¹ GDP experiences a rather persistent decrease, which is

²¹In some specifications there is slight evidence of the price “puzzle”: the contractionary monetary policy shock induces a modest and (marginally) statistically significant increase in the CPI during the first quarter or two.

largest at a horizon of about 12 months, reaching as much as 0.6 percentage point. However, the GDP decrease is (marginally) statistically significant only for the specifications that include the lending standards and the loan spread.

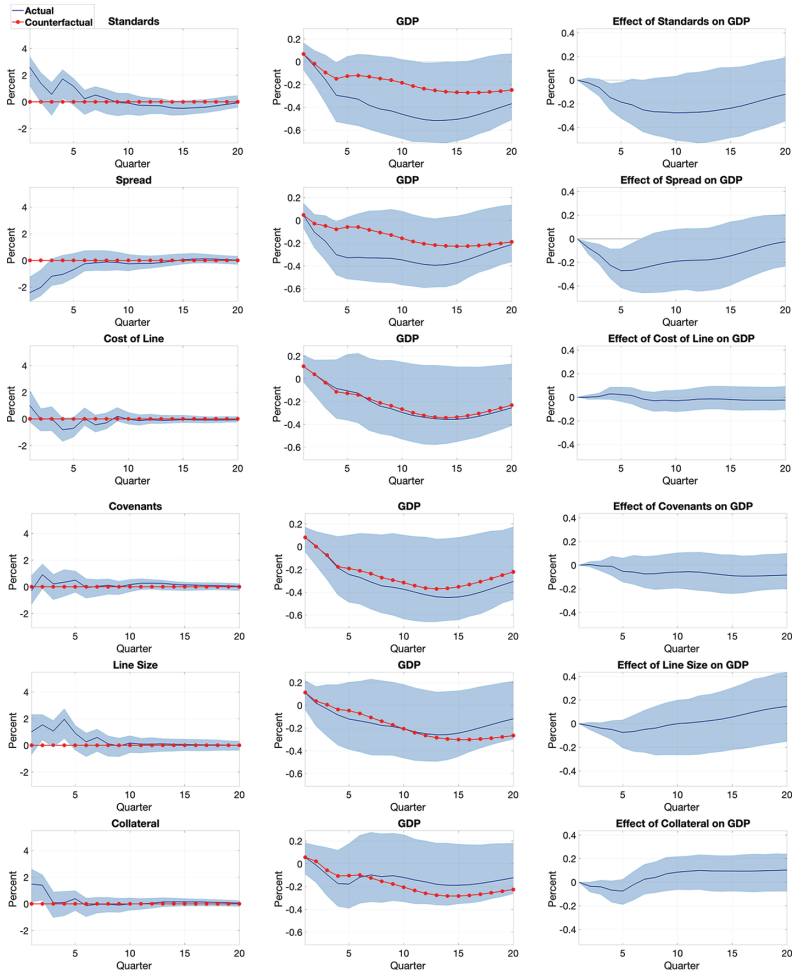
The response of the excess bond premium and the loan terms reflects a change in the credit conditions during the first five quarters after the shock. The excess bond premium and the SLOOS nonprice terms (standards, covenants, line size, and collateral) all tighten (increase). The SLOOS price terms (spread and cost of line) actually *relax* (decrease). These adjustments are consistent with a tightening in credit conditions as predicted by standard economic theory. After a monetary contraction loan supply decreases (tightening of nonprice terms) and the price terms (slowly) adjust to reach the new equilibrium. Given that C&I loan rates do not adjust as fast as government bond rates, this implies a decrease in the C&I loan spread and other price terms.

Furthermore, note the timing of the changes in the credit conditions. While the tightening in the loan contract terms happens immediately upon the shock, the increase in the excess bond premium happens with a lag. I interpret this as evidence of an important interaction between the corporate loan and bond markets. The lagged increase in the excess bond premium suggests that firms turn to the (corporate) bond market to raise funds after they are unable to get funds from banks due to the tightened lending conditions.

Note that this interaction between these two markets is absent in the standard mechanisms à la Kiyotaki and Moore (1997) and à la Bernanke, Gertler, and Gilchrist (1999) commonly used for modeling financial frictions. These mechanisms provide theoretical foundations for thinking about a tightening in loan market credit conditions “as if” it resulted in an increase in lending spreads. My results suggest that the lending spread actually decreases.

At this point it should be clear that several loan contract terms adjust in response to monetary policy surprises. I next consider if these adjustments are relevant for monetary policy *transmission*. Figure 4 presents the IRF counterfactual described in section 4.2. The solid (blue) line in the left and middle panes corresponds to

Figure 4. Contribution of C&I Loan Terms to Monetary Transmission



Notes: The IRFs correspond to one standard deviation of the monetary policy shock. The shaded area represents the 90 percent confidence interval.

the original IRFs.²² The line with (red) dots corresponds to the counterfactual responses where transmission via the loan contract

²²For figures in color, see the online version of the paper, available at <http://www.ijcb.org>.

term has been shut off. The right pane corresponds to the difference between the original response in GDP and its response under the counterfactual. If adjustments in the different loan contract terms are relevant for monetary transmission, then the responses in the right pane should be nonzero and statistically significant.

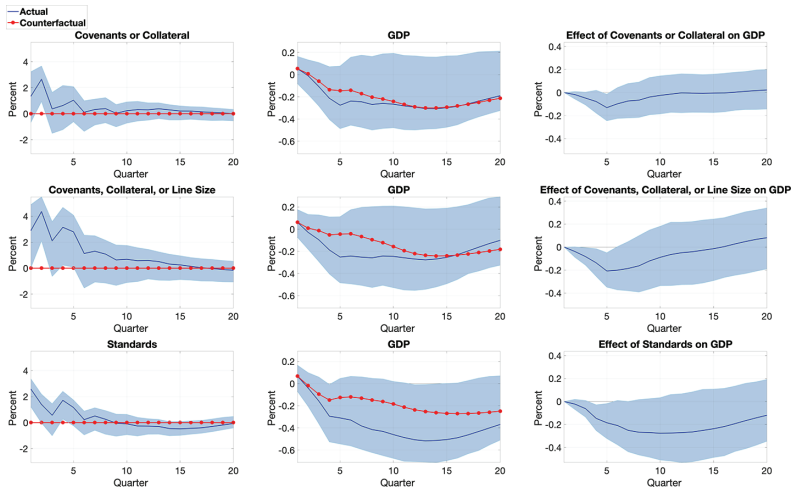
The results from figure 4 suggest that changes in lending standards and loan spreads are relevant for monetary transmission. Given that a tightening in standards together with a relaxation in loan spreads is consistent with a decrease in loan supply, these results imply that the contraction in loan supply following the monetary surprise accounts for a statistically significant decrease in GDP of up to 0.3 percentage point.

Recall that the lending standards capture adjustments in loan contract terms *other* than changes in interest rates; they capture changes in the nonprice terms such as maximum line size, covenants, and collateral requirements (see section 3.2). Therefore, the results from figure 4 suggest that changes in these nonprice terms are relevant for monetary transmission. Nonetheless, nonprice terms are not *separately* relevant for monetary transmission; the individual series in the right column of figure 4 all include zero.

Figure 5 presents further evidence to support the claim that nonprice terms are *collectively* relevant for monetary policy transmission. The VAR specifications used to construct these IRF counterfactuals reflect adjustments in covenants or collateral requirements (first row) and adjustments in covenants, collateral requirements, or the maximum line size (second row).²³ As can be seen from the figure, the effect of the nonprice terms on GDP

²³Effectively, this amounts to creating new variables adding the net percent of banks tightening covenants, the net percent of banks tightening collateral requirements, and the net percent of banks tightening the maximum line size. I then include the transformed version of these variables (i.e., with the principal component removed) as the SLOOS variable in the baseline VAR specifications. Note that this series reflects aggregate tightening in any of the nonprice margins. If a bank tightened more than one margin simultaneously, it would be “double” counted, thus reflecting the simultaneous tightening in multiple margins. If a bank tightened one margin but relaxed another one, it would be “washed out,” thus reflecting the simultaneous tightening in both margins (i.e., no tightening at all). This assumes all three margins are equally relevant. Given that adjustments in collateral requirements seem to matter most, using these series might actually underestimate the results.

Figure 5. Contribution of Adjustments in Covenants or Collateral Requirements to Monetary Transmission



Notes: The IRFs correspond to one standard deviation of the monetary policy shock. The shaded area represents the 90 percent confidence interval.

becomes statistically significant when thinking about them collectively. Furthermore, the results show that collective response of the nonprice terms has a very similar effect on GDP as that predicted by the lending standards (third row).

The previous results are corroborated when the contribution of the different loan terms is evaluated using the forecast error variance or historical decompositions.²⁴

5.2 Robustness

One potential concern is that omitted variables in the VAR specifications might be (partially) driving the previous results. I use the excess bond premium as the indicator for the overall credit conditions precisely to attenuate this concern. Several studies, including Gilchrist and Zakrajšek (2012), have shown that the excess bond

²⁴The results for the forecast error variance and historical decompositions can be found in online appendixes E and F).

premium outperforms every other financial indicator in its forecasting ability for economic activity. Thus the excess bond premium conveniently summarizes the information from variables that might be left out of the VAR specifications.

Furthermore, I show that the results are robust when using other indicators of overall credit conditions.²⁵ To the extent that these indicators contain *less* information about the economy than the excess bond premium, these alternative specifications are more prone to omitted variables by construction. Although the actual GDP responses are somewhat different when the excess bond premium is not included, the responses of the loan terms remain remarkably similar and the results about the contribution of the different loan terms to monetary transmission are virtually unchanged. In other words, the results are robust to omitted variables.

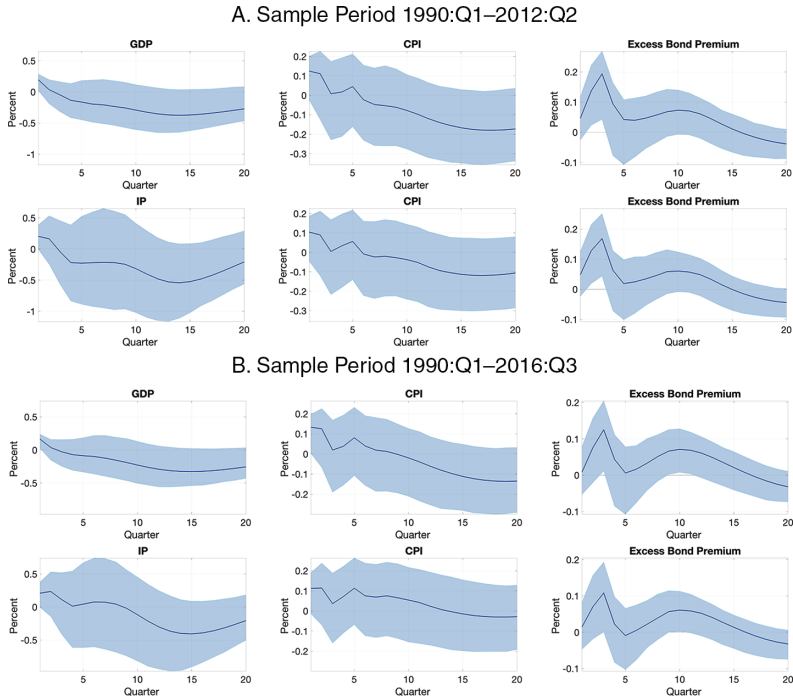
Another potential concern is related to the validity of the SLOOS variables. The previous results rely on the different SLOOS terms actually capturing the margin of adjustment to which they allude. Section 3.2 presented evidence that showed the SLOOS variables are indeed valid and that changes in the lending standards capture changes in the nonprice terms. Furthermore, figure G.3 in online appendix G shows that the response of the *actual* C&I spread is consistent with the response of the SLOOS net percent of banks tightening the spread. I interpret this as further evidence of the validity of the SLOOS terms.

A third concern is related to the identification of the monetary policy shock using the surprise in the three-month-ahead federal funds futures (FF_3) as the external instrument. Panel A in figure 6 shows that my empirical setup yields similar results to those from Gertler and Karadi (2015).²⁶ Keep in mind that they use monthly data, they include the industrial production index (IP) as the measure of macroeconomic activity, they include the mortgage and commercial paper spreads as additional controls, and their sample period ends in 2012:Q2. In order to make the comparison, I use IP instead of real GDP as the macroeconomic indicator (although I keep the quarterly frequency) and I restrict my sample period to 1990:

²⁵See online appendix G.

²⁶See figure 1 on page 61 of their paper.

Figure 6. Reproducing the Results from Gertler and Karadi (2015)



Notes: The IRFs correspond to one standard deviation of the monetary policy shock. The shaded area represents the 90 percent confidence interval.

Q1–2012:Q2. However, I don’t include the mortgage and commercial paper spreads.

The responses of the CPI and EBP are fairly similar to those from Gertler and Karadi (2015); there is no statistically significant change in the CPI, while the EBP experiences a statistically significant increase between 5 and 25 basis points during the first three quarters. The IP response is also qualitatively and quantitatively similar, although Gertler and Karadi (2015) find it to be statistically significant. Altogether, I interpret this as evidence of the monetary policy shock being properly identified within my setup.

Figure 6 also includes the responses using real GDP as the macroeconomic indicator and using the full sample period 1990:Q1–2016:Q3. I provide this for reference and to show that the responses

are similar to the responses from the actual VARs I use in my study (see figure 3).

It must also be noted that I can safely rule out a weak instrument problem when using FF_3 in the identification procedure. To attenuate any remaining concerns, I conduct a robustness check and verify that the main results remain unchanged when using the surprise in the current-month federal funds future (FF_0) as the instrument.²⁷

I also address concerns about overfitting by considering two alternative VAR specifications. In the first one I use only two lags instead of four. In the second one I drop the SLOOS demand controls to reduce the number of regressors.²⁸ In both cases the nonprice terms remain relevant for monetary transmission, and adjustments in these terms lead to significant changes in GDP. Interestingly, changes in collateral requirements become relatively more important within the nonprice terms. This finding suggests that adjustments in collateral requirements are more persistent than adjustments in the other margins and are more strongly associated with changes in loan demand.

Lastly, the results remain robust when using different subsamples in the estimation procedure.²⁹ For instance, the adjustment in the nonprice loan terms becomes even more relevant for monetary transmission when restricting the sample to the pre-crisis period (1990:Q1–2007:Q4). Not only is their contribution to changes in GDP statistically significant, but it almost doubles in magnitude. This is not surprising given that the zero lower bound and other factors might have hindered the overall effectiveness of monetary transmission during the crisis and post-crisis periods.

6. Conclusion

I study the transmission of monetary policy shocks via changes in nonprice contract terms for C&I loans. I use the external instrument approach from Gertler and Karadi (2015) and data from the Senior

²⁷Refer to online appendixes H and I for further details.

²⁸Refer to online appendixes J and K for further details.

²⁹Refer to online appendix L for further details. Additionally, a previous version of this paper used the federal funds surprises and excess bond premium series from Gertler and Karadi (2015), which effectively restricted the sample period to 1990:Q1–2012:Q2. The main results from this earlier version are consistent with the current results.

Loan Officer Opinion Survey to identify changes in loan contract terms.

I find that a surprise monetary contraction leads to a decrease in loan supply characterized by a relaxation in the loan interest rate spread and a tightening in the nonprice terms (maximum line size, covenants, and collateral requirements). The decrease in the spreads is a consequence of loan rates not adjusting as fast as (government) bond rates. The tightening in the nonprice terms is responsible for a statistically significant decrease in GDP of about 0.3 percentage point. Although the contribution of adjustments in collateral requirements is the largest, my results suggest that the changes in nonprice terms are not individually, but collectively, relevant for monetary transmission.

My results also shed light on the interaction between the loan and bond markets for monetary transmission. I find that the tightening in the loan contract terms happens immediately upon the monetary contraction, while the increase in the excess bond premium happens with a lag. The lagged increase in the excess bond premium suggests that firms turn to the (corporate) bond market to raise funds after they are unable to get funds from banks due to the tightened lending conditions.

My study provides empirical support for modeling financial frictions à la Kiyotaki and Moore (1997) and à la Bernanke, Gertler, and Gilchrist (1999) and it uncovers two avenues that could be useful to resolve the critique that these types of financial frictions are quantitatively unimportant. The first one is considering other non-price margins of adjustments in addition to collateral requirements. The second one is explicitly modeling the interaction between the corporate bond and loan markets.

References

- Barakchian, M. S., and C. Crowe. 2013. "Monetary Policy Matters: Evidence from New Shocks Data." *Journal of Monetary Economics* 60 (8): 950–66.
- Bassett, W., M. Chosak, J. Driscoll, and E. Zakrajšek. 2014. "Changes in Bank Lending Standards and the Macroeconomy." *Journal of Monetary Economics* 62 (March): 23–40.

- Bernanke, B. S., M. Gertler, and S. Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, Vol. 1, ed. J. B. Taylor and M. Woodford, 1341–93 (chapter 21). Amsterdam: Elsevier.
- Bernanke, B. S., V. R. Reinhart, and B. P. Sack. 2004. "Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment." *Brookings Papers on Economic Activity* 34 (2): 1–100.
- Christiano, L. J., M. Eichenbaum, and C. L. Evans. 1996. "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds." *Review of Economics and Statistics* 78 (1): 16–34.
- Ciccarelli, M., A. Maddaloni, and J.-L. Peydró. 2015. "Trusting the Bankers: A New Look at the Credit Channel of Monetary Policy." *Review of Economic Dynamics* 18 (4): 979–1002.
- Faust, J., E. T. Swanson, and J. H. Wright. 2004. "Identifying VARs Based on High Frequency Futures Data." *Journal of Monetary Economics* 51 (6): 1107–31.
- Gertler, M., and P. Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7 (1): 44–76.
- Gilchrist, S., and E. Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102 (4): 1692–1720.
- Gonçalves, S., and L. Kilian. 2004. "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form." *Journal of Econometrics* 123 (1): 89–120.
- Gürkaynak, R. S., B. Sack, and E. T. Swanson. 2005. "Do Actions Speak Louder than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1 (1, May): 55–93.
- Hamilton, J. D. 2009. "Daily Changes in Federal Funds Futures Prices." *Journal of Money, Credit and Banking* 41 (4): 567–82.
- Jarociński, M., and P. Karaki. 2020. "Deconstructing Monetary Surprises—The Role of Information Shocks." *American Economic Journal: Macroeconomics* 12 (2): 1–43.
- Jiménez, G., S. Ongena, J.-L. Peydró, and J. Saurina. 2012. "Credit Supply and Monetary Policy: Identifying the Bank Balance-sheet Channel with Loan Applications." *American Economic Review* 102 (5): 2301–26.

- Kiyotaki, N., and J. Moore. 1997. "Credit Cycles." *Journal of Political Economy* 105 (2): 211–48.
- Kuttner, K. N. 2001. "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Markets." *Journal of Monetary Economics* 47 (3): 523–44.
- Lown, C., and D. P. Morgan. 2006. "The Credit Cycle and the Business Cycle: New Findings Using the Loan Officer Opinion Survey." *Journal of Money, Credit and Banking* 38 (6): 1575–96.
- Lown, C. S., D. P. Morgan, and S. Rohatgi. 2000. "Listening to Loan Officers: The Impact of Commercial Credit Standards on Lending and Output." *Economic Policy Review* (Federal Reserve Bank of New York) 6 (2): 1–16.
- Maddaloni, A., and J.-L. Peydró. 2011. "Bank Risk-taking, Securitization, Supervision, and Low Interest Rates: Evidence from the Euro-area and the U.S. Lending Standards." *Review of Financial Studies* 24 (6): 2121–65.
- . 2013. "Monetary Policy, Macroprudential Policy, and Banking Stability: Evidence from the Euro Area." *International Journal of Central Banking* 9 (1, March): 121–69.
- Piazzesi, M., and E. T. Swanson. 2008. "Futures Prices as Risk-adjusted Forecasts of Monetary Policy." *Journal of Monetary Economics* 55 (4): 677–91.
- Ramey, V. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics*, Vol. 2, ed. J. B. Taylor and H. Uhlig, 71–162 (chapter 2). Elsevier.
- Swanson, E. T., and J. C. Williams. 2014. "Measuring the Effect of the Zero Lower Bound on Medium- and Longer-term Interest Rates." *American Economic Review* 104 (10): 3154–85.

International Trade Finance and the Cost Channel of Monetary Policy in Open Economies*

Nikhil Patel

International Monetary Fund

This paper studies the role of international trade finance in the transmission mechanism of monetary policy in a two-country dynamic stochastic general equilibrium (DSGE) framework. The model shows that trade finance can both amplify or mitigate the impact of shocks, depending on the degree to which countries differ in price stickiness and dependence on trade finance. The model is estimated with Bayesian techniques using macroeconomic data from the United States and the euro zone and reveals the impact of trade finance to be quantitatively important, especially for spillover effects of shocks across countries. It significantly alters the interpretation of the sources and propagation of business cycles. In particular, accounting for trade finance makes external shocks much less important for business cycles in the euro area. At the same time, spillover effects of U.S. monetary policy on euro-area output are much larger.

JEL Codes: F44, F41, E44, E52.

*I am grateful to Shang-Jin Wei, Stephanie Schmitt-Grohé, and Martin Uribe for extensive guidance. I would also like to thank Boragan Aruoba and two anonymous referees for extensive comments, and Scott Davis, Michael Devereux, Keshav Dogra, Torsten Ehlers, Yang Jiao, Frederic Mishkin, Emi Nakamura, Jaromir Nosal, Christopher Otrok, Pablo Ottonello, Ricardo Reis, Jon Steinsson, David Weinstein, James Yetman, and seminar participants at various institutions for valuable comments and discussions. Part of this research was conducted when I was visiting the Hong Kong Monetary Authority (HKMA) and Hong Kong Institute for Monetary Research (HKIMR). I am grateful to them for their hospitality and support. The views expressed here are those of the author and do not necessarily correspond to those of the IMF, HKIMR, or HKMA. All errors are the sole responsibility of the author. Author contact: npatel@imf.org.

1. Introduction

While the literature on trade finance is extensive,¹ the implications of trade finance for business cycle fluctuations in macroeconomic models remain understudied. This omission is conspicuous given the fact that open-economy models that are commonly used for policy analysis and forecasting typically give a central role to international trade. Indeed, trade is the primary and in some cases the only channel through which shocks can be transmitted across countries in these models.² This paper studies business cycle implications of trade finance through the lens of an estimated two-country New Keynesian dynamic stochastic general equilibrium (DSGE) model.

The term “trade finance” is used in the literature to describe a number of different financing arrangements. These include direct lending by banks to the exporter and/or the importer, interfirm trade credit, open account (i.e., post-delivery payment), or cash in advance.³ Recognizing that all these mechanisms involve at least one of the parties engaging in borrowing at an interest rate that is potentially affected by changes in monetary policy, trade finance in the paper is introduced by augmenting the cost channel of monetary policy. While there exists a sizable literature that studies different aspects of the cost channel of monetary policy, including extensions to open-economy settings (see, for instance, Gertler, Gilchrist, and Natalucci 2007 and Gilchrist 2003), these models do not distinguish between the external finance dependence of international and intranational trade, a distinction that the international trade literature has strongly emphasized. This paper models this distinction and shows that it is important not only quantitatively but also qualitatively in terms of the sign of the effects that the cost channel of monetary policy can generate.

¹Bekaert and Hodrick (2017) identify trade finance as the “fundamental problem in international trade.” According to the estimates of the Committee on the Global Financial System (CGFS), \$6.5–8 trillion worth of bank-intermediated trade finance was provided during the year 2011, which, at around 10 percent of global gross domestic product (GDP) and 30 percent of global trade, is a fairly sizable number in itself, even though it does not include letters of credit and other forms of trade finance not explicitly involving bank loans.

²See, for instance, Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2003).

³See Ahn, Amiti, and Weinstein (2011) and Schmidt-Eisenlohr (2013).

The standard cost channel of monetary policy typically amplifies the output effect of domestic shocks that hit the economy.⁴ On the other hand, this paper shows that the cost channel when combined with trade finance can either amplify or mitigate the effects of shocks. Consider a monetary contraction in the home economy, which leads to a fall in domestic aggregate demand and prices. If importing firms are constrained to borrow at their respective home interest rates, then foreign imports into the home country become more expensive, whereas imports into the foreign country (i.e., home exports) become cheaper for foreign consumers, leading to a higher demand for the latter and a lower demand for the former. As a result, the trade finance plays the role of cushioning the effect of the original monetary contraction on home output. If on the other hand exporting firms (instead of importing firms) are financially constrained and borrow in their domestic interest rate, then the trade finance constraints can amplify the effect of the monetary contraction on home GDP in the example just described.

Elaborating on these points, the first part of the paper focuses on studying the impact of trade finance on the transmission mechanism of monetary policy shocks through simulations under alternate scenarios. It illustrates how the effect depends critically on parameters characterizing the trade sector in the model, including the degree of price stickiness (and asymmetry across countries in this parameter) and parameters quantifying the external finance dependence of trade flows. Various sources of this asymmetry are identified and their implications are explored. Moreover, because monetary policy has both an exogenous and endogenous component, these additional features not only affect the propagation of monetary policy shocks themselves but also the propagation of all other shocks via the endogenous component of monetary policy.

The general nature of the implications that emerge from the model can be summarized under two polar scenarios depending on whether the countries are symmetric with respect to each other in regard to their external sectors. External sectors could differ due to the degree of external finance dependence, price flexibility, and

⁴See, for instance, Barth and Ramey (2002).

currency denomination of trade finance contracts, all of which in turn could be functions of the nature of export bundles of countries. When the external sectors are symmetric across countries, incorporation of trade finance leads to sharp movements in trade volumes but has negligible impact on GDP. When global interest rates are high, international trade becomes more expensive, which leads to higher import prices for both countries. Both countries shift away from imports and towards their respective domestically produced goods in such a way that the net effect on the GDP of both countries is minimal. On the other hand, when countries are asymmetric in any of these dimensions, the demand shifts do not offset one another, and trade finance can significantly alter the response of GDP to various shocks that hit the economy.

Given that these parameters play a critical role in affecting business cycle fluctuations and for the most part extant literature is not very informative on their values, uncovering values of these parameters and relative differences across trading partners is likely to be a fruitful avenue for future research. The second part of the paper takes the first step in this direction by estimating a two-country DSGE model with trade finance using data from two regions that constitute one of the largest trading relationships in the world—the United States and the euro zone (EZ). The focus of the estimation exercise is threefold: (i) parameter estimation, (ii) model comparison, and (iii) a quantitative analysis of the role played by trade finance in business cycle fluctuations. Regarding parameter estimation, the estimates reveal asymmetries in the degree of price stickiness in imports between the United States and the European Union (EU). In particular, retail prices of U.S. imports are found to be more flexible than their European counterparts.

While open-economy macro models typically give a central role to international trade, by omitting trade finance they ignore an important feature of international trade which has been shown to be important in the trade literature. How significant is this omission, and should there be a move towards incorporating models involving trade finance? To this end, estimation of different versions of the model (in particular, ones with and without trade finance) provide strong evidence in favor of models incorporating trade finance and show that trade finance is indeed quantitatively important in accounting for business cycle fluctuations.

The model makes the simplifying assumption that exporters and importers are not allowed to switch between sources of funding in response to shocks. While there is an extensive literature documenting that firms' sources of funding are sticky, some recent studies focusing specifically on international trade finance have found that that exporters and importers do indeed change their sources of funding in response to shocks.⁵ The implications of such switching are discussed below, while modeling optimal funding choice remains a fruitful avenue for future research.

The remainder of this paper is organized as follows. Section 2 begins with a brief literature review. Section 3 lays out the main features of the model and discusses the equilibrium conditions. Section 4 presents a calibration- and simulation-based exercise to illustrate different features of the model. Section 5 undertakes Bayesian estimation of the model, and section 6 concludes.

2. Related Literature

This paper is linked to several different strands in the literature at the intersection of macroeconomics, monetary economics, and international trade. The incorporation of credit constraints in this paper is motivated by the extensive empirical literature on trade finance and its link to monetary policy. This literature has documented—across countries and time—the higher reliance of international trade on external finance compared with intranational trade. Ju, Lin, and Wei (2013) employ a large bilateral sector-level trade data set for the years 1970–2000 to study the effect of monetary policy tightening on export behavior. They find that the sectors relying more on external finance are disproportionately largely affected by monetary tightening, and that the exporting behavior is affected more than domestic sales. Using monthly data on U.S. imports, Chor and Manova (2012) find that the United States imported less from countries with higher interest rates and tighter credit conditions. Using a panel of 91 countries from 1980 to 1997, Manova (2008) shows that equity market liberalizations are positively associated with higher exports. Manova, Wei, and Zhang (2011) report similar results using

⁵See, for instance, Anràs and Foley (2015), Demir and Javorcik (2018), and Garcia-Marin, Justel, and Schmidt-Eisenlohr (2019).

firm-level data from China. Based on survey data from Italian manufacturing firms, Minetti and Zhu (2011) report that credit rationing affects international sales more than domestic sales. Using a detailed matched firm-level data set for banks and firms in Japan, Amiti and Weinstein (2009) find that the health of the banking sector is much more influential in determining exporting behavior of firms compared with their domestic sales.

On the theoretical front, several explanations for this phenomenon have been explored in the literature. The most common explanation hinges on the fact that international shipments take more time than domestic shipments (both travel time and time taken for documentation and clearances),⁶ which implies that producers have to incur costs of production much before revenues are obtained. Feenstra, Li, and Yu (2014) provide a theoretical model incorporating these ideas. International trade is also likely to be more intensive in external finance because of higher information asymmetries associated with cross-border transactions.

Recognizing the need for trade finance, there is also a growing literature on the optimal financing arrangement. In theoretical frameworks, Ahn (2014) and Schmidt-Eisenlohr (2013) study how the optimal financing arrangement depends on the financial market characteristics of both the source and the destination country. Ahn, Khandelwal, and Wei (2011), Hoefele, Schmidt-Eisenlohr, and Yu (2016), and Niepmann and Schmidt-Eisenlohr (2017) test the occurrences of different financing arrangements in the data against these theories and find the evidence to be broadly consistent. Custom data suggest that open account is the dominant financing form, with a share of around 80 percent of trade reported for Turkey, Chile, and Colombia in Ahn, Khandelwal, and Wei (2011) and Demir and Javorcik (2018).⁷ For the United States, Antràs and Foley (2015) also find a large role for cash in advance when looking at the transaction-level data from a U.S. exporter of frozen and refrigerated food products.

⁶See Djankov, Freund, and Pham (2006) and Hummels and Schaur (2013).

⁷This share, although high, is still less than estimates of the share of open account in domestic transactions in advanced economies. Ellingsen, Jacobson, and von Schedvin (2016), for instance, find the open account share to be close to 100 percent for domestic transactions in Sweden.

An alternative to bank-intermediated trade finance is trade credit, or the direct extension of credit between buyers and suppliers. Although the two are substitutes and one would expect firms to turn from bank-intermediated trade finance to trade credits, the evidence supporting this hypothesis is mixed.⁸

In its exploration of the role of the cost channel of monetary policy in open-economy settings, the paper has several precedents in the closed-economy literature. Using industry-level data from the United States, Barth and Ramey (2002) provide compelling evidence in favor of the cost channel of monetary policy. Dedola and Lippi (2005) report similar conclusions based on a richer data set containing information on 21 manufacturing sectors from five OECD countries. Ravenna and Walsh (2006) highlight the presence of the cost channel on the basis of parameters estimates based on their estimation of the Phillips curve for the United States. They also provide a characterization of the optimal monetary policy problem in the presence of these cost side effects. In advanced economies monetary policy is primarily conducted via open market operations which affect the balance sheets of banks directly. If cost side effects of monetary policy are present, one would expect countries with bank-based systems to be more sensitive to monetary policy shocks. This is exactly what Cecchetti (1999) and Kashyap and Stein (1997) find. Moreover, based on joint BIS-IMF-OECD-World Bank statistics on external debt, Auboin (2007) documents that 80 percent of the providers of trade finance are private banks.⁹

The paper also builds on ideas developed in the literature on vertical specialization and multiple-stage production. Huang and Liu (2001, 2007) and Wong and Eng (2013) are among the many papers that have used these features to explain various empirical stylized facts that standard models have difficulty accounting for. This paper builds a model that would allow multiple-stage trade intermediation to act as an amplification mechanism for shocks due to borrowing constraints. Similar ideas incorporating liquidity constraints have

⁸See Asmundson et al. (2011) and Choi and Kim (2005) as two examples of the mixed evidence.

⁹In the Lehman bankruptcy 6 of the 30 largest unsecured claims against Lehman were letters of credit.

been applied in a closed-economy setting by Bigio and La'O (2013) and Kalemli-Ozcan et al. (2013).

3. Model

The model in this paper builds on the framework used in Galí and Monacelli (2005) and Lubik and Schorfheide (2006), which in turn fit into the New Open Economy Macroeconomics (NOEM) paradigm of Obstfeld, Rogoff, and Wren-Lewis (1996).¹⁰ In particular, it builds on Lubik and Schorfheide (2006) by modeling a cost channel of monetary policy and allowing for trade finance and multiple stages in production of exports. Apart from these features (which are limited to the import-export sector), the rest of the model is identical to Lubik and Schorfheide (2006).

The world economy is assumed to comprise two countries of equal size. Households have preferences over domestic and foreign goods and supply labor to firms. There are two sets of firms in each economy—production firms and trade firms. Prices are assumed to be sticky in both the domestic and import sector. The monetary authority uses the short-term nominal interest rate as its instrument. For brevity, only the home economy is described in detail below. The foreign economy is assumed to be isomorphic.

3.1 Households

The household side of the economy is characterized by a representative consumer with preferences over consumption and leisure given by the following utility function:

$$U(C_t^h, H_t^h, N_t^h) = \frac{1}{1 - \sigma_c} \left(\frac{C_t^h - H_t^h}{A_t} \right)^{1 - \sigma_c} - \frac{1}{1 + \sigma_L} N_t^{h1 + \sigma_L}. \quad (1)$$

Here C_t^h is consumption, N_t^h is the labor supply, and H_t^h ($=\chi C_{t-1}^h$) is the habit stock going into period t . A_t is a non-stationary worldwide productivity shock which evolves according to

$$A_t = Z_t (\gamma A_{t-1}). \quad (2)$$

¹⁰See Lane (2001) for a survey of the NOEM literature.

Z_t is an exogenous component and γ denotes the trend growth rate of world productivity. Agents are thus assumed to derive utility from effective consumption relative to the level of global technology.¹¹ Preferences are characterized by internal habits.¹²

There is a constant elasticity of substitution (CES) aggregator for C_t^h :

$$C_t^h = \left[(1 - \alpha)^{\frac{1}{\eta}} (C_t^{hh})^{\frac{\eta-1}{\eta}} + \alpha^{\frac{1}{\eta}} (C_t^{fh})^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}. \tag{3}$$

Here C_t^{hh} and C_t^{fh} denote the home- and foreign-produced components in the consumption bundle of country h . η is the elasticity of substitution between domestic and foreign aggregates and α parameterizes the home bias in consumption. The associated price index, which is also the consumer price index (CPI) in the home country, is given by

$$P_t^{h,cpi} = \left[(1 - \alpha) (P_t^{hh})^{1-\eta} + \alpha (P_t^{fh})^{1-\eta} \right]^{\frac{1}{1-\eta}}, \tag{4}$$

where P_t^{hh} and P_t^{fh} denote the domestic and import price indexes for the home country. The bundles C_t^{hh} and C_t^{fh} in turn are CES aggregates combining different home- and foreign-produced varieties,

$$C_t^{hh} = \left[\int_j C_t^{hh}(j)^{\frac{\epsilon-1}{\epsilon}} dj \right]^{\frac{\epsilon}{\epsilon-1}}, C_t^{fh} = \left[\int_j C_t^{fh}(j)^{\frac{\epsilon-1}{\epsilon}} dj \right]^{\frac{\epsilon}{\epsilon-1}}, \tag{5}$$

where ϵ is the elasticity of substitution across different varieties produced in the same country.

The associated price indexes are as follows:

$$P_t^{hh} = \left[\int_j P_t^{hh}(j)^{1-\epsilon} dj \right]^{\frac{1}{1-\epsilon}}, P_t^{fh} = \left[\int_j P_t^{fh}(j)^{1-\epsilon} dj \right]^{\frac{1}{1-\epsilon}}. \tag{6}$$

¹¹This assumption is made to ensure that the model has a balanced growth path along which hours worked are stationary, as is the case in the data.

¹²With a representative agent, internal and external habit formulations yield almost identical dynamics. Using micro data, Ravina (2007) argues that the evidence in favor of internal habits is stronger than external habits.

$P_t^{hh}(i)$ and $P_t^{fh}(j)$ denote the prices paid by home consumers for imported varieties i and j , respectively. Markets are assumed to be complete, so that households can trade in a complete set of state-contingent securities in order to smooth consumption fluctuations. While the complete-markets assumption is a strong one, it is used extensively in the literature, and incomplete markets have been shown to generate only minor departures from the complete-markets benchmark (see, for instance, Schmitt-Grohé and Uribe 2003.)

In the presence of complete markets, the household budget constraint is as follows:

$$P_t^{h,cpi} C_t^h + \int_s \mu_{t,t+1}(s) D_{t+1}^h(s) \leq W_t^h N_t^h + D_t^h + T_t^h. \quad (7)$$

D_{t+1} denotes the amount of state-contingent securities purchased by households at price $\mu_{t,t+1}(s)$ which yield one unit of nominal payoff at time $t + 1$ if state s is realized. W_t is the nominal wage, and T_t denotes lump-sum transfers to households. These comprise net transfers from the government as well as dividends from firms and financial intermediaries.

Although as a simplification I model a cashless economy with no explicit mention of money, implicitly there is assumed to be a time-invariant one-to-one relationship between the nominal interest rate and money demand which the central bank can exploit to set the desired nominal interest rate by changing money supply.

As a further simplification, wages are assumed to be flexible and the monetary non-neutrality is induced solely via price stickiness. In a closed-economy setting, Smets and Wouters (2007) show that price stickiness is more important in explaining fluctuations in the U.S. data compared with wage stickiness. Wage stickiness is nevertheless introduced in standard models to provide a “cost-push shock.” In this model, however, the working capital constraints on firms play that role. That said, the main results of the model are robust to the introduction of wage stickiness.¹³

¹³Appendix A extends the model with sticky wages. The main empirical results are unaffected by this extension. It is pertinent to note that the decision to ignore stickiness in wages is made explicitly based on its limited contribution to a model like the one that is being built here. There is strong evidence in favor of wage

The first-order conditions characterizing the household problem are as follows:

$$A_t \lambda_t^h = \left(\frac{(C_t^h - H_t^h)}{A_t} \right)^{-\sigma_c} - \chi \gamma \beta \mathbb{E}_t \left[\frac{A_t}{A_{t+1}} \left(\frac{(C_{t+1}^h - H_{t+1}^h)}{A_{t+1}} \right)^{-\sigma_c} \right] \tag{8}$$

$$(N_t^h)^{\sigma_L} = \lambda_t^h \frac{W_t^h}{P_t^{h, cpi}} \tag{9}$$

$$\beta \mathbb{E}_t \left[\frac{\lambda_{t+1}^h \frac{P_t^{h, cpi}}{P_{t+1}^{h, cpi}}}{\lambda_t^h \frac{P_t^{h, cpi}}{P_{t+1}^{h, cpi}}} \right] = \frac{1}{R_t^h} = \mu_{t, t+1}. \tag{10}$$

λ_t^h is the Lagrange multiplier associated with the budget constraint, which also captures the marginal utility of consumption. Equation (8) is the standard Euler equation with internal habits in consumption. Equation (9) is the labor supply condition which equates the marginal disutility from work to the increase in income, and equation (10) gives the price of state-contingent bonds, which also equals the inverse of the equilibrium gross nominal interest rate. Note that equation (10) uses the assumption that the state-contingent bonds are denominated in the home currency. This is without loss of generality, and the corresponding equation for the foreign country is given by

$$\beta \mathbb{E}_t \left[\frac{\lambda_{t+1}^h \frac{P_t^{f, cpi}}{P_{t+1}^{f, cpi}} \frac{E_t}{E_{t+1}}}{\lambda_t^h \frac{P_t^{f, cpi}}{P_{t+1}^{f, cpi}} \frac{E_t}{E_{t+1}}} \right] = \frac{1}{R_t^f} = \mu_{t, t+1}. \tag{11}$$

E_t denotes the nominal exchange rate, i.e., the price of foreign currency in terms of home currency.¹⁴ Equations (10) and (11) can

stickiness in the form of downward nominal rigidity, and this has first-order implications for open economies—see, for instance, Schmitt-Grohé and Uribe (2011). However, the solution technique used in this paper involves linearization around a deterministic steady state and is neither equipped to deal with large shocks nor with asymmetries like one-sided wage rigidity, so these considerations are beyond the scope of the present paper.

¹⁴Note that as defined here, an increase in the nominal exchange rate corresponds to a depreciation of the home currency.

be used to show that the uncovered interest rate parity condition holds up to a first order.

$$R_t^h = R_t^f \mathbb{E}_t \left(\frac{E_{t+1}}{E_t} \right) \quad (12)$$

3.2 Firms

The production side of the economy is characterized by a continuum of atomistic firms, each of which produces a differentiated product. Labor is the only input in production and the production function of the generic firm is given by¹⁵

$$Y_t^h(j) = A_t A_t^h N_t^h(j). \quad (13)$$

Here A_t is a common worldwide technology component and A_t^h is a country-specific stationary technology shock. Following Christiano, Eichenbaum, and Evans (2005), I assume that firms operate under a working capital constraint and are required to borrow funds at the nominal interest rate to pay a fraction of their wage bill.¹⁶ The cost function of the firm is thus given by

$$\Xi_t^h(j) = R_{L,t}^h W_t^h Y_t^h(j), \quad (14)$$

where $R_{L,t}^h$ is the firm's total interest rate factor. I assume that a fraction u_L^h of the wage bill has to be financed by intraperiod borrowing, which gives the following relationship defining the external financial dependence of goods-producing firms:

$$R_{L,t}^h = (u_L^h R_t^h + 1 - u_L^h). \quad (15)$$

¹⁵The model abstracts from capital mainly for simplicity. This assumption is not uncommon in the New Keynesian literature. Another reason for excluding capital is that the introduction of cost side effects of monetary policy on investment interferes with stability and model indeterminacy as emphasized by Aksoy, Basso, and Martinez (2012).

¹⁶This is a standard channel via which a cost channel for monetary policy can be introduced. See Barth and Ramey (2002) for intra-industry evidence on the cost channel and Ravenna and Walsh (2006) for a theoretical exploration and more empirical evidence.

$u_L^h = 0$ corresponds to the case with no working capital constraints, whereas $u_L^h = 1$ corresponds to the case that is considered in most papers that model the cost channel, including Christiano, Eichenbaum, and Evans (2005) and Ravenna and Walsh (2006).

The market structure is assumed to be monopolistically competitive. Each producer producing a distinct good faces an elasticity of demand ϵ . Prices are assumed to be sticky and pricing contracts are staggered according to the mechanism in Calvo (1983).¹⁷ In each period each firm has the opportunity to reoptimize and set its price with probability $(1 - \theta_h)$. The firms that do not optimize their price are assumed to keep their price unchanged from the previous period.¹⁸ Conditional on having the opportunity to reset its price in period t , firm j would reset its price in order to maximize a discounted value of its lifetime future expected profits conditional on the prices remaining the same. The associated maximization problem is given by

$$P_t^h(j)^* = \mathit{Argmax} \mathbb{E}_t \left[\sum_{k=0}^{\infty} (\theta^h)^k \Omega_{t,t+k} [P_t^h(j)^* Y_{t+k}^h(j) - \Xi_{t+k}^h(j)] \right], \quad (16)$$

where the demand function for each firm is as follows:

$$Y_t^h(j) = \left(\frac{P_t^h(j)^*}{P_t^h} \right)^{-\epsilon} Y_t^h. \quad (17)$$

¹⁷Alternatively, the more realistic quadratic adjustment costs as proposed in Rotemberg (1982) can be assumed. However, the model is solved by considering a first-order approximation around a deterministic steady state, and it can be shown that the dynamics implied by these two mechanisms are identical up to a first-order approximation. In particular, they both lead to the same Phillips curve derived below.

¹⁸Alternatively, one could allow for prices to be indexed to past inflation. As shown by Adolfson et al. (2007) and Smets and Wouters (2007), adding this assumption does not change much in terms of the fit of the model. This is also consistent with the single-equation estimates of Galí, Gertler, and López-Salido (2001).

The first-order conditions associated with this problem yield the following expression for the optimal price conditional on reoptimization:

$$P_t^h(j)^* = \mathbb{E}_t \left[\frac{\sum_{k=0}^{\infty} (\theta^h)^k \Omega_{t,t+k} \left(\frac{\epsilon}{\epsilon-1} \right) P_{t+k}^h MC_{t+k}^h Y_{t+k}^h}{\sum_{k=0}^{\infty} (\theta^h)^k \Omega_{t,t+k} Y_{t+k}^h} \right], \quad (18)$$

where $MC_t^h = \frac{R_L^h W_t^h}{A_t A_t^h P_t^h}$ denotes the real marginal cost facing each firm. The log-linearized version of equation (18) around the symmetric steady state reads¹⁹

$$p_t^h(j)^* = (1 - \beta\theta^h) \sum_{k=0}^{\infty} (\beta\theta^h)^k \mathbb{E}_t(m c_{t+k}^h). \quad (19)$$

This leads to the following forward-looking Phillips curve for PPI inflation:²⁰

$$\pi_t^h = \beta \mathbb{E}_t \pi_{t+1}^h + \frac{(1 - \beta\theta^h)(1 - \theta^h)}{\theta^h} m c_t^h. \quad (20)$$

3.3 Import-Export Sector

In order to introduce a role for trade finance, an import-export sector characterized by the presence of trade firms is explicitly introduced in the model. This international trade sector, which is assumed to be credit constrained, generates a role for trade finance constraints to influence real variables in the economy in addition to incomplete pass-through. In particular, like the domestic firms, the trade firms too are assumed to be credit constrained and are required to borrow to pay for an exogenous (and time-invariant) fraction of their costs. For simplicity, I assume that the trade firms do not employ any labor.

Sequential trade and vertical fragmentation are key features in the trade data that have been successful in explaining many empirical stylized facts.²¹ Following this literature, the import sector is

¹⁹Throughout the paper, lowercase letters are used to denote log-deviations from steady state, i.e., $x_t = \log X_t - \log(\bar{X})$.

²⁰The derivation is standard; see, for instance, Galí (2009).

²¹See, for instance, Huang and Liu (2001, 2007) and Wong and Eng (2013).

assumed to be characterized by a sequence of firms that operate at different stages. Each firm has a production function which transforms the input into output one for one. Each firm however is credit constrained and is required to finance a part of its purchase by borrowing at the risk-free rate. Multiple processing stages in the import sector thus play the role of amplifying the cost effects of monetary policy.

Incorporating these features, the import-export sector is modeled as an n -stage sequential setup. At each stage k , a continuum of atomistic firms operate with the following production technology:

$$Y_{k,t}^{fh}(j) = Y_{k-1,t}^{fh}(j), k \in \{1, 2, \dots, n\}, j \in (0, 1). \quad (21)$$

Note that for simplicity it is assumed that these firms neither employ labor, nor are they subject to productivity shocks as is the case with goods-producing firms. The cost function of each firm is given by

$$\Xi_{k,t}^{fh}(j) = R_t^{fh}(k) P_{k-1,t}^{fh}. \quad (22)$$

Similar to the goods-producing firms, R_t^{fh} is the gross interest factor which characterizes the external finance dependence of the sector. Moreover, in order to allow for incomplete pass-through of exchange rate into import prices, firms at the final stage (n) in the import-export sector are assumed to operate under monopolistic competition like the goods-producing firms. Under these assumptions, the real marginal cost of the import-export sector as a whole can be written as follows:

$$\Phi_t^{fh} = \frac{E_t P_t^f R_t^{fh}}{P_t^{fh}}. \quad (23)$$

Here P_t^{fh} denotes the local currency price of foreign goods that are sold to home consumers. This real marginal cost term can also be interpreted as a law of one price gap. This gap comprises not only incomplete pass-through because of price stickiness but also an additional effect coming from trade finance, which implies that in this model there can be deviations from law of one price even in the absence of market power and flexible prices on the part of the importing firms.

The gross interest rate factor in equation (23) can be written as follows:

$$R_t^{fh} = [u^{fh} R_t^c + (1 - u^{fh})]^n, \quad (24)$$

where n is the number of processing stages and $0 < u^{fh} < 1$ is the fraction of the purchases that have to be financed by external borrowing at each stage.²² R_t^c is the interest rate that is used in trade finance. It would be the home interest rate (R_t^f), the foreign interest rate R_t^f , or a convex combination of the two. While firms are allowed to split their borrowing across domestic and foreign sources, this split is assumed to be time invariant. While this simplifying assumption is potentially restrictive—as it does not allow for optimal choice of funding by firms in response to shocks—the fact that firm’s sources of funding are sticky has been well documented in the literature.²³

Log-linearizing equation (24) yields the following approximate relationship between the number of processing stages, external finance dependence in each stage, and the nominal interest rate:

$$r_t^{fh} \approx n u^{fh} r_t^h. \quad (25)$$

As is evident from equation (25), the impact of changes on nominal interest rate on trade finance depends on both the external finance dependence (u^{fh}) and the number of processing stages (n). The equation also makes it clear however that with this specification it is not possible to identify these two parameters separately in the data. Moreover, the relationship between the risk-free interest rate and the marginal cost of the retail sector may depend on other factors that are not modeled explicitly but may nevertheless play a role. Since the goal of the paper is to study the consequences of this relationship rather than its microfoundations, the model is parameterized in terms of an aggregate parameter (δ^{fh}) which can be understood as the elasticity of marginal cost of import retailers with respect to the risk-free rate, i.e.,

$$r_t^{fh} = \delta^{fh} r_t^h, \quad (26)$$

²²For simplicity, this parameter is assumed to be independent of n as well as t .

²³See, for instance, Degryse et al. (2019), Jiménez et al. (2012), and Khwaja and Mian (2008).

where $\delta^{fh} = f(n, u^{fh}, Z)$ is a function of n , u^{fh} , and other characteristics Z that are not explicitly modeled. Trade finance in the real world (both domestic and international) is operationalized in a number of different ways, including direct lending by banks to the exporter and/or the importer, interfirm trade credit, open account (i.e., post-delivery payment), or cash in advance.²⁴ To the extent that all these mechanisms involve at least one of the parties engaging in borrowing at an interest rate that is directly affected by changes in monetary policy (as captured by equation (26)), it is important to emphasize that even with this parsimonious specification of external finance dependence, the model is general enough to capture all the different trade finance arrangements.

Similar to the case of goods-producing firms, the optimal pricing decisions of the importing firms lead to the following forward-looking Phillips curve for import consumer prices:

$$\pi_t^{fh} = \beta E_t \pi_{t+1}^{fh} + \frac{(1 - \beta\theta^{fh})(1 - \theta^{fh})}{\theta^{fh}} \phi_t^{fh}. \quad (27)$$

As $\theta^{fh} \rightarrow 0$, we have the benchmark case of complete pass-through, with the difference from the standard model being that in addition to exchange rate pass-through, there is also “interest rate pass-through,” a novel channel not considered in the literature so far.

For future reference, the CPI inflation in the home country is given by a weighted sum of π_t^{fh} and π_t^h . In particular,

$$\pi_t^{fh} = (1 - \alpha)\pi_t^h + \alpha\pi_t^{fh}. \quad (28)$$

3.4 Government

There is a government which finances current expenditure by imposing lump-sum taxes on households. For simplicity, I do not allow for government borrowing or lending and all expenditures are financed based on current-period receipts. The government consumption good is assumed to follow the same aggregator as that for the households.

²⁴See Ahn, Amiti, and Weinstein (2011) and Schmidt-Eisenlohr (2013).

The overall government spending process is stochastic and driven by persistent shocks.

$$g_t^h = \rho_g^h g_{t-1}^h + \epsilon_{gt}^h \quad (29)$$

Note that although neither the lump-sum tax nor the assumption of same consumption bundle for households and the government is realistic,²⁵ the sole aim for introducing the government in this model is to have a source for exogenous demand shocks.

3.5 Central Bank

The central bank is assumed to set interest rates according to a modified version of the Taylor rule postulated in Taylor (1993). In particular, I allow for interest rate smoothing and the possibility of nominal exchange rate stabilization in the central bank's reaction function.²⁶

The central bank's reaction function is thus given by

$$i_t^h = \rho_R^h i_{t-1}^h + (1 - \rho_R^h) [\phi_\pi^h \pi_t^h + \phi_y^h \Delta y_t^h + \phi_e^h \Delta e_t] + \epsilon_{rt}^h, \quad (30)$$

where i_t^h denotes the nominal interest rate ($R_t^h = 1 + i_t^h$), Δy_t^h denotes the growth rate of output, and Δe_t denotes the rate of (nominal) depreciation. ϵ_t^h is an idiosyncratic white-noise process to be interpreted as a monetary policy shock.

Finally, the model is closed by imposing the following market clearing condition for each firm in equilibrium:

$$Y_t^h(j) = C_t^{hh}(j) + G_t^{hh}(j) + G_t^{hf}(j) + C_t^{hf}(j) \forall j \in (0, 1). \quad (31)$$

3.6 Terms of Trade and Real Exchange Rate

Terms of trade for a country is defined as the ratio of the price of domestically produced goods at home relative to the price of

²⁵In particular, government consumption is likely to be concentrated towards nontradables and therefore exhibit a higher home bias than households. See Lane (2010) for a discussion of this point.

²⁶The estimation allows for the responses of the central bank to nominal exchange rates to differ across the two countries. Backus et al. (2010) show that this asymmetry can go a long way in explaining the uncovered interest rate parity puzzle.

imported goods.²⁷ In particular, the terms of trade for the home country is defined as follows:

$$tot_t^h = \frac{P_t^h}{P_t^{fh}}. \quad (32)$$

Analogously, terms of trade for the foreign country is defined as

$$tot_t^f = \frac{P_t^f}{P_t^{hf}}. \quad (33)$$

Using equation (23) and its foreign-country counterpart along with (32) and (33) gives

$$\phi_t^{fh} \phi_t^{hf} = tot_t^h tot_t^f R_t^{fh} R_t^{hf}. \quad (34)$$

This equation shows that even under the assumption of perfect competition (so that $\phi_t^{hf} = \phi_t^{fh} = 1$), the home and foreign terms of trade do not equal each other (inversely). In this case, the law of one price gap still exists, but depends only on terms relating to international trade finance.

The real exchange rate (RER) between home and foreign currencies is defined in the standard way by weighting the nominal exchange rate by the ratio of the consumer price indexes in the two countries.

$$S_t = \frac{E_t P^{f,CPI}}{P^{h,CPI}} \quad (35)$$

As with the nominal exchange rate, the real exchange rate is defined in such a way that an increase corresponds to a depreciation of the home currency. Typically in open-economy models, the real exchange rate as defined above is used as a gauge of competitiveness, i.e., a falling RER denotes lower competitiveness of home goods and vice versa. As the next section shows, however, this interpretation of the RER can be flawed in the presence of frictions like trade finance constraints, and the terms of trade is more relevant as a measure of competitiveness.

²⁷Note that typically terms of trade is defined as the ratio of the price of exports to imports. The distinction ceases to matter since most models typically have the feature that export prices are equal to domestic prices. This however is not the case in this model due to imperfect competition as well as trade finance.

3.7 *Equilibrium and Solution Method*

The equilibrium conditions characterized above along with the shock processes comprise a dynamic system with a unique nonstochastic steady state.²⁸ The model is solved by log-linearizing the equilibrium conditions characterizing the model around this nonstochastic steady state. In addition to the monetary policy, productivity, and government spending shocks, the model also features a shock to the labor supply equation and the nominal exchange rate process.

4. Calibration and Model Simulations

This section discusses simulation results based on a calibrated version of the model to outline the dynamics of the key model variables and how they are affected by the presence and degree of trade finance dependence in the wake of exogenous shocks. The model is calibrated to a symmetric two-country case with most parameter values picked from the previous literature—in particular, Lubik and Schorfheide (2006) and Smets and Wouters (2003, 2007)—but the values are kept the same for both home and foreign countries so as to illustrate the mechanics in the model more clearly.²⁹

4.1 *Calibration*

Table 1 shows the values used in the calibration exercise. Although most of the values are standard, there are a couple of parameters that merit further discussion. The intratemporal elasticity of substitution between home and foreign goods is a parameter that, despite extensive empirical research, has failed to yield a consensus, leading to the “elasticity puzzle” (see Ruhl 2008). Typically, the elasticity estimates are found to fall with the level of aggregation, as documented in Disdier and Head (2008) and Hummels (1999). While

²⁸All parameter restrictions required for uniqueness, including the Taylor principle proposed in Woodford and Walsh (2005), are imposed to allow a unique solution. In the estimation, priors are confined to the region so that the posterior distribution also continues to satisfy these constraints.

²⁹These restrictions will be lifted in the empirical section and most parameters will be estimated without imposing these symmetry restrictions.

Table 1. Parameter Calibration for Simulation Exercises

θ^h	0.7	σ_c	2
θ^f	0.7	σ_L	2
θ^{hf}	{0.1,0.7}	h	0.7
θ^{fh}	{0.1,0.7}	η	1
ϕ_π	1.5	α	0.15
ϕ_y	0.5	β	0.99
ϕ_e	0	δ	{0,2}
ρ_R	0.7		

calibrated models typically rely on evidence from the trade literature and pick values greater than 1,³⁰ estimates based on macro data typically yield much lower values, most often less than 1.³¹ Although this paper too finds estimates of elasticity to be small in line with the macro literature, these estimates could be susceptible to the downward aggregation bias discussed in Imbs and Méjean (2012), who show that when elasticities are heterogeneous, aggregation leads to a downward bias. Indeed, the evidence on heterogeneity of elasticities is substantial, as documented in Broda and Weinstein (2006). The value chosen for the simulation results is $\eta = 1$. It is a compromise between the estimate obtained from the micro and macro literatures and is more in line with the latter.³² The main mechanisms highlighted in this paper are not dependent on this choice.

The only asymmetries introduced in the calibration are in the external sectors in the two countries in order to study their interaction with trade finance constraints. The external sectors of the two countries can be asymmetric along several dimensions. Firstly, they could differ in the degree of their external finance dependence, i.e., $\delta^{fh} \neq \delta^{hf}$. As argued above, this implies that the asymmetry is either in the average external finance dependence per stage or in the

³⁰See, for instance, Obstfeld and Rogoff (2005). For micro studies that typically yield values greater than 2, see Broda and Weinstein (2006), Feenstra (1994), and Soderbery (2010).

³¹See, for instance, Justiniano and Preston (2010) and Lubik and Schorfheide (2006).

³²Recently, Drozd, Kolbin, and Nosal (2014) have shown how allowing for dynamic elasticities (i.e., different elasticity in the short versus long run) can help reconcile the business cycle and trade literatures.

number of stages involved in transporting the good from one country to another. For instance, Amiti and Weinstein (2011) find that external finance dependence is much higher for goods shipped by sea than for those shipped by air. Secondly, countries could differ in the degree of their import price pass-throughs, which could be a function of the nature of goods themselves. For instance, Peneva (2009) shows that prices of labor-intensive goods are stickier than those of capital-intensive goods. If countries export goods with substantially different factor intensities, this could lead to an asymmetry in import prices. Lastly, countries can also differ in the interest rate/currency that they are constrained to borrow in. The first two asymmetries are likely to be linked to differences in export bundles of countries. A country exporting high-end luxury products is likely to have lower competitiveness, higher markups, and hence lower price flexibility in its prices than a commodity-exporting country that exports a homogenous product. The third source of asymmetry, the currency denomination of debt, is likely to be an institutional feature that I assume is fixed in the short run.³³ The two parameters governing import price stickiness are varied in the simulations to show how they affect the propagation mechanism of shocks.

In order to determine plausible values for the external finance dependence parameters, I rely on two separate approaches, which yield similar ballpark estimates. Firstly, I consider the model's predication regarding the fall in trade-to-GDP ratios in response to a trade finance shock. Eaton et al. (2011) argue that about 80 percent of the 20–30 percent fall in trade-to-GDP ratio can be accounted for by demand-side effects and heterogeneity in traded versus nontraded goods. This leaves 20 percent of the collapse, or about 4–6 percent fall in trade-to-GDP ratios, unexplained. The first calibration strategy for δ involves matching this response of the trade-to-GDP ratio to an interest rate shock that is simulated in the model. Table 2 shows the peak response of trade-to-GDP ratios under different assumptions on elasticity of substitution and import

³³A large fraction of international trade is conducted in U.S. dollars and hence the dollar is the primary currency not only for settling trade transactions but also in facilitating trade finance. However, local-currency debt in countries like Europe and Japan is also fairly likely—see, for instance, Amiti and Weinstein (2011) and Gopinath, Itskhoki, and Rigobon (2010).

Table 2. Peak Response of Trade-to-GDP Ratio to an Interest Rate Spread Shock of 300 Basis Points

$\eta = 2$	$\delta = 2$	$\delta = 4$
$\theta^{hf} = \theta^{hf} = 0.1$	-10.0366	-23.0467
$\theta^{hf} = \theta^{hf} = 0.7$	-2.8278	-5.8697
$\eta = 0.5$	$\delta = 2$	$\delta = 4$
$\theta^{hf} = \theta^{hf} = 0.1$	-2.5092	-5.7617
$\theta^{hf} = \theta^{hf} = 0.7$	-0.707	-1.467

price flexibility in the model. The size of the shock is 300 basis points, to roughly match the increase in the TED spread during the peak of the 2008 financial crisis. Since there is no consensus on the value of elasticity of substitution (although values closer to and even below 1 are typically preferred by the macro data), a value of δ around 2 seems to be a plausible (if somewhat conservative) value for this parameter. It generates a maximum response of -10 percent, which is on the higher side, but neither this elasticity ($\eta = 2$) nor this pass-through specification seems plausible and is rejected by the data below. Based on the rest of the numbers, it seems to be a conservative estimate, accounting for a decline of trade-to-GDP ratio of less than 3 percent, which is close to but below the 4–6 percent target.

As discussed above, the parameter δ captures not just external financial dependence of sectors but also the number of stages involved in the process from actual production to eventual consumption. The second calibration strategy leverages this interpretation by looking at average propagation lengths (APLs) in the data. The APL between A and B measures the number of stages it takes for the good produced in A to reach B. As an example, consider a world in which global trade comprises an upstream country (say Japan) exporting intermediate goods to a downstream country (say China) which in turn exports them to the consuming country (say the United States). In this simple example, the APL between Japan and the United States would be 2, while the APL between Japan and China would be 1.

More generally, APLs can be computed using input-output tables using the procedure outlined in Dietzenbacher and Romero (2007).

Table 3. Average Propagation Length: Summary Statistics for Benchmark Year 2007

A. Average Propagation Length (APL) Summary Statistics			
Country-Level APL		Country-Sector-Level APL	
No. of Countries	41	No. of Country-Sectors	1,435
Mean APL	2.8465	Mean APL	3.61
Median APL	2.7396	Median APL	3.62
Std. Dev.	0.5	Std. Dev.	0.9
B. APL for Select Country Pairs			
	United States	Germany	China
United States	1.70	2.85	3.65
Germany	2.83	1.62	3.54
China	3.42	3.53	2.48
Source: World Input Output Database (http://www.wiod.org) and author calculations.			

Table 3 displays summary statistics for APLs computed at the country and country-sector level using detailed intercountry input-output data from the World Input-Output Database for the benchmark year 2007.³⁴ While the country-level APLs are likely to be biased downwards since they ignore within-country flows and the heterogeneity is substantial, the values in the range 2 to 5 seem to be reasonable based on these statistics, which are also in line with the range of plausible values obtained using the behavior of trade-to-GDP ratios.

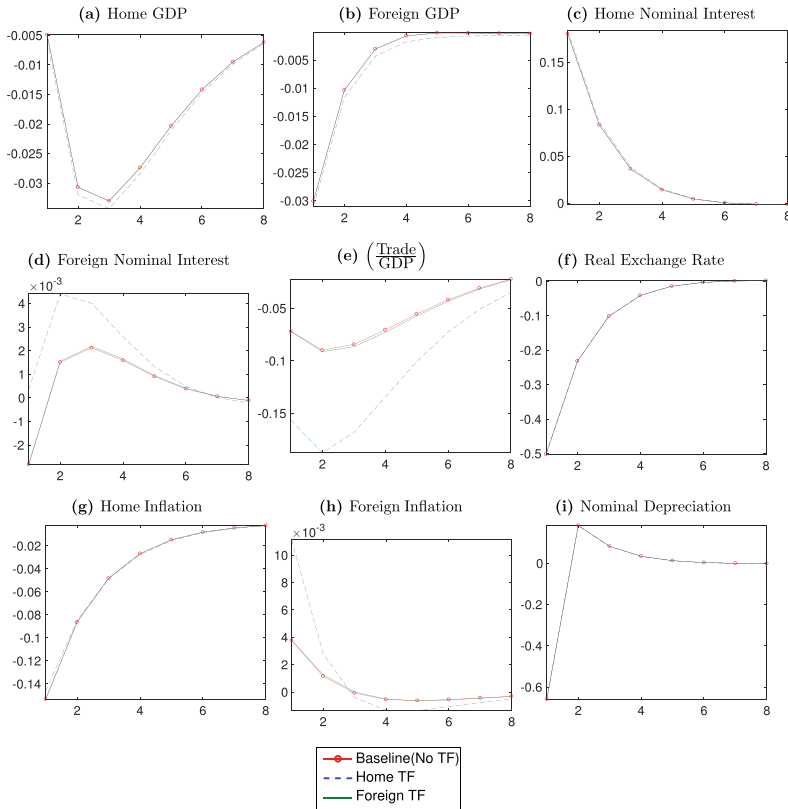
4.2 Model Simulations

Figure 1 shows the impulse response of key macroeconomic variables to a contractionary monetary policy shock (in the form of a 25 basis point increase in the nominal interest rate) in different versions of the model.³⁵ These versions differ only along one dimension—the

³⁴See Timmer and Erumban (2012) for a detailed description of the database and Dietzenbacher and Romero (2007) for a detailed discussion of APL.

³⁵The contractionary monetary shock corresponds to a surprise increase in the nominal interest rate due to a positive shock to $\epsilon_{r_t}^h$ in equation (30). The instantaneous response of the nominal interest rate is less than 25 basis points (the size

Figure 1. Impulse Response to a Home Monetary Contraction: $\theta^{fh} = 0.7, \theta^{hf} = 0.7$



Notes: The impulse responses to a positive 25 basis point shock to the nominal interest rate are computed through simulations using the values in table 1. The horizontal axis measures time in quarters. The vertical axis units are deviations from the unshocked path. Inflation and nominal interest rate are given in annualized percentage points. The other variables are in percentages.

interest rate relevant for trade finance. The blue (dashed) lines correspond to the specification where all borrowing costs related to international trade finance are linked to the home policy rate, whereas the

of the shock), due to the endogenous response of the nominal interest rate to the shock via the interest rate rule (see, for instance, Galí 2009, chapter 3).

green (solid) lines denote the opposite scenario in which borrowing costs are linked to the policy rate of the foreign economy. For comparison, red (solid with dots) lines corresponding to a model without trade finance are also shown.³⁶ The two economies are assumed to be symmetric in all dimensions, including the degree of price stickiness in the import sector ($\theta^{fh} = 0.7, \theta^{hf} = 0.7$).

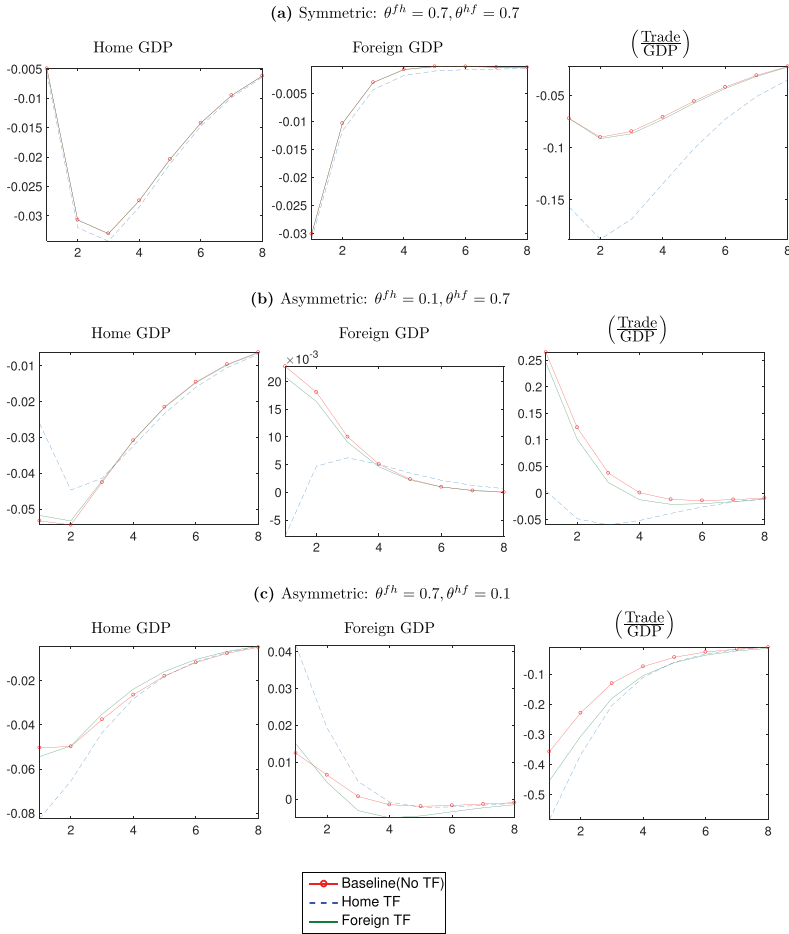
Compared with the model without trade finance, the model in which trade finance is tied to home monetary conditions displays a sharp fall in trade, as captured by the decline in trade-to-GDP ratio (blue dashed line).³⁷ This is a direct consequence of trade becoming more expensive due to a rise in borrowing costs that are linked to the home nominal interest rate. Interestingly, however, the response of both home and foreign GDP is virtually identical under the different models. This reflects the confluence of two effects brought about by the introduction of trade finance which offset one another. As trade becomes more expensive, consumers shift away from imports and towards domestically produced goods. While the former leads to a fall in aggregate demand due to a decline in demand for exports, the latter leads to a rise in aggregate demand due to increased demand for domestically produced goods by consumers in each country as they shift consumption away from imports. On net, these two effects offset each other, such that the impact of trade finance on the response of GDP to monetary shocks remains muted in both the home and the foreign economy.

As shown in figure 2, the symmetry across the two countries is important for this lack of impact of trade finance on GDP. Panel A shows the response of GDP and trade under symmetric price stickiness across countries ($\theta^{fh} = 0.7, \theta^{hf} = 0.7$), the same as in figure 1. Panel B shows how when home import prices are more flexible than foreign prices (a feature that is uncovered in the data in the following section), trade finance begins to significantly alter the impact of home and foreign GDP to home monetary shocks.

³⁶For figures in color, see the online version of the paper, available at <http://www.ijcb.org>.

³⁷The trade-to-GDP ratio is defined as the ratio of total trade divided by total GDP, both measured in nominal terms in a common currency.

Figure 2. Comparison of Impulse Responses to a Home Monetary Contraction under Different Price Stickiness Assumptions



Notes: The impulse responses to a positive 25 basis point shock to the nominal interest rate are computed through simulations using the values in table 1. The horizontal axis measures time in quarters. The vertical axis units are percentage deviations from the unshocked path.

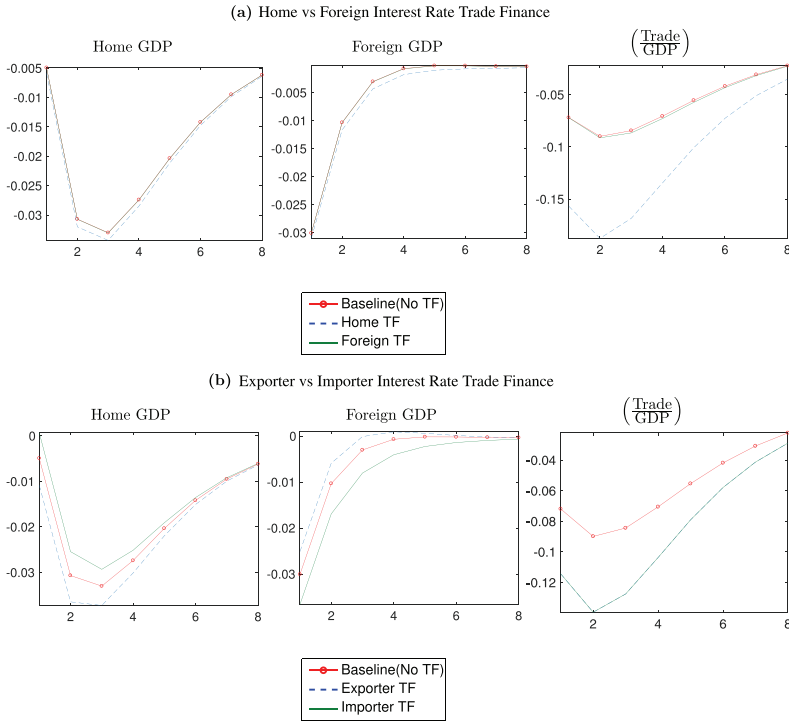
As before, the rise in home interest rates, which govern the borrowing costs for importers worldwide, leads to a sharp rise in the marginal costs of import firms. Since home import prices are more

flexible, home importers pass on this rise to consumers in the form of higher import prices to a larger extent than foreign importers, whose retail prices are stickier. The net result is a sharp fall in the demand for imports in the home economy, which, unlike in the case of symmetric import price stickiness, is not matched by a corresponding fall in demand for home exports coming from the foreign country. Therefore, compared with the baseline model without trade finance, the model with trade finance generates a positive impact on home GDP (which consequently declines by less in response to the monetary shock) and a larger fall in GDP in the foreign economy.

The opposite is true if the price asymmetry is reversed such that the foreign retail price of imports is more flexible than home, as in panel C. In this case, in comparison with the baseline model in response to a home monetary shock, the introduction of trade finance has a positive impact on home GDP and a negative impact on foreign GDP.

Differences in the degree of price stickiness are just one source of asymmetry across countries (albeit an important one for which the following sections provide evidence). In principle, asymmetries in other dimensions can also break the offsetting effects that make trade finance constraints irrelevant as far as GDP is concerned. Panel B in figure 3 shows an example where price stickiness is the same across countries, but they differ in the interest rate that is used to finance international trade (panel A, for reference, is the same as in the previous two figures). The blue lines correspond to the model in which exporters are financially constrained and need to borrow working capital at a cost linked to the risk-free rate of the exporting country. In this case, when home interest rates rise as a consequence of the monetary contraction, foreign imports (home exports) become more expensive for consumers due to the higher borrowing costs of home exporters. This is not offset by a corresponding rise in the price of home imports, since marginal costs of foreign exporters, which are linked to the foreign risk-free rate, do not rise. As a result, compared with the baseline model without trade finance, home GDP falls more, and foreign GDP less, in response to a home monetary contraction. The opposite is true when importers who have working capital requirements that need to be financed at borrowing costs linked to their domestic risk-free rate (green lines).

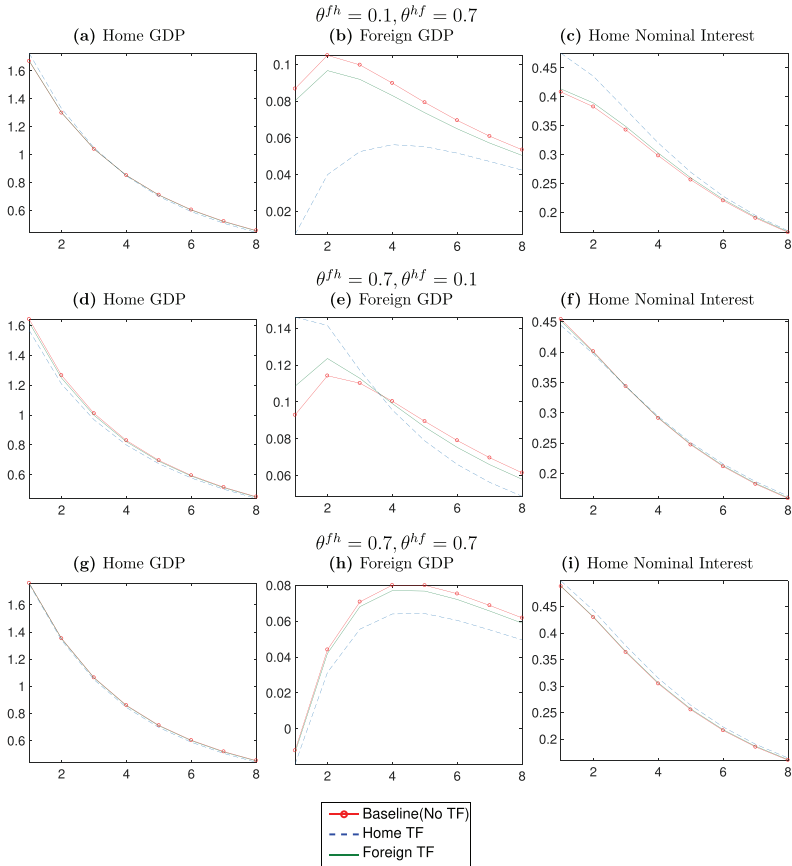
Figure 3. Comparison of Impulse Responses to a Home Monetary Contraction under Different Financing Arrangements



Notes: The impulse responses to a positive 25 basis point shock to the nominal interest rate are computed through simulations using the values in table 1. The price stickiness parameters are fixed at $\theta^{fh} = 0.7, \theta^{hf} = 0.7$ across all sets of simulations reported in this figure. The horizontal axis measures time in quarters. The vertical axis units are percentage deviations from the unshocked path.

To summarize, the main insight from simulation results is that when countries are completely symmetric in terms of their price stickiness and trade financing needs, the introduction of trade finance matters only as far as trade prices and total trade volumes are concerned, but offsetting effects imply that its impact on the response of home and foreign GDP is limited. On the other hand, when countries are asymmetric along any of these dimensions, trade finance exerts a significant influence on the response of home and foreign GDP to shocks.

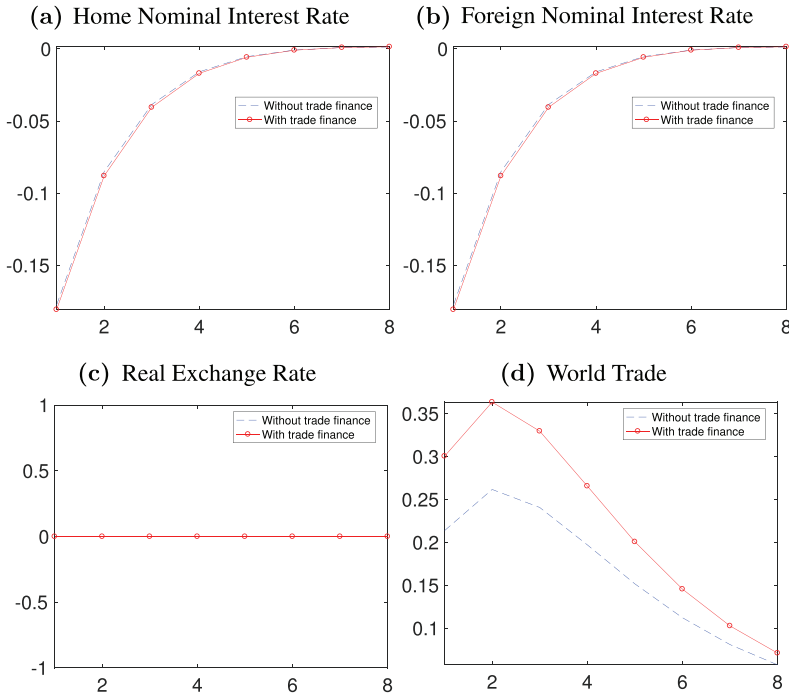
Figure 4. Home Government Spending Shock



Notes: Impulse response to a positive government spending shock in the home country. The vertical axis units are deviations from the unshocked path. Inflation and nominal interest rate are given in annualized percentage points. The other variables are in percentages.

It is important to emphasize that while the impact is most visible in the case of monetary shocks, to the extent that most other shocks generate an endogenous response of the risk-free rate in the economy, the impact of trade finance extends to all other business cycle shocks. Figure 4, for instance, illustrates this for a positive home government spending (demand) shock.

Figure 5. Competitive Devaluations and Trade



Notes: The figure shows impulse responses to a simultaneous interest rate cut by both central banks. The impulse responses are computed through simulations using the values in table 1 ($\theta^{fh} = 0.7, \theta^{hf} = 0.7$). The horizontal axis measures time in quarters. The vertical axis units are deviations from the unshocked path. Inflation and nominal interest rate are given in annualized percentage points. The other variables are in percentages.

4.3 Application: Impact of Competitive Devaluations on Trade

This section discusses the implications of trade finance for competitive devaluations. Figure 5 considers a competitive devaluation scenario in which both the home and the foreign central bank engage in simultaneous interest rate cuts of equal magnitude—a “competitive devaluation” or “currency war” scenario. Panels A and B indicate this action. As is evident from panel C, the actions of the two central banks cancel each other as far as the impact on the exchange rate is

concerned and lead to no change in the real (or nominal) exchange rate. However, as shown in panel D, in a world where trade finance is important, the boost to international trade volumes is much more pronounced. While trade rises in both cases due to the increased aggregate demand in each country which also spills over into demand for imports, the fact that the financing constraints on importers and exporters are loosened in a world in which trade finance is present leads to the much sharper boost in trade. This finding is particularly relevant in an environment where trade conflicts are depressing the outlook for trade. These results show that to the extent that international trade policymakers care about trade over and above its impact on contemporaneous output,³⁸ even if devaluations are matched competitively by trade partners, the boost to exports can be substantially larger than that inferred from models that do not incorporate a role for trade finance.

5. Estimation

As is evident in section 4.2, the role of trade finance in business cycle fluctuations depends critically on parameters characterizing the export-import sectors of countries, and in particular on differences across the two countries. This section uses Bayesian techniques to estimate the model using macroeconomic time-series data from two large open economies—the United States and euro zone.³⁹ Following Smets and Wouters (2003) and others, a full-information likelihood-based estimation procedure is used.

5.1 Data

The model is matched to the data by treating the United States and euro zone as the two countries comprising the world economy. The sample period is 1983:Q1–2007:Q4.⁴⁰ Table 4 lists the variables used

³⁸One reason why this may be so is because there is an extensive literature documenting the productivity gains from trade—see, for instance, De Loecker (2013).

³⁹See appendix B for a brief description of Bayesian estimation and the model comparison exercise.

⁴⁰Since the subsequent period has been characterized by zero and negative interest rates, the monetary policy stance is not well captured by the policy rate.

Table 4. Observables and Data Sources

Interest Rates R^{US} R^{EU}	Effective Federal Funds Rate Euro-Area Nominal Interest Rate
Prices $\pi^{US,CPI}$ $\pi^{US,GDP}$ $\pi^{EU,CPI}$ $\pi^{EU,GDP}$	CPI Inflation, US GDP Deflator Inflation, US CPI Inflation, EU GDP Deflator Inflation, EU
Exchange Rate $\% \Delta E$	Nominal Depreciation Rate of U.S. Dollar against Euro ^a
Output ΔY^{US} ΔY^{EU}	GDP Growth Rate, US GDP Growth Rate, EU
^a Before 2000, a GDP-weighted exchange rate is used from Lubik and Schorfheide (2006)'s publicly available database.	

as observables in the estimation (a more detailed description along with data sources can be found in appendix E). These comprise short-term nominal interest rates, the euro-dollar nominal exchange rate, GDP growth rates, and various inflation rates for the two countries.⁴¹ Compared with previous studies like Lubik and Schorfheide (2006) that have used only one measure of prices (namely the CPI inflation), both CPI- and GDP-deflator-based inflation measures are used. This is done in order to make the likelihood of the model more informative regarding the new features and parameters introduced in the model, and to sharpen the identification of domestic price stickiness parameters (θ^h and θ^f) on the one hand, and import price stickiness parameters (θ^{hf} and θ^{fh}) on the other. The U.S. data are taken from the Bureau of Economic Analysis, and the European data are taken from the European Central Bank's Area Wide Model (AWM) database. Prior to estimation, all the data are seasonally adjusted.

⁴¹Robustness checks also use bilateral trade as well as import price data, and the results are qualitatively similar.

Table 5. Classification of 11 Shocks Used in Benchmark Estimation

U.S. Shocks	Monetary Policy, Productivity, Government Spending, Labor Supply
EU Shocks	Monetary Policy, Productivity, Government Spending, Labor Supply
Common/Global Shocks	Productivity, UIP, Trade Finance

5.2 Shocks

The benchmark estimation allows for 11 shocks. As shown in table 5, the shocks can be classified into three broad categories: U.S. shocks, euro-area shocks, and common or global shocks.⁴²

5.3 Priors

The first five columns of table 6 describe the priors used in the estimation prices. Most of the priors are based on priors and estimates from Lubik and Schorfheide (2006) and Smets and Wouters (2003, 2007). There are two parameters that quantify trade finance dependence which are new in the paper (δ^{hf} and δ^{fh}). The prior mean for these is set equal to 2, based on the calibration exercises in section 4. A fairly high standard deviation is allowed in the prior in order to reflect parameter uncertainty. For the elasticity of substitution (η), a prior of 1 is assumed as a compromise between the macro and micro evidence regarding the magnitude of this parameter as argued before.

5.4 Estimation Results

5.4.1 Parameter Estimates and Model Comparison

Tables 6 and 7 summarize the prior and posterior distribution of the estimated parameters for the model in which all trade is financed by

⁴²The depreciation shock (also labeled “UIP shock”) is common in the literature and is needed to match the dynamics of the nominal exchange rate, which are not explained well by this class of models. This is a standard limitation of models of this type (see, for instance, De Walque, Smets, and Wouters 2005 and Lubik and Schorfheide 2006).

Table 6. Summary of Prior and Posterior Prior and Posterior Distribution of Estimated Parameters

Parameter	Description	Prior			Posterior		
		Distribution	Mean	Std. Dev.	Mean	90% C.I.	
θ^{US}	Calvo Domestic	Beta	0.7	0.05	0.837	0.8	0.874
θ^{US} Import	Calvo Import	Beta	0.5	0.1	0.377	0.229	0.518
θ^{EU} Import	Calvo Import	Beta	0.5	0.1	0.872	0.726	0.986
θ^{EU}	Calvo Domestic	Beta	0.7	0.05	0.75	0.695	0.807
σ_c	Intertemporal Consumption Elasticity	Gamma	1	0.25	4.512	3.309	5.751
σ_L	Labor Supply Elasticity	Gamma	2	0.5	1.541	0.966	2.092
h	Habit Parameter	Beta	0.5	0.1	0.547	0.395	0.697
η	Intratemporal Elasticity	Gamma	1	0.3	0.408	0.25	0.558
ϕ_π^{US}	Taylor-Rule Parameter	Gamma	1.5	0.25	1.926	1.591	2.232
ϕ_y^{US}	Taylor-Rule Parameter	Gamma	0.5	0.25	0.452	0.206	0.68
ϕ_e^{US}	Taylor-Rule Parameter	Gamma	0.1	0.05	0.031	0.01	0.051
ϕ_π^{EU}	Taylor-Rule Parameter	Gamma	1.5	0.25	1.862	1.524	2.219
ϕ_y^{EU}	Taylor-Rule Parameter	Gamma	0.5	0.25	0.546	0.246	0.845
ϕ_e^{EU}	Taylor-Rule Parameter	Gamma	0.1	0.05	0.03	0.008	0.05
ρ_A^{US}	U.S. TFP Persistence	Beta	0.8	0.1	0.996	0.992	0.999
ρ_R^{US}	U.S. Interest Rate Smoothing	Beta	0.5	0.2	0.821	0.789	0.856
ρ_G^{US}	U.S. Government Spending Persistence	Beta	0.8	0.1	0.963	0.941	0.985
ρ_A^{EU}	EU TFP Persistence	Beta	0.6	0.2	0.574	0.259	0.906
ρ_R^{EU}	EU Interest Rate Smoothing	Beta	0.5	0.2	0.867	0.843	0.892
ρ_G^{EU}	EU Government Spending Persistence	Beta	0.8	0.1	0.93	0.891	0.971
ρ_Z	Global Productivity Persistence	Beta	0.66	0.15	0.461	0.258	0.661
$\delta^{EU \rightarrow US}$	Trade Finance Parameter: US	Gamma	2	0.75	2.27	0.991	3.423
$\delta^{US \rightarrow EU}$	Trade Finance Parameter: US	Gamma	2	0.75	1.837	0.735	2.909
ρ_N^{US}	U.S. Labor Supply Shock Persistence	Beta	0.85	0.1	0.81	0.743	0.878
ρ_N^{EU}	EU Labor Supply Shock Persistence	Beta	0.85	0.1	0.894	0.849	0.939

Notes: The results are based on 200,000 MCMC draws (split across two chains) after burn-in with the posterior mode used as the starting value for each parameter. The list of observables is given in table 4.

Table 7. Summary of Priors and Posterior Distributions of Standard Deviations of Shocks

Shock	Prior Distribution	Mean	Std. Dev.	Posterior Mean	90% C.I.	
A^h	Invg.	1.253	0.655	1.167	0.873	1.463
G^h	Invg.	1.253	0.655	0.526	0.451	0.6
R^h	Invg.	0.501	0.262	0.161	0.139	0.183
A^f	Invg.	0.501	0.262	0.464	0.224	0.707
G^f	Invg.	1.253	0.655	0.502	0.432	0.569
R^f	Invg.	0.251	0.131	0.138	0.12	0.156
Z	Invg.	0.627	0.328	0.337	0.236	0.434
ΔE	Invg.	4.387	2.293	4.166	3.673	4.643
N^h	Invg.	0.101	0.262	1.563	1.355	1.755
N^f	Invg.	2	0.5	2.608	1.722	3.472

Notes: “Invg.” denotes the inverse gamma distribution. The last two rows correspond to measurement errors of the corresponding observed variables. h denotes the home country (United States) and f denotes the foreign country (European Union).

borrowing at the U.S. interest rate. (This is the model that is most preferred by the data, i.e., has the highest Bayes factor, as will be discussed later.)

The posterior estimates of the price stickiness parameters imply that the data support a model in which there is asymmetry in the pass-through into import prices across the two countries. While the pass-through into EU import prices is quite low ($\theta^{\text{EU Import}}$ has a posterior mean of 0.87), the corresponding value for the United States is fairly high (posterior mean of $\theta^{\text{US Import}}$ is 0.38).

Given the importance of the import price stickiness parameters in driving the results of the model, table 8 reports additional robustness checks on the estimated values. It imposes priors that imply exactly the opposite price stickiness pattern to the one estimated in the data (table 6), and finds that the results are robust to this change in the priors.

These results are in line with estimates from Lubik and Schorfheide (2006), who also find evidence in favor of this asymmetry. Table 9 shows a comparison of the posterior means for the Calvo parameters from table 6. In their case this difference may also

Table 8. Alternate Priors for Import Price Stickiness

Parameter	Description	Prior			Posterior	
		Distribution	Mean	Std. Dev.	Mean	90% C.I.
<i>Benchmark Estimation (Table 6)</i>						
θ_{US} Import	Calvo Import	Beta	0.5	0.1	0.377	0.229
θ_{EU} Import	Calvo Import	Beta	0.5	0.1	0.872	0.726
<i>Estimation Results under Alternate Priors</i>						
θ_{US} Import	Calvo Import	Beta	0.9	0.15	0.336	0.160
θ_{EU} Import	Calvo Import	Beta	0.1	0.15	0.963	0.927

Notes: The results are based on 200,000 MCMC draws (split across two chains) after burn-in with the posterior mode used as the starting value for each parameter. The list of observables is given in table 4.

Table 9. Comparison of Calvo Parameters with Lubik and Schorfheide (2006)

	Posterior Mean	Lubik and Schorfheide (2006)		
		Posterior Mean	90% C.I.	Prior Month
θ^{US}	0.83	0.62	[0.49, 0.77]	0.5
$\theta^{\text{US Import}}$	0.38	0.45	[0.17, 0.72]	0.5
$\theta^{\text{EU Import}}$	0.87	0.9	[0.82, 1.00]	0.75
θ^{EU}	0.75	0.61	[0.43, 0.81]	0.75

be partly driven by the choice of their prior distribution, which is asymmetric and implies higher price flexibility in the United States than in the EU for both domestic and import prices.⁴³ This paper on the other hand does not impose this asymmetry *ex ante*.

Notwithstanding the fact that the estimates of the price stickiness parameters are in line with the estimates of Lubik and Schorfheide (2006), at first sight they seem to be at odds with the extensive literature on pass-through into import prices which has found the pass-through (in particular, with regard to the nominal exchange rate) into U.S. import prices to be low, pointing to a very low import price flexibility for the United States.⁴⁴ Although a thorough exploration of this apparent discrepancy would require detailed examination of micro data and is beyond the scope of this paper, two possible explanations can be conjectured. Firstly, while the trade literature has focused for the most part on exchange rate pass-through, the asymmetry revealed here is with regard to pass-through of marginal costs into prices more generally, including other components of marginal costs apart from the nominal exchange rate. Secondly, while the trade literature has focused on import prices at the dock, the estimates in the model correspond to the retail price of imports. Understanding the journey of imports from the dock to eventual retail outlets, including the characteristics of the different

⁴³They rely on Angeloni et al. (2006) and Bills and Klenow (2004) to impose a high prior mean for Europe and a lower one for the United States.

⁴⁴See, for instance, Gopinath, Itskhoki, and Rigobon (2010).

markets and intermediaries involved, would be an important part of interpreting these findings.

With regard to δ , the other parameter which governs the strength of the trade finance channel, table 6 shows that the posterior means are 2.27 and 1.87 for $\delta^{EU \rightarrow US}$ and $\delta^{US \rightarrow EU}$, respectively. These are broadly in line with the calibrated values used in section 4.⁴⁵

In terms of the reduced-form interpretation of equation (25), this estimate implies that the elasticity of the trade finance rate with respect to the risk-free rate is around 2, implying that a 1 percentage point increase in the risk-free rate leads to about a 2 percentage point increase in the total cost of trade finance. While the magnitudes are again roughly consistent with the average propagation lengths estimates in the data, the 90 percent interval includes values high enough to imply an inferred external finance dependence greater than 1. This suggests that there remains scope for alternative microfoundations of the parameter δ .

Table 10 reports the log marginal density for various specifications of the model that are estimated, along with the Bayes factor for each model in comparison with the model without trade finance. Assuming the prior probabilities to be the same across models, numbers in each column (i.e., estimates based on the same number of observables) can be interpreted as measures of the posterior odds ratios, with higher numbers (lower absolute values) indicating higher posterior odds for the corresponding model.⁴⁶ The last column reports Bayes factors computed with respect to the baseline model with no trade finance, which by construction has a Bayes factor of 1 with respect to itself. Bayes factors greater than 1 indicate that the respective model is more preferred by the data than the baseline model. According to Jeffreys (1998), a Bayes factor greater than 30 is “very strong” and a Bayes factor greater than 20 is “decisive” evidence.

The table shows that the models with trade financing with U.S. interest rates and importer interest rates carry the highest posterior probability and Bayes factors. The first of these is not surprising,

⁴⁵Given the importance of the δ and θ parameters, appendix C presents some additional robustness checks on the estimates.

⁴⁶Note that this comparison is valid as long as the prior is proper, which is the case throughout this paper.

Table 10. Marginal Likelihood for Different Models

Model	Marginal Data Density	Bayes Factor wrt No Trade Finance
1 No Trade Finance	-1236.04	1
2 Trade Finance: Both Interest Rates	-1233.71	10
3 U.S. Interest Rate Trade Finance	-1227.37	5,825
4 EU Interest Rate Trade Finance	-1236.15	0.9
5 Importer Interest Rate Trade Finance	-1227.42	5,541
6 Exporter Interest Rate Trade Finance	-1232.34	40

Note: The second model, “Trade Finance: Both Interest Rates,” allows for trade finance to be dependent on both home and foreign interest rates.

given the central role that U.S. monetary policy plays in the global economy and given the fact that the dollar is also the primary vehicle currency in which international trade is conducted.⁴⁷ The higher posterior marginal data density of the model with importer interest rate trade finance on the other hand is less easier to motivate, given that the majority of the empirical literature in trade finance has documented the link between exporter monetary policy and volume of exports. However, a close examination of the arguments given for these apply equally to the link between imports and interest rates as well. In fact, in his empirical analysis Schmidt-Eisenlohr (2013) finds the role of importer interest rate to be as important as the exporter one.

What these results indicate in conjunction is that in the data, the trade finance channel seems to be governed by the interaction of U.S. interest rates with U.S. imports. Since European imports play a limited role due to their low price flexibility, the models with U.S. interest rate and importer interest rate financing both seem to be consistent, and the data are not clearly able to distinguish between the two.

⁴⁷For evidence regarding the latter, see Goldberg and Tille (2008).

5.4.2 Variance Decompositions⁴⁸

The simulation results in section 4 show that incorporation of trade finance has a disproportionately large effect on the spillover effects of shocks, as opposed to the effect on the domestic economy. This raises a question of how much the proposed model fundamentally changes our understanding of the importance of U.S. shocks for euro-area business cycles. Table 11 provides a comparison of the variance decomposition of euro-area variables at two horizons for two categories of shocks according to the classification in table 5—domestic U.S. shocks and all external shocks which include U.S. shocks and the global shocks in table 5. The numbers in the table denote the difference between the share (out of 1) of the variance of the row variable explained by the model with trade finance and the one without. Positive numbers therefore convey that U.S. and external shocks are more important in the model with trade finance for the particular variable, whereas negative numbers indicate the opposite.⁴⁹

The differences in the long-run variance decomposition (horizon = ∞) are unanimous in suggesting that ignoring trade finance overstates the importance of U.S. shocks as well as external shocks more generally for the euro area. Indeed, the differences are negative for all the euro-area variables. The results are less clear-cut in the case of short-run variance decompositions (horizon = one quarter), where the model with trade finance does give a higher weight to external shocks for most euro-area quantity variables such as output, consumption, and imports, but not to price variables such as inflation and the real exchange rate.

To summarize, the model with trade finance suggests that while external shocks are more important in explaining the short-run variance of some variables such as output and consumption, overall, external shocks including U.S. domestic shocks are much less

⁴⁸Appendix D provides additional comparisons between the model with and without trade finance based on posterior predictive moments.

⁴⁹For example, the number corresponding to the row “Imported Inflation” and column “Domestic U.S. Shocks” under the “Horizon = ∞ ” panel shows that for euro-area imported inflation, domestic U.S. shocks account for 0.46 (out of 1) more variance according to the model without trade finance, compared with the one with trade finance, i.e., for example, if in the model without trade finance, U.S. shocks account for 0.8 (or 80 percent) of the variance of euro-area import inflation, they account for only 0.34 (or 34 percent) of the variance in the model with trade finance.

Table 11. Variance Decomposition for Euro Area: Difference between Models with and without Trade Finance

Variable (Euro Area)	Domestic U.S. Shocks	All External Shocks	Domestic U.S. Shocks	All External Shocks
	Horizon = ∞		Horizon = 1	
GDP	-0.001	-0.001	0.00	0.114
Inflation	-0.006	-0.478	-0.003	-0.161
Imported Inflation	-0.009	-0.416	-0.008	-0.12
Consumption	-0.006	-0.006	-0.019	0.408
Nominal Interest Rate	-0.037	-0.046	-0.004	-0.004
Imports	-0.459	-0.456	-0.562	0.141
Exports	-0.017	-0.018	-0.115	-0.009
Terms of Trade	-0.1	-0.1	-0.196	-0.19
Real Exchange Rate	-0.11	-0.114	-0.232	-0.36
Nominal Depreciation	-0.013	-0.11	-0.016	-0.095

Notes: The numbers in the table denote the difference between the share of the variance of the row variable explained by the model with trade finance and the one without. Positive numbers therefore convey that the model with trade finance explains a higher share of the variance of the variable, whereas negative numbers indicate the opposite.

important for euro-area business cycles. This is an important take-away for policymakers, as the role of external versus domestic shocks in driving business cycle fluctuations has different implications for their policy frameworks.

5.4.3 Comparison to an Atheoretical Benchmark: DSGE-VAR

This section uses the DSGE-VAR approach to assess the fit of the model with respect to an atheoretical benchmark—an unrestricted VAR.⁵⁰

The approach exploits the fact that estimating a DSGE model is akin to estimating a VAR with cross-equation restrictions, and allows for the extent to which the restrictions can be imposed or relaxed. In particular, there is a hyperparameter $\lambda \geq 0$ such that the DSGE model restrictions are strictly imposed if $\lambda = \infty$, whereas the restrictions are completely ignored if $\lambda = 0$. Estimation of the VAR uses a prior that is centered at the DSGE-model-implied restrictions. The hyperparameter λ scales the covariance matrix of the prior. If it is large, most of the variance is centered around the DSGE model. The prior is combined with the likelihood to obtain the posterior of λ .⁵¹

Table 12 summarizes the estimation results for λ . The fact that the posterior of λ shifts towards zero compared with the prior indicates that some of the restrictions in the model are at odds with the data. This is in line with the results obtained in Lubik and Schorfheide (2006) and Smets and Wouters (2003), and highlights that estimated DSGE models are typically worse than some VAR specifications. That said, the posterior of λ , as well as the marginal likelihood, is higher in the case of the model with trade finance than without. This suggests that even when evaluated against an atheoretical benchmark like an unconditional VAR, the model with trade finance continues to provide a better fit to the data, and the data relaxes less of its cross-equation restrictions compared with the model that lacks a role for trade finance.

⁵⁰For a detailed description of the procedure see, for instance, Del Negro et al. (2007) and Del Negro and Schorfheide (2004, 2006).

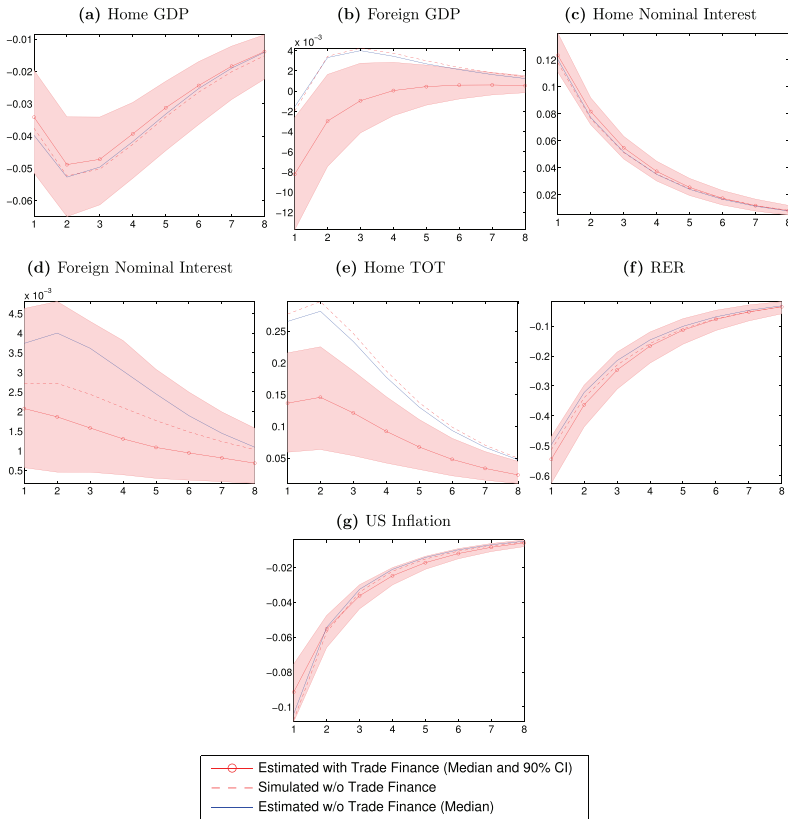
⁵¹The parameters of the DSGE model are also estimated simultaneously in the procedure by projecting the parameters of the estimated VAR back onto the DSGE parameter space.

Table 12. Comparison of the Fit of the Estimated DSGE Model with Unrestricted VAR

Prior and Posterior for DSGE-VAR Hyperparameter λ				
Model	Prior (Mean = 1)	Posterior		Marginal Likelihood
		Mean	90% C.I.	
Without Trade Finance	Uniform [0,2]	0.589	[0.560, 0.615]	-1,799.3
With Trade Finance	Uniform [0,2]	0.661	[0.619, 0.706]	-1,732.9

Notes: The table summarizes the prior and the posterior means of the DSGE-VAR hyperparameter λ which quantifies the extent to which the restrictions of the DSGE model are rejected by the data. The model is estimated with four lags using the nine observables in table 4. See Del Negro and Schorfheide (2006) for a description of the DSGE-VAR procedure.

Figure 6. U.S. Monetary Contraction



Note: Baseline model (dotted line) assumes U.S. interest rate trade finance.

5.4.4 Comparison of Shock Propagation Mechanism across Estimated Versions of the Model

This section illustrates the differences in propagation mechanisms using estimated impulse responses from the model. Figure 6 shows the impulse response of a one-standard-deviation U.S. monetary contraction (median and 90 percent confidence bands) using the estimated model with U.S. trade finance (the model with the higher posterior probability than the standard model). For comparison, the figure also shows two impulse responses corresponding to the standard model. One of these (labeled “Estimated w/o Trade

Finance (Median)”) corresponds to the estimated model without trade finance constraints, and the second (labeled “Simulated w/o Trade Finance”) corresponds to the impulse response from the simulated model with all parameters at the posterior mean from the model with trade finance constraints except the trade finance dependence parameters themselves, which are set to zero. These are two alternate ways of comparing the results with the estimated model with trade finance. Qualitatively, the results in figure 6 are broadly in line with the simulation results. Quantitatively, the figure shows that while the models generate similar predications for the response of domestic GDP, they differ appreciably in the response of foreign GDP and terms of trade.

One implication of this is that for a large open economy like the United States whose business cycle fluctuations are mostly driven by domestic shocks, excluding trade finance from models might be an innocuous omission. On the other hand, if the object of interest is to study spillover effects from foreign shocks (as would typically be the case for a small open economy), ignoring trade finance constraints can lead to severe misrepresentation of the important transmission channels in the model. This is due to the fact that trade finance exerts its influence on shock propagation by affecting terms of trade, which translates into changes in trade volumes. As far as the domestic economy is concerned, it is therefore best seen as an additional channel, while the main effects of the shock are likely to come from the direct domestic impact of shock. On the other hand, as far as the foreign economy and spillover effects are concerned, the entire effect of the domestic shock is transmitted through the external sector, which in turn is affected by trade finance. As a result, incorporation of trade finance matters more for spillover effects of shocks as opposed to domestic effects.

6. Conclusion

An extensive literature in international trade has documented the heavy reliance of international trade flows on external finance and has shown that external financing matters more for international trade as opposed to intranational trade. This paper assesses how this feature affects aggregate business cycle fluctuations and the transmission mechanism of monetary policy. It does so by modeling the

link between trade finance and the cost channel of monetary policy in a two-country New Keynesian model. Unlike the domestic component of the cost channel of monetary policy which has been studied extensively in the literature, the paper shows that the cost channel when combined with trade finance has much richer implications for business cycles, both qualitatively and quantitatively. More specifically, it shows that when external sectors are symmetric across countries, trade finance constraints lead to sharp movements in trade prices and volumes, but do not significantly alter the response of GDP to shocks in either the home country or abroad, due to offsetting effects. On the other hand, if external sectors are asymmetric, trade finance constraints significantly change the response of GDP to both monetary and nonmonetary shocks. The paper identifies various sources of such asymmetry (including differences in import price flexibility) and studies their implications. The parameter estimates provide compelling evidence for asymmetry in import price flexibility across the two countries. In particular, U.S. retail import prices are found to be more flexible than their European counterpart.

Using Bayesian techniques, the paper estimates a two-country DSGE model with macroeconomic time-series data from the United States and the euro zone, two regions which share one of the largest bilateral trade relationships in the world. Based on model comparison exercises, models that appropriately incorporate trade finance constraints are shown to be preferred by the data. Furthermore, trade finance is found to have a larger impact on spillover effects of shocks rather than the effects on the country of origin. For example, euro-zone output contracts sharply in response to U.S. monetary contractions once the model is allowed to appropriately account for the trade finance channel, while the impact is indistinguishable from zero in the model without trade finance. In addition, appropriately accounting for trade finance in the model significantly reduces the importance of external (including domestic U.S.) shocks for euro-area business cycles. This is an important takeaway for policymakers, as the role of external versus domestic shocks in driving business cycle fluctuations has different implications for their policy frameworks. These results also carry important implications for the theory of competitive devaluations. By relaxing financing constraints on exporters and importers worldwide, devaluations are likely to boost trade volumes even more strongly than inferred by standard models.

An important limitation of this paper that future research warrants addressing is with respect to the rigidity in modeling the choice of trade finance by firms. Firms in the model are not allowed to switch between sources of funding in response to shocks. While this assumption is well grounded in the large body of empirical work documenting the stickiness in firms' funding sources (Degryse et al. 2019, Jiménez et al. 2012, and Khwaja and Mian 2008), and may be justifiable for the sample period considered in the paper which coincides with the great moderation period when the shocks hitting the economy were not too large, an extension of the model to endogenize firms' funding choices is bound to provide a more comprehensive understanding of the role of trade finance for business cycle fluctuations.⁵² If firms could switch instantaneously, completely, and costlessly from one funding source to another, then the impact of trade finance on business cycles could be mitigated. The mitigation would be particularly pronounced, and may even lead to a rise in trade, if the interest rates of the two countries move in opposite directions in response to a shock, since in that case exporters and importers would benefit from a decrease in the cost of financing by switching to the lower interest rate. The fact that the data overwhelmingly find a role for trade finance is indicative that while firms may switch between funding sources, it may not be reasonable to assume that this switch is instantaneous and costless.

Appendix A. Model with Sticky Wages

The household problem is to maximize utility given by

$$\max \sum_{j=0}^{\infty} (\beta \theta_w^h)^j E_t(U_{t+j}(C_{t+j}, H_{t+j}, N_{t+j}(h))) \quad (\text{A.1})$$

subject to the per-period budget constraint given by

$$P_t^{,cpi} C_t + \int_s \mu_{t,t+1}(s) D_{t+1}(s) \leq W_t N_t + D_t + T_t \quad (\text{A.2})$$

⁵²Indeed, recent empirical work looking specifically at exporting firms has found that many of them do in fact switch sources of funding in response to shocks—see, for instance, Antràs and Foley (2015), Demir and Javorcik (2018) and Garcia-Marin, Justel, and Schmidt-Eisenlohr (2019).

and the labor demand schedule given by

$$N_t(j) = \left(\frac{W_t(j)}{W_t^h} \right)^{-\eta} N_t \forall t. \tag{A.3}$$

Here $(1 - \theta_w)$ denotes the time-invariant probability of re-adjusting wages in a given period.

The first-order condition implies the following expression for the wage negotiated by households who optimize in a given period:

$$W_t^* = \frac{\sum_j (\beta \theta_w)^j E_t (N_{t+j}(h) U_N(t+j))}{\sum_j (\beta \theta_w)^j E_t \left(N_{t+j}(h) U_C(t+j) \left(\frac{\eta-1}{\eta} \right) \frac{1}{p_{c,t+j}} \right)}, \tag{A.4}$$

which linearizes to

$$\hat{w}_t^* = (\beta \theta_w) E_t (\hat{w}_{t+1}^*) + (1 - \beta \theta_w) \left(\hat{U}_N(t) - \hat{U}_c(t) + \hat{p}_c(t) \right). \tag{A.5}$$

The aggregate wage evolves according to the following equation:

$$\hat{w}_t = (1 - \theta_w) \hat{w}_t^* + \theta_w \hat{w}_{t-1}. \tag{A.6}$$

Combining (A.5) and (A.6), we can write the Phillips-curve analogue of real wage inflation as follows:

$$\begin{aligned} \hat{w}_t = & \frac{\beta \theta_w}{1 + \beta \theta_w^2} E_t \hat{w}_{t+1} + \frac{\theta_w}{1 + \beta \theta_w^2} \hat{w}_{t-1} \\ & + \frac{(1 - \beta \theta_w)(1 - \theta_w)}{1 + \beta \theta_w^2} (\hat{U}_N(t) - \hat{U}_c(t) + \hat{p}_c(t)). \end{aligned}$$

Appendix B. Bayesian Estimation Preliminaries

Let \mathbb{M} denote a generic model and let $\theta_{\mathbb{M}}$ be the vector of parameters associated with it. Let \mathbb{Y} denote the data that is used to estimate the model (note that \mathbb{Y} does not have an \mathbb{M} subscript, i.e., it is assumed that the data used is the estimation routine is constant across models). Bayesian estimation proceeds by specifying a prior distribution over $\theta_{\mathbb{M}}$ which is denoted here by $\mathbb{P}(\mathbb{M}, \theta_{\mathbb{M}})$. The prior is then combined with the likelihood computed using the data to form the posterior distribution of parameters as follows:

$$\mathbb{P}(\theta_{\mathbb{M}} | \mathbb{M}, \mathbb{Y}) \propto \mathbb{P}(\mathbb{Y} | \mathbb{M}, \theta_{\mathbb{M}}) \mathbb{P}(\mathbb{M}, \theta_{\mathbb{M}}). \tag{B.1}$$

Draws from the posterior distribution are generated by applying the Gibbs sampler using standard Markov chain Monte Carlo (MCMC) techniques.⁵³

B.1 Model Selection

The marginal density of the data given the model \mathbb{M} is given by

$$\mathbb{P}(\mathbb{Y}|\mathbb{M}) = \int_{\theta_{\mathbb{M}}} \mathbb{P}(\mathbb{Y}|\mathbb{M}, \theta_{\mathbb{M}})\mathbb{P}(\mathbb{M}, \theta_{\mathbb{M}}). \quad (\text{B.2})$$

This quantity has the interpretation of being the probability of observing the data given the true model is \mathbb{M} . In order to compare two models \mathbb{M}_1 and \mathbb{M}_2 , first the prior odds are specified for both models. These are then combined with the marginal densities to obtain posterior odds ratios which are used for the purpose of model comparison.

$$PO_{1|2} = \frac{\mathbb{P}(\mathbb{Y}|\mathbb{M}_1)\mathbb{P}(\mathbb{M}_1)}{\mathbb{P}(\mathbb{Y}|\mathbb{M}_2)\mathbb{P}(\mathbb{M}_2)} \quad (\text{B.3})$$

One advantage of the Bayesian framework is that the models do not have to be nested.⁵⁴ Throughout this paper, a non-informative prior is assumed on the models ($\mathbb{P}(\mathbb{M}_1) = \mathbb{P}(\mathbb{M}_2) = 0.5$) so that the ratio of marginal data densities is equal to the posterior odds ratio, which in this case is also equal to the frequently quoted statistic called the Bayes factor.

Appendix C. Bayesian Estimation Robustness Checks

The parameters quantifying import price flexibility as well as the elasticity of marginal cost with respect to the risk-free rate are critical in determining the role played by trade finance in propagation of business cycle shocks. This section conducts a series of robustness checks with regard to these parameters. Table C.1 reports posterior

⁵³See Koop, Poirier, and Tobias (2007) for an overview of MCMC techniques.

⁵⁴Note however that in order for the data densities to be comparable, the data used in estimating the two models should be the same and the priors should be proper (i.e., they should define a valid distribution that integrates to one). These conditions will be imposed throughout the paper in order to keep the model comparisons valid.

Table C.1. Posterior Means of Key Parameters under Different Model Assumptions/Restrictions

	$\theta^{\text{US Import}}$	$\theta^{\text{EU Import}}$	$\delta^{\text{EU} \rightarrow \text{US}}$	$\delta^{\text{US} \rightarrow \text{EU}}$
$\sigma_c = 1$	0.31	0.72	2.02	1.68
$\eta = 1$	0.33	0.96	2.40	1.94
Domestic Cost Channel	0.33	0.84	2.36	1.89
Sticky Wages	0.37	0.84	2.12	1.79

Note: The prior mean and standard deviation of the parameters is the same as that in the benchmark case (table 6) except when indicated in the first column.

means of these parameters under different variations of the model. For each of the cases reported in table C.1, the prior mean and standard deviation of the parameters is the same as that in the benchmark case (table 6) except when indicated in the first column.

Since the elasticity of intertemporal substitution is estimated to be somewhat higher in comparison with the literature in the baseline case, the first row considers a model with log utility. The second row considers another restriction on the model by fixing the intratemporal elasticity of substitution between domestic and foreign bundles. As argued before, there is little consensus in the value of this parameter in the literature, and a value of 1 can be considered a compromise between the trade and business cycle literatures.⁵⁵ The third row considers a model in which the cost channel of monetary policy is operational even in the domestic sector, i.e., even the goods-producing firms are required to borrow in order to finance their wage bill. This is typically how the cost channel of monetary policy has been modeled in the literature so far.⁵⁶ As is evident from the results reported in the table, the estimates of the main parameters of interest are robust to all these departures from the baseline version of the model.

⁵⁵A more thorough approach would be to allow for dynamic elasticities as discussed in Crucini and Davis (2013) and Drozd, Kolbin, and Nosal (2014). However, this approach is not undertaken since the main message of the paper is robust to the value of the elasticity used.

⁵⁶See, for instance, Barth and Ramey (2002), Christiano, Eichenbaum, and Evans (2005), and Ravina (2007).

Table D.1. Comparison of Data- and Model-Generated Variances

		Data	Model with Trade Finance	Model without Trade Finance
1	$\pi^{US,CPI}$	1.55	0.85	1.2
2	$\pi^{EU,CPI}$	2.78	4.92	2.75
5	ΔY^{US}	0.31	0.39	0.4
6	ΔY^{EU}	0.22	0.34	0.31
7	$\% \Delta E$	19.99	18.58	21.42

Appendix D. Posterior Predictive Moments

The estimation results in the main text show that the model with trade finance provides a better fit to the data, as measured by the marginal density. To help shed some light on which moments of the data the two models help match better, this section discusses some of the posterior predictive moments generated by the two models and compares them with their counterparts in the data.⁵⁷ Table D.1 presents a comparison of the variances of output and inflation for the two countries as well as the nominal exchange rate depreciation. While the model with trade finance yields variances for U.S. output and the nominal exchange rate depreciation that are closer to the data, the model without trade finance performs better with respect to the other variances. The model with trade finance begins to outperform the model without trade finance more systematically when going to higher order and cross-moments. As an example, table D.2 presents a comparison of the autocorrelations of the different variables generated by the two models with the data. An entry of “1” indicates that the corresponding value of the model with trade finance was closer to the data, while “0” indicates that the value for the model without trade finance is closer to the corresponding value in the data. In total, the model with trade finance generates moments that are closer to the data in 42 out of the 72 possible instances.

⁵⁷See An and Schorfheide (2007) for a similar approach.

Table D.2. Comparison of Autocorrelations between Models with and without Trade Finance and the Data

Lags ↓	$\pi^{US,CPI}$	$\pi^{EU,CPI}$	R^{US}	R^{EU}	ΔY^{US}	ΔY^{EU}	% ΔE	$\pi^{EU,GDP}$	$\pi^{US,GDP}$
1	0	0	0	0	0	1	0	1	0
2	0	1	0	0	0	1	0	1	1
3	1	1	0	0	0	1	1	1	1
4	1	1	0	0	0	1	1	1	1
5	1	1	0	0	1	1	1	0	1
6	1	1	1	0	1	1	0	1	1
7	1	0	1	0	1	1	1	1	1
8	1	0	1	0	0	1	0	0	1

Note: "1" indicates that the corresponding value of the model with trade finance was closer to the data, while "0" indicates that the value for the model without trade finance was closer to the corresponding value in the data.

Appendix E. Data

E.1 Correlations and Plots

This appendix provides the details and sources for the data used in the empirical part of the paper. Unless otherwise mentioned, the data are at quarterly frequency from 1983:Q1–2007:Q4. They are seasonally adjusted and demeaned before estimation.

U.S. Data:

- R^{US} : Effective federal funds rate, nominal, annualized, percentage
- ΔY^{US} : Quarter-to-quarter growth rate of GDP per capita computed as follows:

$$\Delta Y_t^{US} = 100 \left[\log \left(\frac{GDP_t}{POP_t} \right) - \log \left(\frac{GDP_{t-1}}{POP_{t-1}} \right) \right]$$

Note: Nominal GDP is converted to real using the GDP deflator.

- CPI Inflation:

$$\pi_t^{CPI,US} = 400 [\log (CPI_t) - \log (CPI_{t-1})]$$

- GDP Deflator Inflation:

$$-\pi_t^{GDP,US} = 400 [\log (GDPDEF_t) - \log (GDPDEF_{t-1})]$$

- Import Price Inflation (used only in robustness checks, not used in benchmark estimation)

$$-\pi_t^{IM,US} = 400 [\log (P_{IM,t}) - \log (P_{IM,t-1})]$$

Data Sources: The data for the U.S. block are taken from the Bureau of Economic Analysis (BEA) National Income and Product Accounts (NIPA). The data on population are taken from Ramey (2011)'s publicly available data set.

EU Data:

- R^{EU} : Effective federal funds rate, nominal, annualized, percentage

- ΔY^{EU} : Quarter-to-quarter growth rate of GDP per capita computed as follows:

$$\Delta Y_t^{EU} = 100 \left[\log \left(\frac{GDP_t}{POP_t} \right) - \log \left(\frac{GDP_{t-1}}{POP_{t-1}} \right) \right]$$

Note: Nominal GDP is converted to real using the GDP deflator.

- CPI Inflation:

$$\pi_t^{CPI,EU} = 400 [\log(CPI_t) - \log(CPI_{t-1})]$$

- GDP Deflator Inflation:

$$-\pi_t^{GDP,EU} = 400 [\log(GDPDEF_t) - \log(GDPDEF_{t-1})]$$

- Nominal Exchange Rate Depreciation:

$$-\Delta E_t = \log(E_t) - \log(E_{t-1})$$

Data Sources: The data for the EU block are taken from the European Central Bank (ECB) Area Wide Model (AWM) database. The nominal effective exchange rate series before 2000 is taken from Lubik and Schorfheide (2006)'s publicly available database.

Trade Data:

- Bilateral trade data between the United States and the European Union at quarterly frequency are taken from the International Monetary Fund's (IMF's) Direction of Trade Statistics (DOTS). The database only covers merchandise trade and is used in this paper as a proxy for total trade.

$$\Delta \frac{trade}{GDP} = 100 \left[\log \left(\frac{Exports_t + Imports_t}{GDP_t^{US}} \right) - \log \left(\frac{Exports_{t-1} + Imports_{t-1}}{GDP_{t-1}^{US}} \right) \right] \quad (E.1)$$

$$\Delta \frac{Import}{GDP} = 100 \left[\log \left(\frac{Imports_t}{GDP_t} \right) - \log \left(\frac{Imports_{t-1}}{GDP_{t-1}^{US}} \right) \right] \quad (E.2)$$

References

- Adolfson, M., S. Laséen, J. Lindé, and M. Villani. 2007. "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-through." *Journal of International Economics* 72 (2): 481–511.
- Ahn, J. 2014. "Understanding Trade Finance: Theory and Evidence from Transaction-level Data." Unpublished, International Monetary Fund.
- Ahn, J., M. Amiti, and D. E. Weinstein. 2011. "Trade Finance and the Great Trade Collapse." *American Economic Review* 101 (3): 298–302.
- Ahn, J., A. K. Khandelwal, and S.-J. Wei. 2011. "The Role of Intermediaries in Facilitating Trade." *Journal of International Economics* 84 (1): 73–85.
- Aksoy, Y., H. S. Basso, and J. C. Martinez. 2012. "Investment Cost Channel and Monetary Transmission." *Central Bank Review* 11 (2): 1–13.
- Amiti, M., and D. E. Weinstein. 2009. "Exports and Financial Shocks." NBER Working Paper No. 15556.
- . 2011. "Exports and Financial Shocks." *Quarterly Journal of Economics* 126 (4): 1841–77.
- An, S., and F. Schorfheide. 2007. "Bayesian Analysis of DSGE Models." *Econometric Reviews* 26 (2–4): 113–72.
- Angeloni, I., L. Aucremanne, M. Ehrmann, J. Galí, A. Levin, and F. Smets. 2006. "New Evidence on Inflation Persistence and Price Stickiness in the Euro Area: Implications for Macro Modeling." *Journal of the European Economic Association* 4 (2–3): 562–74.
- Antràs, P., and C. F. Foley. 2015. "Poultry in Motion: A Study of International Trade Finance Practices." *Journal of Political Economy* 123 (4): 853–901.
- Asmundson, I., T. Dorsey, A. Khachatryan, I. Niculcea, and M. Saito. 2011. "Trade and Trade Finance in the 2008-09 Financial Crisis." IMF Working Paper No. 11/16.
- Auboin, M. 2007. "Boosting Trade Finance in Developing Countries: What Link with the WTO?" WTO Staff Working Paper No. ERS-D-2007-4.
- Backus, D. K., F. Gavazzoni, C. Telmer, and S. E. Zin. 2010. "Monetary Policy and the Uncovered Interest Parity Puzzle." NBER Working Paper No. 16218.

- Barth, M. J., III, and V. A. Ramey. 2002. "The Cost Channel of Monetary Transmission." In *NBER Macroeconomics Annual 2001*, Vol. 16, ed. B. S. Bernanke and K. Rogoff, 199–256 (chapter 4). MIT Press.
- Bekaert, G., and R. Hodrick. 2017. *International Financial Management*. Cambridge University Press.
- Bigio, S., and J. La'O. 2013. "Financial Frictions in Production Networks." Working Paper, University of Chicago Booth.
- Bils, M., and P. J. Klenow. 2004. "Some Evidence on the Importance of Sticky Prices." *Journal of Political Economy* 112 (5): 947–85.
- Broda, C., and D. E. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics* 121 (2): 541–85.
- Calvo, G. A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–98.
- Cecchetti, S. G. 1999. "Legal Structure, Financial Structure, and the Monetary Policy Transmission Mechanism." *Economic Policy Review* (Federal Reserve Bank of New York) 5 (2, July): 9–28.
- Choi, W. G., and Y. Kim. 2005. "Trade Credit and the Effect of Macro-financial Shocks: Evidence from US Panel Data." *Journal of Financial and Quantitative Analysis* 40 (04): 897–925.
- Chor, D., and K. Manova. 2012. "Off the Cliff and Back? Credit Conditions and International Trade during the Global Financial Crisis." *Journal of International Economics* 87 (1): 117–33.
- Christiano, L. J., M. Eichenbaum, and C. L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1): 1–45.
- Crucini, M. J., and J. S. Davis. 2013. "Distribution Capital and the Short- and Long-run Import Demand Elasticity." NBER Working Paper No. 18753.
- De Loecker, J. 2013. "Detecting Learning by Exporting." *American Economic Journal: Microeconomics* 5 (3): 1–21.
- De Walque, G., F. Smets, and R. Wouters. 2005. "An Estimated Two-country DSGE Model for the Euro Area and the US Economy." Mimeo, European Central Bank.
- Dedola, L., and F. Lippi. 2005. "The Monetary Transmission Mechanism: Evidence from the Industries of Five OECD Countries." *European Economic Review* 49 (6): 1543–69.

- Degryse, H., O. De Jonghe, S. Jakovljevic, K. Mulier, and G. Schepens. 2019. "Identifying Credit Supply Shocks with Bank-Firm Data: Methods and Applications." *Journal of Financial Intermediation* 40 (October): Article 100813.
- Del Negro, M., and F. Schorfheide. 2004. "Priors from General Equilibrium Models for VARs*." *International Economic Review* 45 (2): 643–73.
- . 2006. "How Good is What You've Got? DSGE-VAR as a Toolkit for Evaluating DSGE Models." *Economic Review* (Federal Reserve Bank of Atlanta) 91 (2): 21–37.
- Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters. 2007. "On the Fit of New Keynesian Models." *Journal of Business and Economic Statistics* 25 (2): 123–43.
- Demir, B., and B. Javorcik. 2018. "Don't Throw in the Towel, Throw in Trade Credit!" *Journal of International Economics* 111 (March): 177–89.
- Dietzenbacher, E., and I. Romero. 2007. "Production Chains in an Interregional Framework: Identification by Means of Average Propagation Lengths." *International Regional Science Review* 30 (4): 362–83.
- Disdier, A.-C., and K. Head. 2008. "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *Review of Economics and Statistics* 90 (1): 37–48.
- Djankov, S., C. L. Freund, and C. S. Pham. 2006. "Trading on Time." Policy Research Working Paper No. 3909, World Bank.
- Drozd, L., S. Kolbin, and J. B. Nosal. 2014. "Long-run Price Elasticity of Trade and the Trade-Comovement Puzzle." Working Paper.
- Eaton, J., S. Kortum, B. Neiman, and J. Romalis. 2011. "Trade and the Global Recession." NBER Working Paper No. 16666.
- Ellingsen, T., T. Jacobson, and E. L. von Schedvin. 2016. "Trade Credit: Contract-level Evidence Contradicts Current Theories." Research Paper No. 139, Riksbank.
- Feenstra, R. C. 1994. "New Product Varieties and the Measurement of International Prices." *American Economic Review* 84 (1): 157–77.
- Feenstra, R. C., Z. Li, and M. Yu. 2014. "Exports and Credit Constraints under Incomplete Information: Theory and Evidence from China." *Review of Economics and Statistics* 96 (4): 729–44.

- Galí, J. 2009. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.
- Galí, J., M. Gertler, and J. D. López-Salido. 2001. "European Inflation Dynamics." *European Economic Review* 45 (7): 1237–70.
- Galí, J., and T. Monacelli. 2005. "Monetary Policy and Exchange Rate Volatility in a Small Open Economy." *Review of Economic Studies* 72 (3): 707–34.
- Garcia-Marin, A., S. Justel, and T. Schmidt-Eisenlohr. 2019. "Trade Credit and Markups."
- Gertler, M., S. Gilchrist, and F. M. Natalucci. 2007. "External Constraints on Monetary Policy and the Financial Accelerator." *Journal of Money, Credit and Banking* 39 (2–3): 295–330.
- Gilchrist, S. 2003. *Financial Markets and Financial Leverage in a Two-Country World-Economy*, Vol. 228. Banco Central de Chile.
- Goldberg, L. S., and C. Tille. 2008. "Vehicle Currency Use in International Trade." *Journal of International Economics* 76 (2): 177–92.
- Gopinath, G., O. Itskhoki, and R. Rigobon. 2010. "Currency Choice and Exchange Rate Pass-through." *American Economic Review* 100 (1): 304–36.
- Hoefele, A., T. Schmidt-Eisenlohr, and Z. Yu. 2016. "Payment Choice in International Trade: Theory and Evidence from Cross-country Firm-level Data." *Canadian Journal of Economics/Revue canadienne d'économique* 49 (1): 296–319.
- Huang, K. X., and Z. Liu. 2001. "Production Chains and General Equilibrium Aggregate Dynamics." *Journal of Monetary Economics* 48 (2): 437–62.
- . 2007. "Business Cycles with Staggered Prices and International Trade in Intermediate Inputs." *Journal of Monetary Economics* 54 (4): 1271–89.
- Hummels, D. 1999. "Toward a Geography of Trade Costs." GTAP Working Paper No. 17.
- Hummels, D. L., and G. Schaur. 2013. "Time as a Trade Barrier." *American Economic Review* 103 (7): 2935–59.
- Imbs, J., and I. Méjean. 2012. "Elasticity Optimism."
- Jeffreys, H. 1998. *The Theory of Probability*. Oxford University Press.

- Jiménez, G., S. Ongena, J.-L. Peydró, and J. Saurina. 2012. "Credit Supply and Monetary Policy: Identifying the Bank Balance-Sheet Channel with Loan Applications." *American Economic Review* 102 (5): 2301–26.
- Ju, J., S. Lin, and S.-J. Wei. 2013. "The Credit Channel of Monetary Policy: Solving the Causality Challenge by Using the Impossible Trinity." Working Paper.
- Justiniano, A., and B. Preston. 2010. "Can Structural Small Open-economy Models Account for the Influence of Foreign Disturbances?" *Journal of International Economics* 81 (1): 61–74.
- Kalemli-Ozcan, S., S.-J. Kim, H. S. Shin, B. E. Sørensen, and S. Yesiltas. 2013. "Financial Shocks in Production Chains." Unpublished Manuscript, Princeton University.
- Kashyap, A. K., and J. C. Stein. 1997. "The Role of Banks in Monetary Policy: A Survey with Implications for the European Monetary Union." *Economic Perspectives* (Federal Reserve Bank of Chicago) 21 (3): 2–18.
- Khwaja, A. I., and A. Mian. 2008. "Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market." *American Economic Review* 98 (4): 1413–42.
- Koop, G., D. J. Poirier, and J. L. Tobias. 2007. *Bayesian Econometric Methods*, Vol. 7. Cambridge University Press.
- Lane, P. R. 2001. "The New Open Economy Macroeconomics: A Survey." *Journal of International Economics* 54 (2): 235–66.
- . 2010. "External Imbalances and Fiscal Policy." In *European Economy Occasional Papers 66*, ed. S. Barrios, S. Deroose, S. Langedijk, and L. Pench, 17–30. European Commission.
- Lubik, T., and F. Schorfheide. 2006. "A Bayesian Look at the New Open Economy Macroeconomics." In *NBER Macroeconomics Annual 2005*, Vol. 20, ed. M. Gertler and K. Rogoff, 313–82 (chapter 5). MIT Press.
- Manova, K. 2008. "Credit Constraints, Equity Market Liberalizations and International Trade." *Journal of International Economics* 76 (1): 33–47.
- Manova, K., S.-J. Wei, and Z. Zhang. 2011. "Firm Exports and Multinational Activity under Credit Constraints." NBER Working Paper No. 16905.
- Minetti, R., and S. C. Zhu. 2011. "Credit Constraints and Firm Export: Microeconomic Evidence from Italy." *Journal of International Economics* 83 (2): 109–25.

- Niepmann, F., and T. Schmidt-Eisenlohr. 2017. "International Trade, Risk and the Role of Banks." *Journal of International Economics* 107 (July): 111–26.
- Obstfeld, M., and K. S. Rogoff. 2005. "Global Current Account Imbalances and Exchange Rate Adjustments." *Brookings Papers on Economic Activity* 2005 (1): 67–146.
- Obstfeld, M., and K. S. Rogoff. 1996. *Foundations of International Macroeconomics*, Vol. 30. Cambridge, MA: MIT Press.
- Peneva, E. 2009. "Factor Intensity and Price Rigidity: Evidence and Theory." FEDS Paper No. 2009-07.
- Ramey, V. A. 2011. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economics* 126 (1): 1–50.
- Ravenna, F., and C. E. Walsh. 2006. "Optimal Monetary Policy with the Cost Channel." *Journal of Monetary Economics* 53 (2): 199–216.
- Ravina, E. 2007. "Habit Formation and Keeping Up with the Joneses: Evidence from Micro Data." Available at SSRN 928248.
- Rotemberg, J. J. 1982. "Monopolistic Price Adjustment and Aggregate Output." *Review of Economic Studies* 49 (4): 517–31.
- Ruhl, K. J. 2008. "The International Elasticity Puzzle." University of Texas at Austin.
- Schmidt-Eisenlohr, T. 2013. "Towards a Theory of Trade Finance." *Journal of International Economics* 91 (1): 96–112.
- Schmitt-Grohé, S., and M. Uribe. 2003. "Closing Small Open Economy Models." *Journal of International Economics* 61 (1): 163–85.
- . 2011. "Pegs and Pain." NBER Working Paper No. 16847.
- Smets, F., and R. Wouters. 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1 (5): 1123–75.
- . 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97 (3): 586–606.
- Soderbery, A. 2010. "Investigating the Asymptotic Properties of Import Elasticity Estimates." *Economics Letters* 109 (2): 57–62.
- Taylor, J. B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy* 39 (December): 195–214.

- Timmer, M., and A. Erumban. 2012. "The World Input-Output Database (WIOD): Contents, Sources and Methods." WIOD Background document available at <http://www.wiod.org>.
- Wong, C.-Y., and Y.-K. Eng. 2013. "International Business Cycle Co-movement and Vertical Specialization Reconsidered in Multistage Bayesian DSGE Model." *International Review of Economics and Finance* 26 (April): 109–24.
- Woodford, M., and C. E. Walsh. 2005. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.

Policy Performance and the Behavior of Inflation Expectations*

Eda Gülşen^a and Hakan Kara^b

^aCentral Bank of Turkey

^bBilkent University

This paper investigates the changing behavior of inflation expectations in response to the macroeconomic and policy environment. Using a panel of professional forecasters covering 13 years of inflation-targeting period in a major emerging economy, we present evidence on the behavioral shifts in the inflation expectations associated with evolving macroeconomic and policy performance. The rapidly changing nature of the policy setting and ample data variation in our data set constitute a suitable background to explore this question. We use a unique survey which includes matched policy rate and fixed-horizon inflation expectations at the individual level. Moreover, the paper employs a novel technique where direct feedback from the survey participants is used to determine the baseline empirical model governing expectations dynamics. Interpretation of the empirical findings jointly with the feedback from the survey respondents indicate that the anchoring power of inflation targets depend on the policy performance. The weights attached to inflation targets in forming expectations are strongly associated with the size of the inflation deviation from the targets. As the targets become less credible through time, the survey participants assign increasingly higher weight to past inflation and the relationship between exchange rates and inflation expectations becomes stronger. Overall, our results imply

*We would like to thank Refet Gürkaynak, participants at the Workshop on Advances in Applied Macro-Finance, Turkish Economic Association Conference, and the Central Bank of Turkey internal seminar for useful comments. We also thank the editor Boragan Aruoba and two anonymous referees for helpful suggestions. We are grateful to the data governance and statistics department of the Central Bank of Turkey for the data and fruitful collaboration in conducting the feedback study. The views expressed in this paper are those of the authors and do not necessarily represent the official views of the Central Bank of Turkey.

that expectations behavior might display significant and rapid shifts with the underlying economic and policy performance. Therefore, policymakers in advanced and emerging economies should not take the current stability of inflation expectations for granted.

JEL Codes: C51, C53, E31, E37, E58.

Another gap in our knowledge about the nature of the inflation process concerns expectations. . . Perhaps most importantly, we need to know more about the manner in which inflation expectations are formed and how monetary policy influences them.

– Janet Yellen (2016)

1. Introduction

Inflation expectations constitute an integral part of the monetary theory and policy (Blinder et al. 2008, Galí 2008). The behavior of inflation expectations is often the key input for forecasting and policy analysis models used by policymakers. Anchored longer-term inflation expectations is the hallmark of effective and credible monetary policy. Expectations drive a wide range of economic variables, which, in turn, affect real economic activity and inflation dynamics. Therefore, understanding inflation expectations and their interaction with monetary policy is important from an academic and policy perspective.

This paper seeks to understand how the behavior of inflation expectations shifts in response to policy performance. With the widespread adoption of price-stability-oriented policies during the past decades, inflation expectations have been increasingly anchored in many economies (Gürkaynak, Levin, and Swanson 2010). One important question is whether this success should be taken for granted in designing future monetary policy. Recently, this question is particularly of more relevance, given the excessive reliance on monetary expansion through unconventional tools and the tendency towards curbed central bank independence across the globe. Our study aims to shed some light on this question by utilizing a unique data set on inflation expectations. Using a panel of expectations covering 13 years of inflation-targeting period from Turkey, we investigate the changing behavior of inflation expectations in

response to macroeconomic and policy environment. Turkish macroeconomic conditions and policy framework, which has been subject to frequent changes during the past decade, provides an ideal laboratory for the analysis of time-varying aspects of the expectations behavior.

Using a unique survey data set and rolling panel regressions, we explore several questions pertaining to the behavioral aspects of inflation expectations: How do agents form their inflation expectations in relation with the macroeconomic and policy environment? Do expectations dynamics change through time and across policy regimes? How do inflation expectations respond to shifts in the monetary policy framework and the policy performance? Answering these interrelated questions would not only yield insights into our main question of interest but also improve general understanding of the behavior of inflation expectations, which, in turn, may contribute to build more realistic models and formulate sound policy responses.

In order to conduct an analysis on inflation expectations, we need a quantitative measure of expectations. This paper employs the survey compiled by the Central Bank of the Republic of Turkey (CBRT), called “Survey of Expectations.” The survey comprises one- and two-year-ahead fixed-horizon inflation expectations at the monthly frequency since 2006 along with some key macrovariable forecasts, incorporating a rich variety of responses at the individual level. A unique property of the survey is including policy rate expectations at the micro level, which allows us to extract forecaster-specific monetary policy surprises—a rare feature for such surveys.

Using individual-level survey data helps to identify the relationships through cross-sectional variation. Moreover, survey-based measures of inflation expectations reflect direct forecasts by economic agents, thus they have low sensitivity to varying market liquidity and do not require any adjustment or inflation risk compensation as opposed to market-based measures. These advantages may become more relevant in an emerging economy with relatively less developed financial markets and volatile risk premium. However, surveys may also have some weaknesses compared with market-based measures (Armantier et al. 2017). Because of the absence of direct financial consequences and limited ability to process information, survey responses may suffer from cheap-talk

problems, weak incentives, herd behavior, strategic misreporting, as well as sticky information and/or inattention issues.¹ Notwithstanding these shortcomings, exploring the behavioral aspects of survey-based expectations on a micro basis and identifying the major shifts through time has the potential to provide important insights for the design and formulation of monetary policy (Coibon et al. 2020).

Determinants of inflation expectations and their interaction with the monetary policy have been studied extensively in the literature. A significant fraction of the previous work has concentrated on the variations of empirical closed-economy New Keynesian models across advanced economies (Mankiw, Reis, and Wolfers 2004; Coibon, Gorodnichenko, and Kamdar 2018), whereas our playground is an open emerging economy with rapidly evolving policy environment and imperfect credibility of institutions. Some related papers have explored the role of the policy framework in the behavior of inflation expectations, assessing the significance of the inflation-targeting regime in affecting expectations dynamics across countries (see, e.g., Brito and Bystedt 2010; Gürkaynak, Levin, and Swanson 2010). Another strand of the literature, closer to our work, has investigated the changing behavior of inflation expectations through time within a particular economy.²

Our paper's contribution to the literature can be summarized in four dimensions: First, we use a unique monthly data set including matched monetary policy and inflation expectations at the individual level, which is a valuable feature especially for estimating the impact of policy surprises on inflation expectations and their evolution through time. Availability of matched inflation and policy rate expectations at the micro level is a rare asset for expectation surveys. Second, we link the documented changes in the behavior of expectations to several aspects such as operational framework and credibility gap, showing that the role of nominal anchors may shift quickly depending on the policy performance. Third, we

¹See e.g., Keane and Runkle (1990); Manski (2004); Pesaran and Weale (2006); Inoue, Kilian, and Kiraz (2009); and Marinovic, Ottaviani, and Sørensen (2013).

²Some examples are Celasun, Gelos, and Prati (2004), Carvalho and Minella (2012), and Cortes and Paiva (2017), for emerging economies; Blanchflower and MacCoille (2009), Strohsal, Melnick, and Nautz (2016), and Ciccarelli, Garcia, and Montes-Galdón (2017) for advanced economies. See also Köse et al. (2019) for a comprehensive literature survey on the dynamics of inflation expectations.

adopt a novel methodology where direct feedback from the survey participants is received regarding the construction of their inflation forecasts, where the results are used to build the base for the empirical model and to complement the main findings. Fourth, we use a macro data set with ample variation in variables of interest, which helps to identify key relationships. High volatility in inflation expectations and macroeconomic variables in Turkey provides substantial variation to explore the shifts in the dynamics of inflation expectations.

Overall, both the rich content of our data set and the rapidly changing nature of the Turkish economic context present a suitable background to study the behavior of inflation expectations and their interaction with the macroeconomic and the policy environment.

To our knowledge, this is the first study to employ individual-level direct policy surprises to investigate the response of inflation expectations to monetary policy surprises. The literature has used event studies (Bernanke and Kuttner 2005; Gürkaynak, Sack, and Swanson 2005), structural vector autoregressive (SVAR) models (Christiano, Eichenbaum, and Evans 1999), or a combination of both (Gertler and Karadi 2015) to identify the impact of monetary shocks on the inflation expectations. These papers, by nature, implicitly assume that monetary policy surprises are identical for each agent. Moreover, SVARs and other structural models often impose strong identifying assumptions. Using individual-level monetary policy surprises directly extracted from surveys might provide complementary evidence to the existing work on identifying the effect of monetary policy on inflation expectations.

More recently, some studies have explored the impact of monetary policy surprises using survey data. These papers have mostly focused on the effect of unconventional monetary policy (quantitative easing and forward guidance) on economic agents' expectations. However, none of these studies use *direct* monetary policy surprises at the *individual level*. For example, Boneva et al. (2016) explore the impact of asset purchase amounts on firms' inflation expectations, but they implicitly assume that the unexpected component of the quantitative easing is identical for all firms. Altavilla and Giannone (2017) extract the revision in agents' monetary policy expectations from their bond yield forecasts at the individual level, which provides a micro but *indirect* measure for policy effects at the individual

level. Eminidou, Zachariadis, and Andreou (2020) utilize an estimated monetary policy reaction function to extract consumer-level monetary policy surprises; yet, their measure is *indirect and model dependent*. Our study, on the other hand, uses individual-level direct policy surprises, enabling us to assess the impact of monetary policy on the inflation expectations without imposing model-dependent identifying assumptions, which is a unique feature compared with the related work in the literature.

Given this background, we run full-sample and rolling panel regressions to explore the dynamics of inflation expectations and their interaction with the economic environment. Our estimates suggest that the inflation expectations are significantly related to macrovariables such as exchange rates, oil prices, inflation realizations, and inflation targets, as well as individual-level policy surprises, consistent with the previous literature on emerging economies.³ More importantly, rolling regressions reveal that the parameters governing the expectations formation process change considerably through time, possibly responding to the shifting performance of the policy framework and sliding external conditions. Empirical results indicate that the weight attached to inflation targets by forecasters is inversely related to the size of the target breaches. Moreover, we document that the sensitivity of inflation expectations to monetary policy surprises varies significantly with the policy framework.

The findings are suggestive of a significant change in the expectation behavior, possibly associated with the policy performance. Despite the fairly anchored inflation expectations during the initial years of the inflation-targeting framework, expectations behavior changes rapidly through time with the persistent breaches of the targets on the upside. The relationship between exchange rates and inflation expectations becomes stronger and survey participants assign increasingly higher weight to past inflation through time. These findings are supported by the direct feedback survey we conducted among the participants, which indicates that, as of the end of the sample period, inflation target ceases to be a key parameter in

³See, for example, Carvalho and Minella (2012) for Brazil; Pedersen (2015) for Chile; and Kara and Küçük (2010), Çiçek, Akar, and Yücel (2011), and Başkaya, Gülşen, and Kara (2012) for Turkey.

driving medium-term expectations. Taken together, the results point to a significant weakening in the credibility and the anchoring power of inflation targets through time, associated with the underlying policy and economic performance.

Our findings imply that the existing stability of inflation expectations across the globe should not be taken for granted. The credibility and the ability to shape expectations around an inflation target may change rapidly depending on the policy performance. Recent overshoots of inflation targets in many economies and the tendency towards more discretionary policies in other jurisdictions warrant caution in this respect.

The remainder of the paper is organized as follows: The next section explains the main features of the expectation survey used in the paper and summarizes the evolution of inflation expectations throughout the sample period. The third section presents the empirical model and the changing behavior of inflation expectations along with some robustness analysis. The last section presents final remarks and some reflections.

2. An Overview of Inflation and Inflation Expectations in Turkey

Turkish economy and inflation dynamics have witnessed a comprehensive transition after 2001 with the adoption of a floating exchange regime along with an implicit inflation-targeting regime. Following a successful disinflation period between 2002 and 2005, which brought inflation down to single digits after many decades of high double-digit inflation, explicit inflation targets were adopted in 2006 to lock in the gains from disinflation. The period between 2006 and 2010 can be described as a standard inflation-targeting regime where the central bank used a single policy rate with a medium-term forecast horizon. The policy framework has evolved into a more flexible form of inflation targeting through time. Following the global financial crisis and the European debt crisis, multiple instruments were used to deal with the consequences of excessive global liquidity and the volatility in capital flows, with financial stability being adopted as a supplementary goal. To this end, the period between 2011 and 2015 involved unconventional interest rate corridor policies along with the active use of reserve requirement tools, where credit and exchange

rate served as intermediate variables. Monetary policy operational framework reverted to a relatively more conventional setup after 2016 when leading central banks started implementing exit strategies from quantitative easing policies. These frequent shifts in the background policy framework provide ample variation to identify the changes in the expectation behavior associated with the monetary framework.

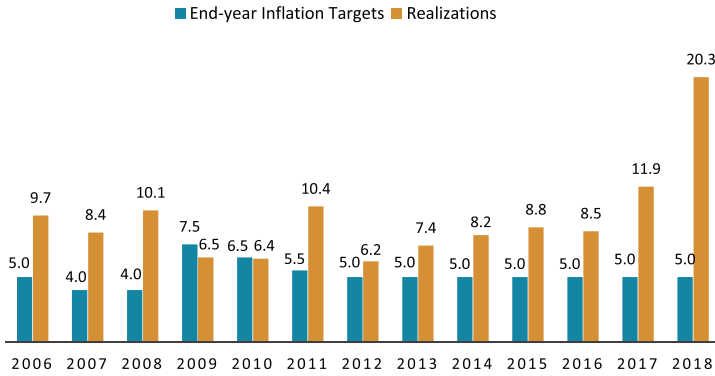
Another interesting feature of our data set is the variation in inflation targets, which is typically absent in many inflation-targeting countries. Since 2006, consumer price index (CPI) inflation targets have been announced by the CBRT in each December for a three-year horizon. During the initial years, the multi-year targets were set constant at 4 percent. However, targets were revised on the upside in June 2008, where 2009–11 inflation targets were set at 7.5, 6.5, and 5.5 percent, respectively. The inflation target has stayed at 5 percent thereafter (figure 1). Deviation of inflation from the targets has also showed considerable variation. The targets were breached consistently on the upside at varying degrees, except for the years 2009 and 2010. The size and the volatility of the deviation of inflation from the targets, coupled with the variation in the targets, allow us to explore whether and how the performance of the inflation-targeting framework has affected the anchoring role of the targets.

2.1 The Survey

The CBRT launched the “Survey of Expectations” in August 2001 to measure and monitor expectations for inflation and some key macroeconomic variables.⁴ Expectations behavior analyzed in this paper pertains to the forecasts collected through this survey. The survey participants include commercial banks, asset management and investment banks, insurance and factoring companies, pension funds, large firms and conglomerates, economists, and other professionals. Financial institutions constitute a large fraction (around 80 percent) of the survey participants. The data governance and statistics department of the CBRT regularly monitors the quality of

⁴The most recent set of the survey questions can be found at the CBRT website.

Figure 1. End-Year Inflation Targets and Realization, 2006–18

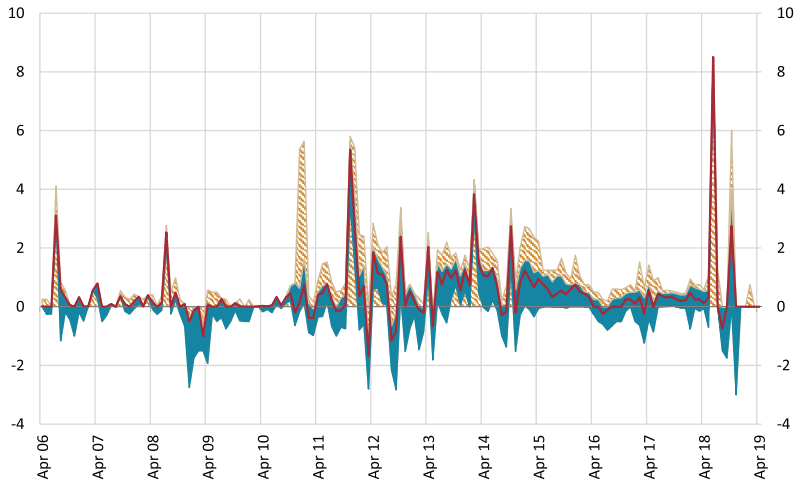


Source: CBRT, TurkStat.

the survey and contacts the participants to ensure a satisfactory participation rate. The survey is distributed to around 100 participants every month comprising professionals and institutions. The response rate has varied between 60 and 70 percent since 2006.⁵ For the financial sector and large firms, the survey is sent directly to a representative of the institutions—typically the chief economist or the head of research. In a recent feedback study covering survey participants, around three-fourths of the respondents stated that their reported forecasts are institutional projections, implying that the responses largely reflect the institutions’ official forecasts, possibly incorporating multiple cross-checks. Given this structure, the forecast production process should be less prone to the criticisms cited in the literature such as herd behavior, cheap-talk problems, and strategic misreporting.

Because forecasts are largely interpreted as institutions’ views rather than individuals’ own projections, changes in the specific survey representatives should have limited impact on the behavior of the institutions’ forecasts. Still, the turnover may have some effect on the behavior of forecasts, as each individual is likely to add his/her

⁵Gülşen and Kara (2019) provide more detail on the survey response rates through time.

Figure 2. Distribution of Monetary Policy Surprises

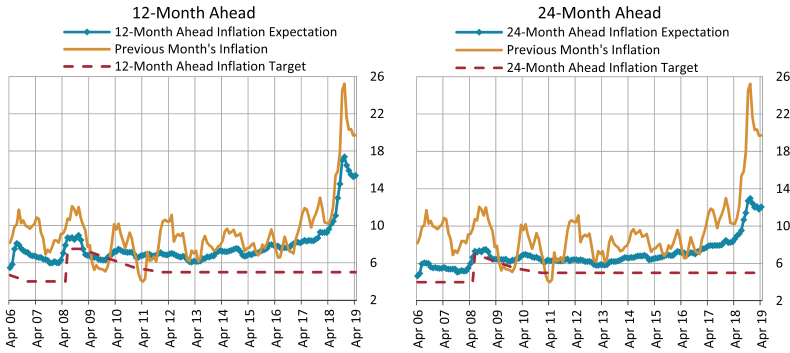
Source: CBRT.

Notes: The vertical axis reports the distribution of monetary policy surprises across survey respondents. For the April 2006–May 2018 period, individual-level monetary policy surprise is calculated as the difference between survey participants' end-of-month expectation and the realization for interbank market rate for the corresponding month. Since June 2018, survey expectations on one-week repo rate are used to calculate monetary policy surprises. A positive (negative) value for the surprise implies monetary policy is tighter (easier) than expected. The solid line is the median of the monetary policy surprise distribution for each month. The shaded areas comprise 50 percent and 90 percent of the cross-sectional distribution.

own judgment in forming expectations. Nonetheless, this effect is likely to be small on average, because in our sample only one-tenth of the survey respondents change institutions per year.

One of the strengths of the survey is that it has quantitative fixed-horizon inflation forecasts along with monetary policy expectations matched at the individual level. This unique feature enables us to explore the response of inflation expectations to the monetary policy surprises without imposing model-dependent identifying assumptions. As shown in figure 2, the distribution of the monetary policy surprises is quite dispersed across participants except for the periods of sharp and unpredicted movements in the policy rate during extreme market volatility. It is also interesting to observe

Figure 3. Inflation, Expectations, and Targets



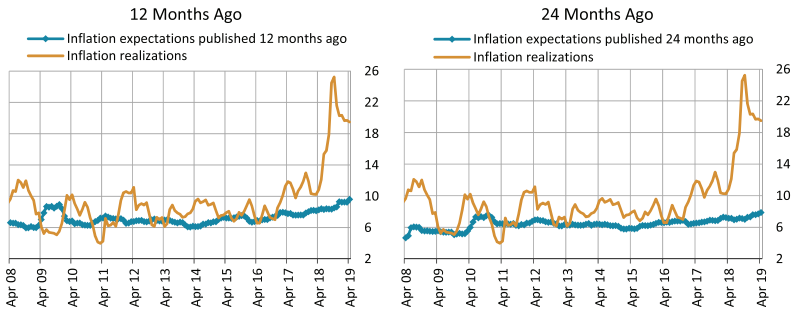
Source: CBRT, TurkStat.

Notes: All the inflation, target, and corresponding expectations series reflect annualized figures. The darker line (blue in color version of figure online) shows mean inflation forecasts by participants in the CBRT’s Survey of Expectations. Until 2013, the survey was conducted twice a month. Starting from January 2013, participants are surveyed once a month. We use second-half-of-the-month results before January 2013. Monthly inflation target series are computed by linear interpolation of the year-end inflation targets.

that the cross-sectional dispersion increased considerably after 2010 with the implementation of the unconventional interest corridor policy. This picture suggests that exploiting the variation in surprises across forecasters may provide additional insights into the existing literature on estimating the impact of monetary policy surprises. Substantial variation in both cross-sectional and time-series dimensions facilitates the identification of the impact of policy shocks even in narrow moving-window estimates.

A cursory look at the historical plot of average inflation expectations reveals that expectations have been below the realized inflation but above the targets most of the time (figure 3). Moreover, inflation turned out to be consistently higher than expectations during the past decade (figure 4). The gap between inflation and the target has widened markedly at the end of the sample, which is likely to have affected the expectations formation process due to weaker anchoring role of the targets. In fact, inflation expectations have drifted upwards and moved closer to realized inflation after 2013, possibly related to persistent overshoots of the inflation targets. These observations suggest that anchoring power of the targets may

**Figure 4. Forecast Performance of the Survey:
Inflation Expectations and Realizations**



Source: CBRT, TurkStat.

have waned through time—a recurrent theme that will be investigated throughout the paper.

3. Formation of Inflation Expectations

This section employs empirical specifications to explore the behavior of inflation expectations and their evolution through time. Deciding on the set of explanatory variables in an empirical model governing expectation dynamics is not a trivial task because inflation expectations of the professional agents may respond to a large array of variables affecting inflation outlook. Recent literature has suggested that, because of the reasons such as limited capacity for processing information, agents may choose a small set of variables to form their information set (Sims 2003). Existing studies on emerging economies typically adopt some version of an open-economy Phillips curve to explore the formation of inflation expectations, augmented by country-specific explanatory variables (Celasun, Gelos, and Prati 2004; Carvalho and Minella 2012). In this paper, we pursue a novel approach by utilizing the results of a direct “feedback survey” to determine the set of candidate explanatory variables, where the survey participants are asked to reveal the variables they use in constructing inflation forecasts. Doing so allows us to adopt a more tailored approach in choosing the variables of interest used in the main regressions, addressing possible endogeneity issues

that may originate from omitted-variables and/or common factor problems.

3.1 A Survey of Survey Respondents: Which Variables Are Important in the Conduct of Inflation Forecasts?

Before turning to the empirical model, we summarize the results of the direct feedback from survey participants.⁶ The feedback is collected by simply asking the survey participants to fill out the degree of importance they attach to certain variables when they forecast annual inflation at one- and two-year horizons. Specifically, we have provided the participants with a list of macrovariables and made the following request: “Please mark the variables you use when constructing your (one- and two-year) inflation forecasts and their degree of importance.” The participants are asked to choose among four options: “high,” “medium,” “low,” and “no” importance. Next, the feedback is quantified and aggregated for each variable by assigning grades to individual responses from 3 to 0, representing the range from high importance to no importance, respectively.

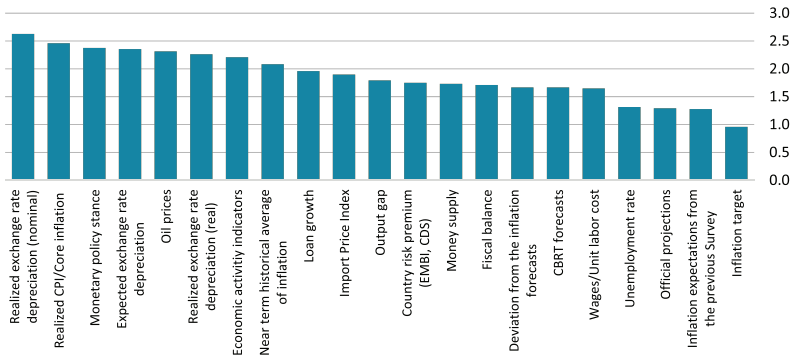
Figure 5 summarizes the results. The horizontal axis depicts the variables that appeared in the list provided to the participants as candidate variables having the potential to influence inflation forecasts. The vertical axis shows the score of each variable averaged across all participants. The quantitative scores provide a metric to assess the degree of relative importance of each variable in driving inflation forecasts.⁷ The closer is the score to 3, the more important is the variable in shaping overall inflation expectations. For example, nominal exchange rate depreciation (USD/TL) makes the top among all variables with a value of 2.63 out of 3, whereas inflation target gets the lowest score with 0.96.⁸

⁶The survey was designed and conducted in June 2019 jointly with the data governance and statistics department of the CBRT. The questions were distributed to around 80 people, which constitutes the whole sample, and 50 of the respondents have provided direct feedback on the variables they use in forecasting inflation.

⁷We have tried different specifications in quantifying the feedback responses, but the ranking of the variables did not change in any meaningful way.

⁸Participants were also asked to state other relevant variables (not listed in the feedback forms) used in forecasting inflation, but they have not revealed any significantly important variable that would change the ranking in figure 5.

Figure 5. Importance of Variables Driving Inflation Forecasts: Survey Participants' Scores (out of 3)



Source: CBRT.

Notes: Values in the vertical axis show the average score of the corresponding variable in driving inflation forecasts across survey participants. The responses of survey participants are quantified by assigning grades from 3 to 0, for “high importance,” “medium importance,” “low importance,” and “no importance,” respectively.

Feedback results from the survey respondents show that the top six variables driving inflation forecasts of the professionals are exchange rates, inflation outturn, monetary policy stance, oil prices, economic activity, and near-term historical average of inflation. Each of these variables has an average score of more than 2 out of 3. These variables will constitute the base for the regressor set in our empirical models. Note that the participants attach high scores to various forms of exchange rate variables (nominal, real, and expected); but given the possible collinearity between these variables, we decided to use only one of the exchange rate variables, namely the nominal depreciation, which has the highest rank among the whole list.

Interestingly, the survey respondents seem to assign a very low weight to the inflation target when forming their inflation expectations. This observation suggests that the inflation target does not serve as an anchor among the survey respondents. We should note that the reported direct feedback is very recent, which represents the expectation formation process at the end of the sample period. Whether the targets had a low weight in shaping the expectations

during the (relatively more successful) initial periods of the inflation-targeting period is an important question to be explored. Therefore, we will include the targets in our empirical specifications to assess the changing nature of expectations and their interaction with the background policy setting. The evolution of the estimated coefficients and the results of the feedback survey will be jointly used for cross-check purposes to support our main hypothesis.

3.2 The Empirical Model

Our aim is to explain the movements in inflation expectations at the individual level. The cross-sectional dimension of our data set captures around 70 participants per month, while the time dimension is about 150 months, which includes a rich panel of forecasters to identify some of the key factors driving inflation expectations. The empirical strategy will be running panel regressions of expectations on the relevant macroeconomic and policy variables and tracking the evolution of the key coefficients through rolling windows.

In light of the feedback from the survey participants and considering the related empirical literature, we construct the following model to explain inflation expectations:

$$\begin{aligned} \pi_{i,t|t+k}^e &= \beta_1 \pi_{t-1} + \beta_2 \pi_{t-1}^{MA12} + \beta_3 \pi_{t|t+k}^{target} + \beta_4 MP_{i,t-1}^{surprise} \\ &\quad + \beta_5 \Delta basket_{t-1} + \beta_6 \Delta ipi_{t-2} + \beta_7 \Delta oil_{t-1} \\ &\quad + \beta_8 D_{Target Revision} + \mu_i + \varepsilon_{it}. \end{aligned} \quad (1)$$

The dependent variable $\pi_{i,t|t+k}^e$ shows k -month-ahead inflation forecast (expectation) of participant i at time t .⁹ The specific lag structure chosen for the explanatory variables reflects the information set available to the survey participants when constructing the forecasts.¹⁰ The first two variables on the right-hand side pertain to

⁹During the initial years of the survey, the longest-term inflation forecast was one year. After the introduction of an explicit inflation-targeting regime in 2006, the survey questions were further expanded to include medium-term (two-year-ahead) inflation forecasts. In order to incorporate the two-year-ahead inflation forecasts, we start the sample at year 2006.

¹⁰Using lagged variables may also help to address potential endogeneity issues between expected inflation and other macrovariables as argued by Mehrotra and Yetman (2018).

observed levels of past inflation: π_{t-1} is the annual inflation rate of the previous month, which is the latest inflation figure observed by the time of the survey. π_{t-1}^{MA12} is the moving average of the annual inflation rate of the previous 12 months. Note that we use both previous month's inflation and last 12 months' average inflation to capture the sensitivity to past inflation components. The idea is that survey participants attach some weight to the most recent level of inflation, but they also consider the history of inflation in forming their forecasts.¹¹ Adding this latter variable to the set of regressors is also justified by the direct feedback from the survey respondents (see figure 5). $\pi_{t|t+k}^{target}$ represents the CBRT's k -month-ahead inflation targets known to the forecaster at time t , which is constructed by interpolating the end-year inflation targets.

$MP_{i,t-1}^{surprise}$ denotes the individual-level monetary policy surprise variable. This variable is constructed by taking into account the changes in the CBRT's operational framework. For the April 2006–May 2018 period, the policy surprise variable is calculated as the difference between participant i 's end-of-month expectation and the realization for the interbank market rate. During this period, the overnight interbank rate is used to represent the monetary policy stance, rather than the official policy rates, because interbank rates have occasionally deviated from the official policy rates during the implementation of unconventional interest rate corridor policy. Related research shows that the de facto policy stance has been represented by the interbank rates during this period (Binici, Kara, and Özlü 2019). Since June 2018 the CBRT reverted to a relatively more conventional interest rate corridor system in which the one-week repo auction rate represents the policy rate. Therefore, after this period, we use survey expectations on the one-week repo rate to calculate monetary policy surprises. A positive (negative) value of $MP_{i,t-1}^{surprise}$ implies monetary policy surprise on the tightening (easing) side. The coefficient of this variable in the rolling regressions will be of particular interest, as part of our aim will be to track

¹¹In fact, R^2 of a simple ordinary least squares (OLS) regression of actual 12-month-ahead inflation to one-month lagged and MA(12) inflation is 0.75, where most of the variation is explained by the MA(12) term. Therefore, past inflation variables we use in the regressions have strong predictive power for future inflation.

the interaction of monetary policy framework with the expectations behavior. Having an individual-level measure of the policy surprise is a valuable feature of the data set. To our knowledge, the CBRT survey is the only official broad-coverage survey asking the expectations of policy rates jointly with fixed-horizon inflation forecasts for a reasonably long period (13 years) at the monthly frequency.

$\Delta basket_{t-1}$ is the annual percentage change in the monthly average currency basket (representing euro and U.S. dollar in equal weights). A positive value in this variable indicates depreciation of Turkish lira. We use Δipi_{t-2} as a measure of economic activity, which is constructed using the three-month moving average of annual percentage change of the seasonally and calendar-adjusted industrial production index. This variable is lagged two months because industrial production data are publicly available with a two-month lag.¹² We apply three-month moving-average transformation to smooth excessive volatility in the monthly industrial production. Moreover, Δoil_{t-1} shows a six-month percentage change of monthly average crude oil price in U.S. dollars.¹³

$D_{Target Revision}$ is a dummy variable controlling for the announcement effect of the target revision in June 2008. The dummy variable takes the value of 1 for June 2008 and 0 otherwise. Finally, μ_i represents individual fixed effects, used to avoid any bias due to time-invariant individual characteristics that may be correlated with the independent variables. We use Driscoll and Kraay (1998) standard errors to account for cross-sectional and time correlation in the errors.

The use of forecaster-level microdata helps to address some of the endogeneity issues related to reverse causality problems in the empirical models using aggregate data, as discussed in Boneva et al. (2016). Individual expectations are affected by the inflation and other macrovariables but cannot significantly influence these variables. Therefore, employing a forecaster-level dependent variable eases the simultaneity problems inherent in macro relationships.

¹²We have also used one-month lagged or contemporaneous values of the industrial production for robustness purposes but the results remained intact.

¹³We use different data transformations for oil and exchange rates (six months and one year percentage change, respectively) in order to avoid possible collinearity between the U.S. dollar and oil prices.

While this addresses the reverse causality issue, expectations and the explanatory variables may still be driven by a common factor, which may be another source of endogeneity. Controlling for all the relevant variables on the right-hand side alleviates the common factor issue, but this is not a trivial task. Relative strength and the novel feature of our approach compared with similar studies is that we are able to relate the choice of explanatory variables to direct evidence, thanks to the availability of feedback from the survey respondents, which should minimize the omitted-variable problem.

Our particular focus when interpreting the empirical results will be on the role of inflation target, past inflation, exchange rates, and monetary policy in driving inflation expectations, as well as their changing nature through time. Table 1 shows panel regression results of the baseline empirical model for one-year and two-year inflation expectations. The high R^2 values, which are 0.8 for 12-month and 0.7 for 24-month expectations, suggest that the model is able to explain a sizable portion of the variations in inflation expectations. Moreover, both the sign and the magnitude of the coefficients on the explanatory variables are reasonable in economic terms. Inflation expectations have a positive and strongly significant relationship with the past inflation terms and the targets. The coefficients on the economic activity, exchange rates, and oil prices are positive and significant. The coefficient on the monetary policy surprise has a negative and significant sign, implying that tighter-than-expected monetary policy leads to lower inflation expectations.¹⁴ However, these effects are not economically significant in the sense that the fit of the regression seems almost identical when we use the median surprise or altogether drop the individual-level surprises (not reported). Still, tracking the sign and statistical significance of the coefficients through time provides valuable information regarding the behavior of inflation expectations.

Although the coefficients on policy surprises seem to be in line with the textbook response, this may not reflect the expectations

¹⁴Note that, under a completely credible inflation target, expectations of inflation at long-enough horizons should not respond to shocks, including policy surprises. In our case, we use one- and two-year expectations due to data limitations for longer-term expectations. One- or two-year-ahead inflation may be within the horizon where policy is perceived to be effective, but not enough to fully offset the impact of shocks and bring inflation back to target at all times.

Table 1. Drivers of Survey-Based Inflation Expectations (April 2006–April 2019)

Dependent Variable: k -month-ahead annual inflation expectations of participant i at time t ($\pi_{i,t t+k}^e$)				
$k =$	(1) 12-Month	(2) 24-Month	(3) 12-Month	(4) 24-Month
CPI Inflation $_{t-1}$	0.335*** (0.029)	0.205*** (0.042)	0.366*** (0.037)	0.259*** (0.035)
MA12 Inflation $_{t-1}$	0.371*** (0.042)	0.229*** (0.080)		
Inflation Target $_{t t+k}$	0.358*** (0.098)	0.665** (0.282)	1.087*** (0.197)	1.482*** (0.320)
Policy Surprise $_{i,t-1}$	-0.025* (0.014)	-0.042** (0.021)	-0.040*** (0.009)	-0.057*** (0.020)
Nom. Depreciation $_{t-1}$	0.034*** (0.002)	0.025*** (0.007)	0.036*** (0.003)	0.025*** (0.006)
IPI Growth $_{t-2}$	0.039*** (0.013)	0.030*** (0.008)	0.042*** (0.013)	0.024*** (0.009)
Oil Price Growth $_{t-1}$	0.010*** (0.002)	0.008*** (0.003)	0.006*** (0.002)	0.005** (0.003)
Dummy _{Target Revision}	-0.087 (0.177)	-1.090* (0.606)	-0.269** (0.112)	-1.102** (0.428)
MA12 Target Deviation $_{t-1}$			1.229*** (0.328)	1.229** (0.478)
MA12 Target Dev. $_{t-1}$ * Inflation Target $_{t t+k}$			-0.186*** (0.051)	-0.222** (0.091)
Observations	8,182	7,943	8,182	7,943
R ²	0.799	0.688	0.803	0.695

Notes: MA12 Target Deviation $_{t-1}$ shows the 12-month moving average of the deviation of annual inflation rate from the inflation target. *, **, and *** represent statistical significance at levels of 10, 5, and 1 percent, respectively. Driscoll and Kraay (1998) standard errors are given in parentheses.

behavior for all episodes, given that the interest-setting framework has shown considerable shifts during our sample period. In the next section, we will run moving-window estimates to understand if the response of the expectations to the interest rate decisions have shown behavioral shifts through time.

Empirical results in table 1 suggest that agents pay significant attention to past inflation terms, represented by the latest inflation print and the near history (as represented by the MA12 term). The coefficient on inflation targets may be interpreted as a measure

of the degree of anchoring in inflation expectations. For one-year-ahead inflation expectations, the coefficient on the inflation target is smaller than the sum of the coefficients on the past inflation variables.¹⁵ Regarding two-year-ahead expectations, the coefficient of the target is higher than the sum of the coefficients on past inflation terms. The finding that longer-term expectations are more sensitive to inflation targets makes sense, given the role of inflation targets in the policy regime. These findings are in line with Mehrotra and Yetman (2018) who argue that, as the forecast horizon shortens, newly arriving public information such as past inflation realizations become more relevant in driving inflation expectations. Overall, full-sample results suggest that inflation targets on average seem to have served at least as a partial anchor for medium-term expectations.

Recall that our direct evidence extracted from the feedback survey indicated that the survey participants do not rank the inflation target as a significant variable in forming their inflation forecasts as of the end of the sample period. On the other hand, the empirical results in table 1 suggest that agents attach a reasonable and highly significant weight to inflation targets for the whole sample period. Taken together, these observations suggest that the role of targets in anchoring expectations may have changed through time, which will be further explored in the upcoming sections.

One candidate explanation for the changing weight of the inflation targets may be related to the sizable and persistent deviations of inflation from the targets, which may have undermined the anchoring role of the targets. In order to further investigate this hypothesis, in the last two columns of table 1, we explore whether the anchoring degree of the targets depends on the past performance in meeting the targets. To this end, we ask the following question: Does the inflation-targeting performance—measured by the gap between inflation realizations and the target—affect the sensitivity of expectations to the targets? In order to test this hypothesis, we interact the inflation targets with the difference between realized inflation and the target in the baseline specification averaged over the past

¹⁵Although the inflation target has a relatively large coefficient in the baseline regression (especially for two-year expectations), its partial R^2 in explaining inflation expectations (reported in table A.1 of the appendix) is relatively small due to low variation of the targets.

year (table 1, columns 3 and 4). The answer is a clear yes, as depicted by the highly significant negative coefficient of the interaction terms shown at the last row. The results reveal that the higher is the gap between inflation and the target, the lower is the weight attached to targets. Sensitivity of expectations to the inflation-targeting performance seems to be higher for medium-term expectations (last column of table 1). These results support the view that persistent upside breaches of the inflation targets have weakened the anchoring power of the targets through time. This finding is also consistent with the direct evidence obtained from the survey participants, who have ranked the inflation target as the least important variable in driving their forecasts in a recent feedback survey (figure 5).

3.3 The Interaction between Exchange Rates and the Expectation Formation Process

We now turn to the interaction of inflation expectations behavior with the movements in exchange rates (table 2). Table 2 runs the baseline regressions by interacting key variables with an “exchange rate depreciation dummy,” which takes the value of 1 for the periods where the exchange rate depreciated in the past 12 months and 0 otherwise. In total, we have 34 appreciation and 123 depreciation periods in our sample. Almost all the appreciation points take place before 2013, which was a period of relatively better performance in reaching the inflation targets.

The coefficient of the interaction term is significant and positive for the past inflation and negative for the inflation targets. In other words, during depreciation episodes, the weight attached to past inflation is higher and the weight on the inflation target is lower, compared with appreciation periods. These results reveal that the targets might be perceived as less of an anchor during depreciation episodes, possibly pointing to some interaction between the credibility of the inflation targets and the exchange rate depreciation. Expectations seem to be more sensitive to exchange rate movements during depreciation periods. These findings suggest that exchange rate depreciation periods coincide with weaker anchoring of inflation expectations. Overall, the behavior of inflation expectations seems to be sensitive to exchange rate movements, suggesting a strong

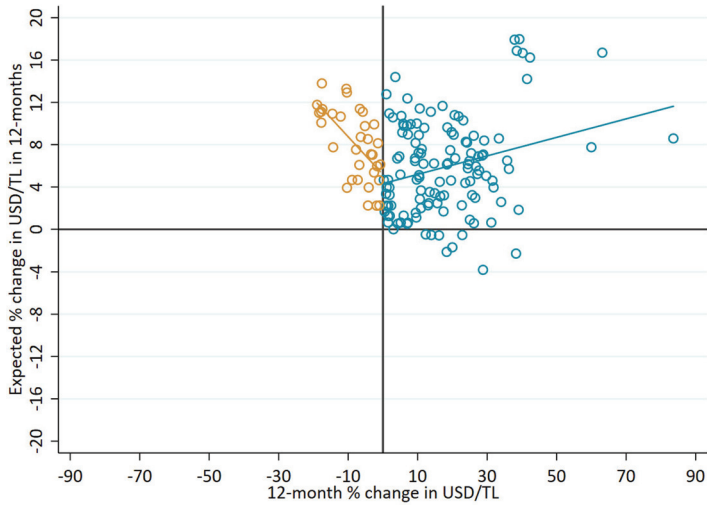
Table 2. Exchange Rate Movements and the Behavior of Expectations (April 2006–April 2019)

Dependent Variable: k -month-ahead annual inflation expectations of participant i at time t ($\pi_{i,t t+k}^e$)		
$k =$	(1) 12-Month	(2) 24-Month
CPI Inflation $_{t-1}$	0.047 (0.034)	-0.017 (0.068)
CPI Inflation $_{t-1}$ * Depr. Dummy	0.272*** (0.056)	0.215*** (0.078)
MA12 Inflation $_{t-1}$	0.405*** (0.038)	0.238*** (0.051)
Inflation Target $_{t t+k}$	0.840*** (0.049)	1.080*** (0.079)
Inflation Target $_{t t+k}$ * Depr. Dummy	-0.647*** (0.146)	-0.588*** (0.161)
Policy Surprise $_{i,t-1}$	-0.026* (0.015)	-0.043*** (0.010)
Nom. Depreciation $_{t-1}$	0.012*** (0.005)	0.010* (0.006)
Nom. Depreciation $_{t-1}$ * Depr. Dummy	0.020*** (0.006)	0.017** (0.007)
IPI Growth $_{t-2}$	0.033** (0.015)	0.026*** (0.005)
Oil Price Growth $_{t-1}$	0.011*** (0.002)	0.008*** (0.002)
Depr. Dummy	1.237 (1.127)	1.120 (1.410)
Observations	8,182	7,943
R ²	0.809	0.695

Notes: Depr. Dummy is a dummy variable that takes 1 for the periods of Turkish lira depreciation, i.e., Nom. Depreciation $_t$ is positive. *, **, and *** represent statistical significance at levels of 10, 5, and 1 percent, respectively. Driscoll and Kraay (1998) standard errors are given in parentheses.

interaction between the exchange rates and the expectations formation process. This result may reflect that exchange rates play a more important role in driving inflation expectations, beyond the dimension of pass-through to domestic prices. In fact, Coibion and Gorodnichenko (2015) argue that in countries with high inflation, economic agents could routinely use exchange rates as a statistic summarizing the stance of monetary and fiscal policies as well as other macroeconomic conditions to infer the rate of inflation.

Figure 6. Realized vs. Expected Exchange Rate Movements (April 2006–April 2019)



Source: CBRT.

In order to further explore the behavioral asymmetry with respect to the exchange rate movements, we look at how the relation between realized and expected exchange rate changes differs during appreciation and depreciation episodes. Figure 6 depicts the scatter-plot of past 12-months’ exchange rate (USD/TL) depreciation rate versus expected depreciation rate in the next 12 months by survey participants. The dots at the right side of the vertical axis indicate that weaker TL observed in the past year prompts expectations of further depreciation in the next 12 months, as most of the observations are in the first quadrant. On the other hand, as depicted by the dots at the left side of the vertical axis, survey respondents expect past appreciation periods to be somewhat reversed by future depreciation periods. In other words, appreciations are perceived as more temporary. These observations may help to explain why the linear relation between exchange rate movements and inflation expectations exhibit asymmetry. To the extent that the actual behavior of price setters mimics that of survey participants, such an asymmetric pattern in the expectation behavior may also lead to asymmetry in the realized exchange rate pass-through to inflation.

3.4 Formation of Expectations: Do the Financial and Real Sectors Differ?

Next, we investigate whether expectations formation differs between real- and financial-sector participants (table 3). This question is addressed by estimating the baseline empirical model with a binary dummy that takes the value of 0 or 1 denoting whether the participant is a representative of the financial or real sector, respectively.¹⁶ We interact the dummy with each regressor and interpret the estimated coefficients. The results suggest that coefficients for both groups are of similar size, yet there are some statistically significant discrepancies. Financial-sector participants significantly put one-third more weight on inflation targets for the medium-term horizon than do real-sector participants. Response of the two-year-ahead financial-sector forecasts to the target revision in June 2008 is stronger and the difference is statistically significant. Regarding the sensitivity of expectations to the exchange rates and economic activity, there are also statistically significant differences across two groups, where financial participants seem to respond more strongly to the release of macrovariables, especially for the medium term. Overall, the results suggest that the financial sector's and the real sector's attentiveness to new information released by the central bank shows some heterogeneity, which echoes the point made by Blinder et al. (2008): Central banks, which largely focus on the financial markets in designing their communication strategy, need to develop alternative tools for communicating with the general public.

3.5 Has the Behavior of Inflation Expectations Changed through Time?

As explained in section 2, Turkish inflation dynamics and monetary policy framework has gone through significant changes during the past decade, especially after the global financial crisis, which might have significant implications for the inflation expectations formation process. We will seek to identify the changes in the behavior

¹⁶The real-sector participants are typically chief financial officers (CFOs) or chief economists of large conglomerates.

Table 3. Financial- and Real-Sector Expectations (April 2006–April 2019)

Dependent Variable: k -month-ahead inflation expectations of participant i at time t ($\pi_{i,t t+k}^e$)		
$k =$	(1) 12-Month	(2) 24-Month
CPI Inflation $_{t-1}$	0.336*** (0.028)	0.206*** (0.029)
CPI Inflation $_{t-1}$ * Real-Sector Dummy	0.006 (0.014)	0.009 (0.009)
MA12 Inflation $_{t-1}$	0.372*** (0.047)	0.235*** (0.063)
MA12 Inflation $_{t-1}$ * Real-Sector Dummy	-0.007 (0.031)	-0.016 (0.033)
Inflation Target $_{t t+k}$	0.364*** (0.108)	0.722*** (0.185)
Inflation Target $_{t t+k}$ * Real-Sector Dummy	-0.017 (0.052)	-0.238* (0.122)
Policy Surprise $_{i,t-1}$	-0.021 (0.019)	-0.036** (0.015)
Policy Surprise $_{i,t-1}$ * Real-Sector Dummy	-0.009 (0.026)	-0.011 (0.025)
Nom. Depreciation $_{t-1}$	0.034*** (0.003)	0.027*** (0.005)
Nom. Depreciation $_{t-1}$ * Real-Sector Dummy	-0.006*** (0.002)	-0.007* (0.004)
IPI Growth $_{t-2}$	0.037*** (0.013)	0.031*** (0.006)
IPI Growth $_{t-2}$ * Real-Sector Dummy	0.008*** (0.002)	-0.014*** (0.004)
Oil Price Growth $_{t-1}$	0.010*** (0.003)	0.008*** (0.003)
Oil Price Growth $_{t-1}$ * Real-Sector Dummy	-0.001 (0.001)	-0.001 (0.001)
Dummy _{Target Revision}	-0.094 (0.190)	-1.336*** (0.352)
Dummy _{Target Revision} * Real-Sector Dummy	-0.060 (0.074)	0.816*** (0.229)
Observations	7,712	7,473
R ²	0.797	0.682

Notes: This table shows the regression results of the baseline empirical model (equation (1)) for survey participants from financial and real sector. Real-Sector Dummy is a dummy variable that takes a value of 1 if the survey participant is from the real sector and takes 0 if it is from the financial sector. *, **, and *** represent statistical significance at levels of 10, 5, and 1 percent, respectively. Driscoll and Kraay (1998) standard errors are given in parentheses.

of inflation expectations by estimating moving-windows regressions and tracking the relevant coefficients in time.

We modify the baseline model (equation (1)) slightly to conduct the rolling regressions. The reason for using a different setup is the lack of variation in inflation targets since 2012. In other words, in our baseline model, one of the explanatory variables is constant during the last six years of the sample period. While this may not be a problem for the entire sample, it creates complications with short-horizon moving-window estimates. In order to circumvent this problem, we employ a modified version of the baseline model in equation (1), by simply replacing the *inflation-level* variables with the “gap” terms. Accordingly, our modified empirical model takes the following form:

$$\begin{aligned}
 (\pi_i^e - \pi^{target})_{t|t+k} &= \alpha_1 (\pi - \pi^{target})_{t-1} + \alpha_2 (\pi - \pi^{target})_{t-1}^{MA12} \\
 &\quad + \alpha_3 MP_{i,t-1}^{surprise} + \alpha_4 \Delta basket_{t-1} + \alpha_5 \Delta ipi_{t-2} \\
 &\quad + \alpha_6 \Delta oil_{t-1} + \mu_i + \varepsilon_{it}.
 \end{aligned} \tag{2}$$

Our transformed dependent variable is now k -month-ahead inflation expectations minus the corresponding inflation target, which we denote as $((\pi_i^e - \pi^{target})_{t|t+k})$. We call this variable “the credibility gap,” representing the gap between central bank’s k -period-ahead inflation target and private agents’ forecasts of inflation for the same horizon. Past inflation terms on the right-hand side are also transformed into the gap form. Instead of inflation levels, we use the gap between realized inflation and the target as explanatory variables. Accordingly, $(\pi - \pi^{target})_{t-1}$ shows the deviation of previous month’s inflation rate from the corresponding target and $(\pi - \pi^{target})_{t-1}^{MA12}$ denotes past 12-month moving average of this deviation. Here, once again we assume that survey participants, when constructing their forecasts, not only consider the most recent inflation figures, but also take into account an average of near history performance (represented by the MA12 term). Other explanatory variables are exactly in the same form as in equation (1), except that we dropped the target change dummy, as the new form of the dependent variable makes it redundant. Table A.2 in the appendix shows the regression results for the credibility gap. As expected, the coefficients are almost identical to the results of the level specification.

In order to track the time-varying behavior of inflation expectations, we run five-year moving-window regressions. Figure 7 shows the evolution of coefficients on (i) the sum of past inflation terms (previous month's inflation and 12-month average inflation), (ii) exchange rate depreciation, and (iii) central bank policy surprises at the individual level.

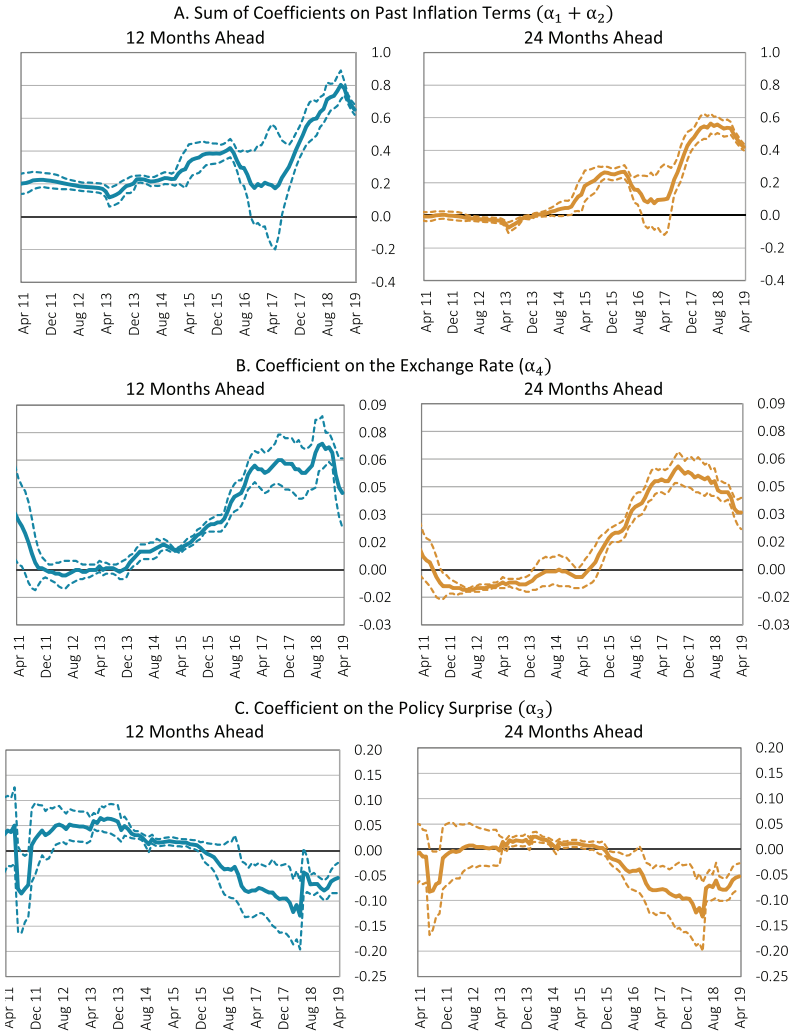
Several implications emerge from the rolling regression results. The sum of the coefficients on past inflation components exhibits a marked upward movement towards the end of the sample period (figure 7A). In other words, survey participants tend to attach increasingly higher weight to the previous inflation figures when forming expectations. Considering the significant upside breaches towards the end of the sample period, this finding is consistent with the negative sign of the interaction term in table 1. The anchoring role of inflation targets seems to have weakened as the gap between inflation and targets has widened.

Private forecasters attach higher weights to the past inflation in recent years, and the shift has become more noticeable after 2017—a period marked by persistent double-digit inflation. Given the sharp exchange rate depreciation of the Turkish lira towards the end of the sample period, these results are also consistent with the findings presented in table 2, which implies higher sensitivity of expectations to past inflation during depreciation periods.

The results depicted in figure 7B reveal that the relationship between exchange rate and inflation expectations has strengthened after 2013, which coincides with the persistent depreciation in the Turkish lira during this period. Higher inflation and inflation volatility, combined with the asymmetric pass-through effects may have altered the observed relationship between exchange rates and inflation expectations. Although the causality may run in both directions, this finding is notable, as it implies a stronger feedback between exchange rates and inflation expectations in driving the inflation process.

On the other hand, it is interesting to note that the upward trend in the sensitivity of expectations to past inflation and exchange rates seems to have partly reversed course towards the end of the sample period, as depicted by the decline in the coefficients in figure 7A and 7B during the recent period. These changes broadly coincide with a tighter monetary policy stance (the central bank increased

Figure 7. Five-Year Rolling Regressions for the Credibility Gap



Notes: Dates in the x-axis show the last month of the 60-month (five-year) rolling windows. Dashed lines show 90 percent confidence intervals with Driscoll and Kraay (1998) standard errors.

the base policy rate sharply in September 2018) and the adoption of a more conventional policy framework by mid-2018, although more observation is needed to make a firmer assessment on the drivers and significance of this behavioral shift.

Evolution of the coefficients on the monetary policy surprises across time provides useful insights regarding how monetary policy interacts with expectations under different policy frameworks. Under a conventional framework, a positive monetary policy shock would lower medium-term inflation expectations by signaling a tighter-than-expected policy stance. In fact, the full-sample estimations shown in table 1 and table A.2 reveal a negative and significant coefficient for the policy surprises. However, moving-window estimates depicted in figure 7C reveal that the coefficients showing the impact of monetary policy surprises on the inflation expectations vary across time, and these changes largely coincide with the shifts in the monetary policy framework. Adoption of an unconventional interest rate corridor policy in 2011 and the gradual exit from this framework after 2016 may explain some of the changing relationships. Between 2011 and 2015, the CBRT used a relatively complicated and high-frequency interest rate policy to smooth exchange rate fluctuations (Kara 2015). Moving-window regression coefficients suggest that, during this period, the response of the medium-term (two-year) inflation expectations to monetary policy surprises are insignificant and short-term (one-year) expectations respond with a wrong (positive) sign. This result makes sense because during this period, monetary policy surprises are likely to be perceived as short-term reactions to exchange rate volatility rather than a response to medium-term inflationary pressures. On the other hand, the sign of the policy surprise coefficient turns negative after 2016, following the attempts of gradually reverting to a more conventional monetary policy framework (figure 7C). With the normalization of monetary policy strategy towards the end of the sample period, a surprise tightening (easing) seems to be associated with a decrease (increase) in medium-term inflation expectations, as predicted by the conventional theory. Our unique data set including matched forecasts for inflation and the policy rate at the individual level, as well as the frequently changing nature of the background monetary policy framework, enables us to make these assessments with a reasonable precision.

The finding of an upside response of inflation expectations to tightening surprises in some occasions is not specific to our study. For example, Andrade and Ferroni (2018) argue that the “wrong sign” of the policy surprise coefficients in the case of the European Central Bank is due to the fact that policy surprises are perceived as news about future macroeconomic conditions, rather than a stronger or weaker commitment for the price stability objective. In our case, the economic agents may have perceived the high-frequency interest rate hikes as a signal of future exchange rate pressures during the period of unconventional interest rate corridor framework, which may have contaminated the relationship between policy surprises and inflation expectations.

Overall, the results suggest that the expectation dynamics have exhibited notable changes throughout the sample period, possibly associated with the underlying policy and economic performance.

One important question is whether the change in the actual inflation process mimicked the changes in the expectation dynamics. In order to contrast the pattern of changing expectation behavior with the inflation process itself, we have regressed actual inflation on lagged inflation and exchange rates along with similar control variables used in the empirical model for inflation expectations. The regression results are reported in figure 8 with five-year rolling windows. The coefficients on past inflation and the exchange rate depreciation in this regression rise sharply after 2017.¹⁷ More interestingly, a comparison of figure 7B with figure 8B suggests that the sensitivity of inflation expectations to exchange rates started to increase before the rise in the estimated exchange rate pass-through.¹⁸

Our analysis so far suggests that the behavioral shift in inflation expectations might be attributed to the performance of achieving the inflation objectives. A complementary possible explanation for the increased prominence of past inflation and exchange rates in

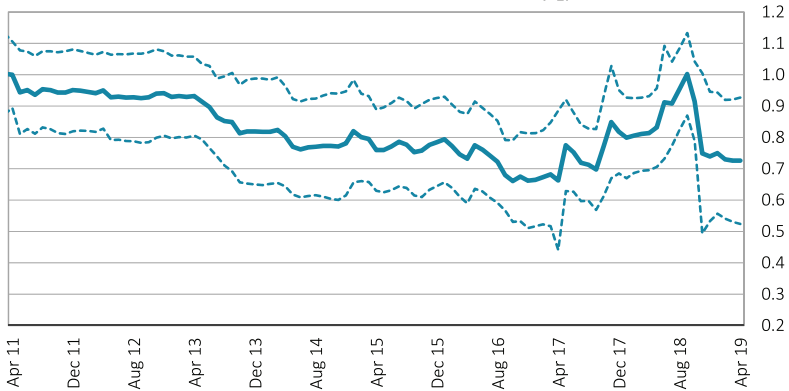
¹⁷A recent CBRT Inflation Report box presents similar findings using a time-varying parameter model of the inflation process developed in Kara, Ögünc, and Sarıkaya (2017). For details, see CBRT (2019).

¹⁸The structural break dates based on supremum Wald and Lagrange multiplier tests suggest that a significant shift in inflation dynamics has materialized around June 2016.

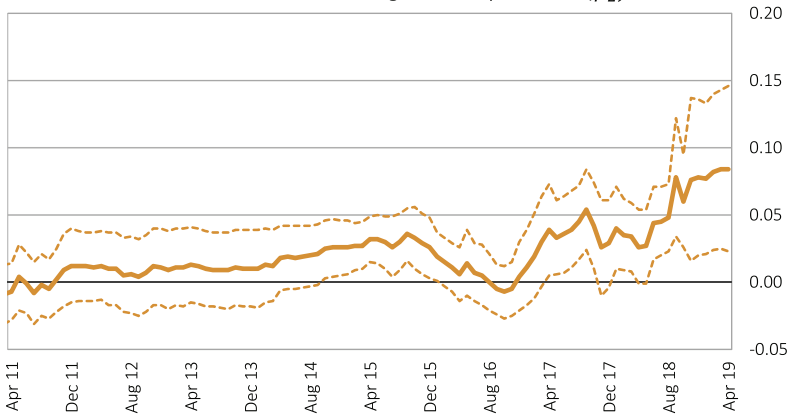
Figure 8. Coefficient of Lagged Inflation and Exchange Rate Depreciation in Explaining Annual Inflation (five-year rolling regressions)

$$\pi_t = \beta_1\pi_{t-1} + \beta_2\Delta basket_{t-1} + \beta_3ipi_{t-2} + \beta_4\Delta oil_{t-1} + \varepsilon_t$$

A. Coefficient of Lagged Inflation (β_1)



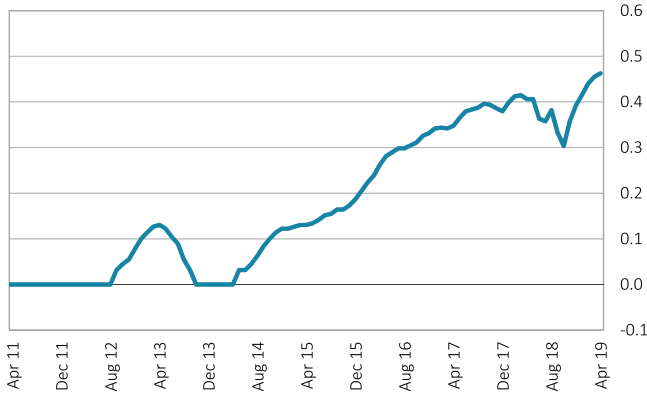
B. Coefficient of Exchange Rate Depreciation (β_2)



Notes: Dates in the x-axis show the last month of the 60-month (five-year) rolling windows. Dashed lines indicate 95 percent confidence intervals.

driving inflation expectations in recent years may be related to higher attentiveness of participants to these variables with the heightened volatility during the corresponding period (Coibion and Gorodnichenko 2015). In fact, figure 9 reveals that the individual-level correlation between expected exchange rate depreciation and

Figure 9. Co-movement of Individuals' 12-Month-Ahead Inflation Expectations and Expected TL Depreciation in 12 Months (five-year rolling correlation coefficients)



Source: CBRT.

expected inflation at the one-year horizon strengthened considerably towards the end of the sample period.

3.6 Robustness Analysis

In this subsection, we present some modifications and extensions to our baseline empirical model to see whether main findings remain robust against different specifications. To this end, we modify the main model in two dimensions: First, we use alternative definitions for key variables of interest, also considering the results of the direct feedback from participants. To this end, we add core inflation (instead of headline inflation), 24 months moving average of past inflation (instead of 12 months moving average), real effective exchange rate (instead of nominal exchange rate), and import prices (instead of oil prices). Second, we use additional explanatory variables that may be important in driving expectations dynamics implied by our feedback survey from respondents. Accordingly, we conduct alternative regressions by adding the following variables: (i) risk premium (monthly change in the Emerging Markets Bond Index (EMBI) spread), (ii) fiscal balance (primary budget balance to GDP ratio), (ii) money supply (rate of annual change

in M1), and (iv) wage inflation (annual growth in hourly labor cost index).¹⁹ Tables A.3 and A.4 in the appendix summarize the robustness results for one-year-ahead and two-year-ahead inflation expectations, respectively. Despite some minor discrepancies regarding the size of coefficients, our main conclusions are robust to all alternative specifications. The coefficients and the signs of the variables in the baseline model remain broadly the same. We also conduct moving-window estimates to see whether the main findings on the behavioral changes in expectations stay robust against alternative specifications. Moving-window estimates of the key parameters (past inflation, exchange rate, and policy surprise) are depicted in figures A.1 and A.2 in the appendix for the baseline and eight alternative models, with each column corresponding to a different specification. Although the size of the coefficients varies across models, their pattern and the evolution remain broadly robust. We still see parameters changing significantly through time associated with the background macroeconomic conditions and policy setting. The role of exchange rates and the past inflation terms seem to have strengthened through time. Policy rate surprises become insignificant during the implementation of the unconventional interest rate corridor between years 2011 and 2015, slightly gaining significance towards the end of the sample period.

Overall, our main results hold firmly across different specifications. Moreover, the robustness exercises show that direct feedback provided by the survey participants (summarized in figure 5) is highly consistent with the empirical results, confirming the usefulness of such feedback in supporting empirical research.

4. Concluding Remarks

We have investigated time-varying aspects of inflation expectation dynamics, seeking to explore how the behavior of expectations interacts with the policy setting and the macroeconomic performance. With its rapidly evolving macroeconomic and external conditions and highly volatile inflation process, the Turkish economy provides a

¹⁹Money supply series start from December 2006. Wage data are available at quarterly frequency and start from the first quarter of 2008, which is transformed into monthly series by assuming constant annual growth within the quarter.

genuine laboratory for exploring this question. Using individual-level data on a new survey of private forecasts, we document the changing dynamics of inflation expectations in response to the macroeconomic and policy environment. Our empirical model, which is built on direct feedback from survey participants, reflects a novel contribution to the related literature. The results imply that monitoring not only the *level* but also the *behavior* of inflation expectations may provide valuable insights for the formulation and the design of monetary policy.

The empirical evidence we provide on the expectations dynamics reveals that the behavior of inflation expectations may be highly sensitive to the underlying policy performance. Our results suggest that Turkish inflation expectations have been increasingly associated with the movements in exchange rates and past inflation through time, possibly associated with the changing macroeconomic landscape and the weakened anchoring power of the official targets through time. We support these findings by direct evidence from a recent feedback study conducted with the survey respondents, which reveals that towards the end of the sample period inflation target ceases to serve as an anchor in driving private inflation forecasts. These results indicate that the anchoring role of inflation targets can weaken considerably through time if the targets are breached for an extended period.

Overall, the Turkish experience offers important insights for other countries. The long-achieved credibility and strong anchoring of inflation targets across many emerging and advanced economies during the past decades should not be taken for granted. Credibility and the ability to shape expectations may shift quickly depending on the policy performance. The world experience and the literature so far has been on the benign examples where central banks gained credibility and inflation expectations became more anchored. Our study indicates that credibility may be gained yet lost quickly if promises are not delivered. The Turkish case, which shows that this may revert even after a period of successful inflation targeting, yields an important lesson for developing economies, which seem to be reverting to their previous ailments, and for developed economies, which face difficulties in raising inflation to their targets but have not suffered major credibility losses, yet.

Although our findings suggest that changes in the expectations formation process are related to the policy performance, we do not

attempt to provide concrete evidence on why the performance of inflation targeting was far from stellar. Explaining the fundamental factors driving the inflation target overshoots or exchange rate depreciations during our sample period is beyond the scope of this paper. It should be noted that for the Turkish case, the significant changes in the behavior of inflation expectations coincided with a period of heightened concerns on central bank instrument independence, which may have accelerated the behavioral shift in inflation expectations. In that sense, deeper research is needed to unveil the specific underlying mechanisms leading to changes in the expectations behavior. It would be particularly an interesting extension for future work to explore to what extent the changes in the expectations dynamics are driven by the perceptions of sliding external outlook as opposed to domestic factors including macro policy setting and the role of strong institutions.

Appendix. Robustness Regressions

Table A.1. Partial R²s for the Covariates in the Baseline Model of Table 1 (April 2006–April 2019)

	(1)	(2)
	$\pi_{i,t t+12}^e$	$\pi_{i,t t+24}^e$
π_{t-1}	0.201	0.078
π_{t-1}^{MA12}	0.110	0.049
$\pi_{t t+k}^{target}$	0.043	0.071
$MP_{i,t-1}^{surprise}$	0.002	0.003
$\Delta basket_{t-1}$	0.106	0.070
Δipi_{t-2}	0.034	0.023
Δoil_{t-1}	0.040	0.026
No. of Observations	8,207	7,947

Notes: Values in column 1 and column 2 show the square of partial correlation coefficients of the corresponding variable with the 12-month-ahead and 24-month-ahead inflation expectations, respectively. Partial correlation coefficients measure the strength of a relationship between the corresponding variable and inflation expectations, while controlling for the effect of other variables. All the correlation coefficients are significant at the 1 percent significance level.

**Table A.2. Drivers of Credibility Gap (expected deviation from the inflation targets)
(April 2006–April 2019)**

Dependent Variable: Credibility gap for participant i at time t ($\pi_i^e - \pi^{target}$) _{$i t+k$}		
$k =$	(1) 12-Month	(2) 24-Month
Target Deviation _{$t-1$}	0.362*** (0.065)	0.217*** (0.038)
MA12 Target Deviation _{$t-1$}	0.179*** (0.068)	0.078 (0.075)
Policy Surprise _{$t,t-1$}	-0.064*** (0.021)	-0.077*** (0.019)
Nom. Depreciation _{$t-1$}	0.037*** (0.008)	0.028*** (0.008)
IPI Growth _{$t-2$}	0.040** (0.019)	0.017* (0.010)
Oil Price Growth _{$t-1$}	0.013*** (0.003)	0.010*** (0.002)
Observations	8,182	7,943
R ²	0.759	0.632

Notes: *, **, and *** represent statistical significance at levels of 10, 5, and 1 percent, respectively. Driscoll and Kraay (1998) standard errors are given in parentheses.

Table A.3. Robustness Analysis for 12-Month-Ahead Inflation Expectations (full-sample regressions)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable: 12-month-ahead annual inflation expectations of participant i at time t ($\pi_{i,t t-12}^e$)									
CPI Inflation $_{t-1}$	0.335*** (0.029)		0.293*** (0.026)	0.387*** (0.035)	0.306*** (0.028)	0.335*** (0.029)	0.336*** (0.030)	0.348*** (0.026)	0.331*** (0.033)
Core Inflation $_{t-1}$		0.426*** (0.058)							
MA12 Inflation $_{t-1}$	0.371*** (0.042)	0.125*** (0.055)		0.350*** (0.039)	0.378*** (0.033)	0.370*** (0.041)	0.358*** (0.034)	0.377*** (0.033)	0.354*** (0.032)
MA24 Inflation $_{t-1}$			0.675*** (0.063)						
Inflation Target $_{t t+12}$	0.358*** (0.098)	0.573*** (0.120)	0.274** (0.107)	0.340*** (0.106)	0.373*** (0.099)	0.351*** (0.102)	0.217*** (0.076)	0.296*** (0.082)	0.289*** (0.108)
Policy Surprise $_{i,t-1}$	-0.025* (0.014)	-0.083*** (0.012)	-0.038** (0.016)	-0.009 (0.018)	-0.026 (0.018)	-0.028 (0.017)	-0.024* (0.012)	-0.030** (0.014)	-0.021* (0.011)
Nom. Depreciation $_{t-1}$	0.034*** (0.002)	0.019*** (0.006)	0.035*** (0.004)		0.039*** (0.002)	0.033*** (0.002)	0.028*** (0.003)	0.027*** (0.003)	0.022*** (0.003)
Real Depreciation $_{t-1}$				-0.033*** (0.003)					

(continued)

Table A.3. (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable: 12-month-ahead annual inflation expectations of participant i at time t ($\pi_{i,t t+12}^e$)									
IPI Growth $_{t-2}$	0.039*** (0.013)	0.022 (0.017)	0.047*** (0.014)	0.027** (0.013)	0.029*** (0.010)	0.037** (0.015)	0.030** (0.014)	0.029*** (0.010)	0.051*** (0.010)
Oil Price Growth $_{t-1}$	0.010*** (0.002)	0.016*** (0.001)	0.007*** (0.001)	0.009*** (0.002)	0.010*** (0.002)	0.010*** (0.002)	0.008** (0.002)	0.010*** (0.002)	0.011*** (0.003)
Import Price Growth $_{t-1}$			0.037*** (0.008)						
DummyTarget Revision	-0.087 (0.177)	-1.263*** (0.396)	-0.341*** (0.113)	-0.315** (0.156)	-0.233 (0.183)	-0.060 (0.194)	0.518** (0.203)	0.032 (0.126)	-0.438** (0.176)
Δ EMBI $_{t-1}$						0.001 (0.001)			
Prim. Balance (%GDP) $_{t-2}$							-0.212*** (0.016)		
Money Growth $_{t-1}$								0.024* (0.012)	
Wage Growth $_{t-3}$									0.078*** (0.015)
Observations	8,182	8,182	8,182	8,182	8,182	8,182	8,182	7,738	7,029
R ²	0.799	0.739	0.800	0.790	0.800	0.800	0.816	0.813	0.838

Notes: *, **, and *** represent statistical significance at levels of 10, 5, and 1 percent, respectively. Driscoll and Kraay (1998) standard errors are given in parentheses.

Table A.4. Robustness Analysis for 24-Month-Ahead Inflation Expectations (full-sample regressions)

	Dependent Variable: 24-month-ahead annual inflation expectations of participant i at time t ($\pi_{i,t t+24}^e$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CPI Inflation $_{t-1}$	0.205*** (0.015)		0.181*** (0.014)	0.247*** (0.015)	0.184*** (0.012)	0.205*** (0.015)	0.205*** (0.015)	0.220*** (0.016)	0.210*** (0.017)
Core Inflation $_{t-1}$		0.219*** (0.038)							
MA12 Inflation $_{t-1}$	0.229*** (0.049)	0.110 (0.070)		0.213*** (0.039)	0.232*** (0.046)	0.229*** (0.048)	0.207*** (0.018)	0.227*** (0.042)	0.216*** (0.029)
MA24 Inflation $_{t-1}$			0.419*** (0.059)						
Inflation Target $_{t t+12}$	0.665*** (0.108)	0.659*** (0.083)	0.647*** (0.113)	0.672*** (0.120)	0.678*** (0.113)	0.666*** (0.111)	0.296*** (0.059)	0.503*** (0.087)	0.076 (0.077)
Policy Surprise $_{i,t-1}$	-0.042*** (0.008)	-0.082*** (0.009)	-0.049*** (0.009)	-0.026*** (0.010)	-0.043*** (0.010)	-0.042*** (0.009)	-0.035*** (0.006)	-0.044*** (0.008)	-0.041*** (0.006)
Nom. Depreciation $_{t-1}$	0.025*** (0.003)	0.021*** (0.006)	0.026*** (0.003)		0.029*** (0.003)	0.025*** (0.003)	0.020*** (0.002)	0.019*** (0.003)	0.016*** (0.003)
Real Depreciation $_{t-1}$				-0.023*** (0.003)					
IPI Growth $_{t-2}$	0.030*** (0.004)	0.017** (0.009)	0.037*** (0.006)	0.021*** (0.004)	0.022*** (0.004)	0.030*** (0.005)	0.015*** (0.005)	0.018*** (0.004)	0.021*** (0.006)
Oil Price Growth $_{t-1}$	0.008*** (0.002)	0.011*** (0.001)	0.007*** (0.001)	0.008*** (0.002)	0.008*** (0.002)	0.008*** (0.002)	0.006*** (0.001)	0.008*** (0.001)	0.008*** (0.001)

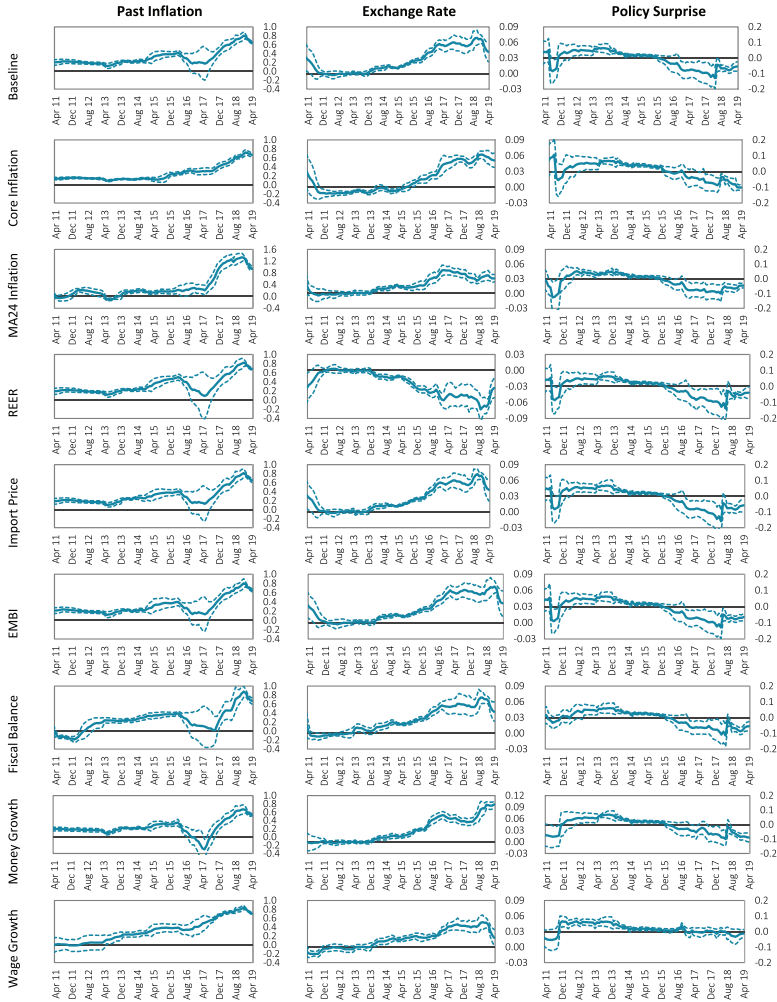
(continued)

Table A.4. (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable: 24-month-ahead annual inflation expectations of participant i at time t ($\pi_{i,t t+24}^e$)									
Import Price Growth $_{t-1}$					0.027*** (0.004)				
Dummy _{Target Revision}	-1.090*** (0.197)	-1.366*** (0.248)	-1.329*** (0.227)	-1.311*** (0.221)	-1.160*** (0.227)	-1.093*** (0.201)	0.076 (0.182)	-0.790*** (0.153)	-0.410*** (0.125)
ΔEMBI_{t-1}						-0.000 (0.001)			
Prim. Balance (%GDP) $_{t-2}$							-0.287*** (0.018)	0.024** (0.011)	
Money Growth $_{t-1}$									0.038*** (0.012)
Wage Growth $_{t-3}$									6.849
Observations	7,943	7,943	7,943	7,943	7,943	7,943	7,943	7,531	0.721
R ²	0.688	0.644	0.688	0.677	0.685	0.688	0.734	0.704	

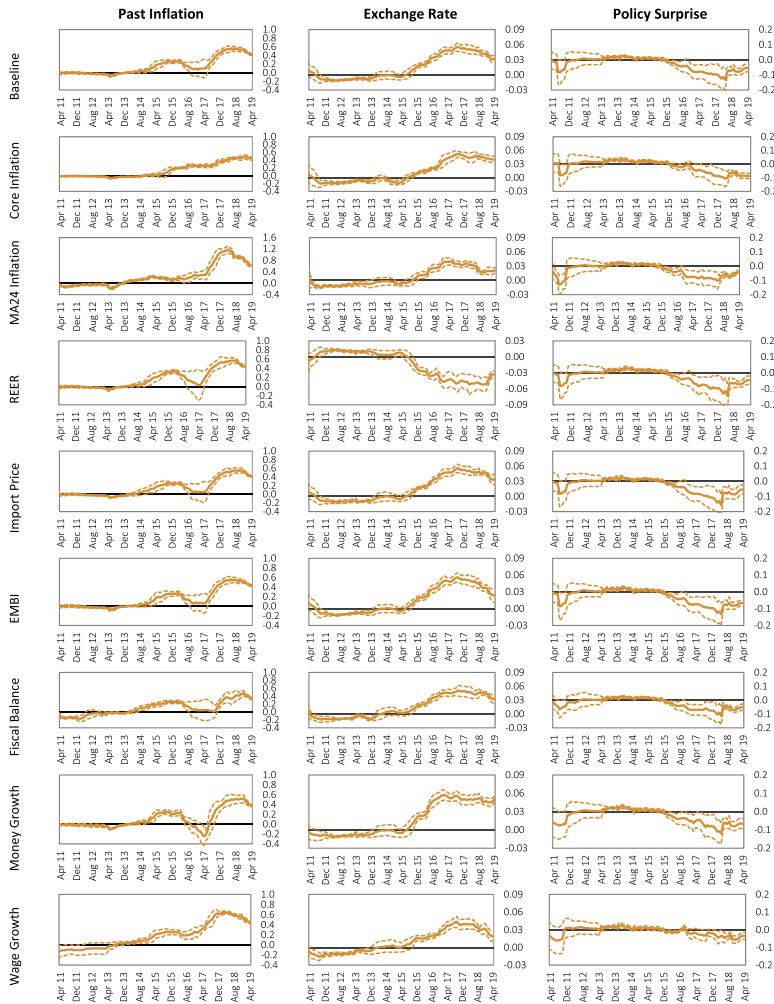
Notes: *, **, and *** represent statistical significance at levels of 10, 5, and 1 percent, respectively. Driscoll and Kraay (1998) standard errors are given in parentheses.

Figure A.1. Robustness Analysis for the Evolution of the Coefficients: 12-Month-Ahead Inflation Expectations



Notes: The graphs show five-year rolling window estimates of the coefficients on past inflation, exchange rate, and policy surprises for the baseline model in equation (2) and its modifications with additional variables listed in the first column. Dashed lines show 90 percent confidence intervals with Driscoll and Kraay (1998) standard errors.

Figure A.2. Robustness Analysis for the Evolution of the Coefficients: 24-Month-Ahead Inflation Expectations



Notes: The graphs show five-year rolling window estimates of the coefficients on past inflation, exchange rate, and policy surprises for the baseline model in equation (2) and its modifications, with additional variables listed in the first column. Dashed lines show 90 percent confidence intervals with Driscoll and Kraay (1998) standard errors.

References

- Altavilla, C., and D. Giannone. 2017. "The Effectiveness of Non-standard Monetary Policy Measures: Evidence from Survey Data." *Journal of Applied Econometrics* 32 (5): 952–64.
- Andrade, P., and F. Ferroni. 2018. "Delphic and Odyssean Monetary Policy Shocks: Evidence from the Euro Area." Working Paper No. 2018-12, Federal Reserve Bank of Chicago.
- Armantier, O., G. Topa, W. Van der Klaauw, and B. Zafar. 2017. "An Overview of the Survey of Consumer Expectations." *Economic Policy Review* (Federal Reserve Bank of New York) 23 (2): 51–72.
- Başkaya, Y. S., E. Gülşen, and H. Kara. 2012. "Inflation Expectations and Central Bank Communication in Turkey." *Central Bank Review* 12 (2): 1–10.
- Bernanke, B. S., and K. N. Kuttner. 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *Journal of Finance* 60 (3): 1221–57.
- Binici, M., H. Kara, and P. Özlü. 2019. "Monetary Transmission with Multiple Policy Rates: Evidence from Turkey." *Applied Economics* 51 (17): 1869–93.
- Blanchflower, D. G., and C. MacCoille. 2009. "The Formation of Inflation Expectations: An Empirical Analysis for the UK." NBER Working Paper No. 15388.
- Blinder, A., M. Ehrmann, M. Fratzscher, J. De Haan, and D. Jansen. 2008. "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence." *Journal of Economic Literature* 46 (4): 910–45.
- Boneva, L., J. Cloyne, M. Weale, and T. Wieladek. 2016. "The Effect of Unconventional Monetary Policy on Inflation Expectations: Evidence from Firms in the United Kingdom." *International Journal of Central Banking* 12 (3, September): 161–95.
- Brito, R. D., and B. Bystedt. 2010. "Inflation Targeting in Emerging Economies: Panel Evidence." *Journal of Development Economics* 91 (2): 198–210.
- Carvalho, F. A., and A. Minella. 2012. "Survey Forecasts in Brazil: A Prismatic Assessment of Epidemiology, Performance, and Determinants." *Journal of International Money and Finance* 31 (6): 1371–91.

- Celasun, O., R. G. Gelos, and A. Prati. 2004. "Obstacles to Disinflation: What is the Role of Fiscal Expectations?" *Economic Policy* 19 (40): 441–81.
- Central Bank of the Republic of Turkey (CBRT). 2019. "A Model-Based Analysis of Recent Inflation Developments." *Inflation Report 2018-IV*, Box 3.1, 36–39.
- Christiano, L. J., M. Eichenbaum, and C. L. Evans. 1999. "Monetary Policy Shocks: What Have We Learned and to What End?" In *Handbook of Macroeconomics*, Vol. 1, ed. J. B. Taylor and M. Woodford, 65–148 (chapter 2). Amsterdam: Elsevier.
- Ciccarelli, M., J. A. Garcia, and C. Montes-Galdón. 2017. "Unconventional Monetary Policy and the Anchoring of Inflation Expectations." Working Paper No. 1995, European Central Bank.
- Coibion, O., and Y. Gorodnichenko. 2015. "Inflation Expectations in Ukraine: A Long Path To Anchoring?" *Visnyk of the National Bank of Ukraine* (233): 6–23.
- Coibion, O., Y. Gorodnichenko, and R. Kamdar. 2018. "The Formation of Expectations, Inflation, and the Phillips Curve." *Journal of Economic Literature* 56 (4): 1447–91.
- Coibion O., Y. Gorodnichenko, S. Kumar, and M. Pedemonte. 2020. "Inflation Expectations as a Policy Tool?" *Journal of International Economics* 124 (May): Article 103297.
- Cortes, G., and C. A. C. Paiva. 2017. "Deconstructing Credibility: The Breaking of Monetary Policy Rules in Brazil." *Journal of International Money and Finance* 74 (June): 31–52.
- Çiçek, S., C. Akar, and E. Yücel. 2011. "Türkiye’de Enflasyon Beklentilerinin Çapalanması ve Güvenilirlik." *ktisat letme ve Finans* 26 (304): 37–55.
- Driscoll, J. C., and A. C. Kraay. 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics* 80 (4): 549–60.
- Eminidou, S., M. Zachariadis, and E. Andreou. 2020. "Inflation Expectations and Monetary Policy Surprises." *Scandinavian Journal of Economics* 122 (1): 306–39.
- Galí, J. 2008. *Monetary Policy, Inflation and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton, NJ: Princeton University Press.
- Gertler, M., and P. Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7 (1): 44–76.

- Gülşen, E., and H. Kara. 2019. "Measuring Inflation Uncertainty in Turkey." *Central Bank Review* 19 (2): 33–43.
- Gürkaynak, R. S., A. Levin, and E. Swanson. 2010. "Does Inflation Targeting Anchor Long-Run Inflation Expectations? Evidence from the U.S., UK, and Sweden." *Journal of the European Economic Association* 8 (6): 1208–42.
- Gürkaynak, R. S., B. Sack, and E. T. Swanson. 2005. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1 (1, May): 55–93.
- Inoue, A., L. Kilian, and F. C. Kiraz. 2009. "Do Actions Speak Louder than Words? Household Expectations of Inflation Based on Micro Consumption Data." *Journal of Money, Credit and Banking* 41 (7): 1331–63.
- Kara, H. 2015. "Interest Rate Corridor and the Monetary Policy Stance." CBT Research Notes in Economics No. 1513.
- Kara, H., and H. Küçük. 2010. "Inflation Expectations in Turkey: Learning to be Rational." *Applied Economics* 42 (21): 2725–42.
- Kara, H., F. Ögünç, and Ç. Sarıkaya. 2017. "Inflation Dynamics in Turkey: A Historical Accounting." CBT Research Notes in Economics No. 1703.
- Keane M., and D. Runkle. 1990. "Testing the Rationality of Price Forecasts: New Evidence from Panel Data." *American Economic Review* 80 (4): 714–35.
- Köse, M. A., H. Matsuoka, U. Panizza, and D. Vorisek. 2019. "Inflation Expectations: Review and Evidence." Policy Research Working Paper No. 8785, World Bank.
- Mankiw, N. G., R. Reis, and J. Wolfers. 2004. "Disagreement about Inflation Expectations." *NBER Macroeconomics Annual 2003*, Vol. 18, ed. M. Gertler and K. Rogoff, 209–48 (chapter 4). The MIT Press.
- Manski, C. 2004. "Measuring Expectations." *Econometrica* 72 (5): 1329–76.
- Marinovic, I., M. Ottaviani, and P. N. Sørensen. 2013. "Forecasters' Objectives and Strategies." In *Handbook of Economic Forecasting*, Vol. 2B, ed. G. Elliott and A. Timmermann, 689–720 (chapter 12). Amsterdam: North-Holland.
- Mehrotra, A., and J. Yetman. 2018. "Decaying Expectations: What Inflation Forecasts Tell Us about the Anchoring of Inflation

- Expectations.” *International Journal of Central Banking* 14 (5, December): 55–101.
- Pedersen, M. 2015. “What Affects the Predictions of Private Forecasters? The Role of Central Bank Forecasts in Chile.” *International Journal of Forecasting* 31 (4): 1043–55.
- Pesaran, M. H., and M. Weale. 2006. “Survey Expectations.” In *Handbook of Economic Forecasting*, Vol. 1, ed. G. Elliott, C. W. J. Granger, and A. Timmermann, 715–76 (chapter 14). Amsterdam: North-Holland.
- Sims, C. A. 2003. “Implications of Rational Inattention.” *Journal of Monetary Economics* 50 (3, April): 665–90.
- Strohsal, T., R. Melnick, and D. Nautz. 2016. “The Time-Varying Degree of Inflation Expectations Anchoring.” *Journal of Macroeconomics* 48 (June): 62–71.
- Yellen, J. 2016. “Macroeconomic Research After the Crisis.” Speech given at “The Elusive ‘Great’ Recovery: Causes and Implications for Future Business Cycle Dynamics” 60th annual economic conference sponsored by the Federal Reserve Bank of Boston, Boston, Massachusetts, October 14.

Which Credit Gap Is Better at Predicting Financial Crises? A Comparison of Univariate Filters*

Mathias Drehmann and James Yetman
Bank for International Settlements

The credit gap, defined as the deviation of the credit-to-GDP ratio from a one-sided HP-filtered trend, is a useful indicator for predicting financial crises. Basel III therefore suggests that policymakers use it as part of their countercyclical capital buffer frameworks. Hamilton (2018), however, argues that you should never use an HP filter, as it results in spurious dynamics, has endpoint problems, and its typical implementation is at odds with its statistical foundations. Instead he proposes the use of linear projections. Some have also criticized the normalization by GDP, since gaps will be negatively correlated with output. We agree with these criticisms. Yet, in the absence of clear theoretical foundations, *all* proposed gaps are but indicators. It is therefore an empirical question which measure performs best as an early-warning indicator for crises. We run a horse race using expanding samples on quarterly data from 1970 to 2017 for 41 economies. We find that credit gaps based on linear projections in real time perform poorly when based on country-by-country estimation, and are subject to their own endpoint problem. But when we estimate as a panel, and impose the same coefficients on all economies, linear projections perform marginally better than the baseline credit-to-GDP gap, with somewhat larger improvements

*We thank an anonymous referee, Claudio Borio, Itamar Caspi, Stijn Claessens, Jim Hamilton, Daniel Leff, Moritz Schularick, and seminar participants at the BIS and the Reserve Bank of New Zealand for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank for International Settlements. Author contact: Drehmann: Centralbahnplatz 2, CH-4002 Basel, Switzerland, mathias.drehmann@bis.org; Yetman: Representative Office for Asia and the Pacific, 78th Floor, Two IFC, 8 Finance Street, Central, Hong Kong, james.yetman@bis.org.

concentrated in the post-2000 period and for emerging market economies. The practical relevance of the improvement is limited, though. Over a 10-year horizon, policymakers could expect one less wrong call on average.

JEL Codes: E44, G01.

1. Introduction

Excessive credit growth has long been recognized as integral to financial booms and busts (Minsky 1982; Kindleberger 2000). However, what constitutes growth being “excessive” remains undefined. Borio and Lowe (2002) propose a credit-to-GDP gap measured by the deviations of the credit-to-GDP ratio from a one-sided Hodrick-Prescott (HP) filter with a large smoothing parameter (400,000 for quarterly data). Borio and Drehmann (2009), Drehmann et al. (2010), and Drehmann, Borio, and Tsatsaronis (2012) revisit the gap in light of the crisis and do extensive comparisons of its early-warning indicator (EWI) properties for systemic banking crises with other variables. They identify the credit-to-GDP gap as the best single EWI across those that they examine. Their work underpins the choice of the Basel Committee for Banking Supervision (BCBS) to single out the credit-to-GDP gap as a useful guide for setting countercyclical capital buffers (BCBS 2010b).

But the credit-to-GDP gap is only one possible indicator of excessive credit growth.¹ Following the work of Jordà, Schularick, and Taylor (2011), for example, the academic literature has mainly relied on medium-term growth rates in credit-to-GDP. In addition, the HP-based gap has been challenged on conceptual grounds. We address two such challenges here.

Most importantly, many have criticized the use of the HP filter to derive the gap. It has long been known that the HP filter has serious problems. These are succinctly summarized by Hamilton (2018). In particular, the HP filter results in spurious dynamics that

¹Since Borio and Lowe (2002), Bank for International Settlements (BIS) authors have always been careful in emphasizing this point (for an overview see Drehmann and Tsatsaronis 2014). This is also one reason why there is no mechanical link between the credit gap and the countercyclical capital buffer under the Basel III rules (BCBS 2010b).

are not found in the underlying data, results in filtered data with properties that differ between the middle and ends of the sample, and its typical implementation is at odds with its statistical foundations.² Hamilton therefore concludes that you should never use the HP filter for any purpose, including for deriving credit-to-GDP gaps. He proposes the use of linear projections as an alternative to derive deviations from trends.

In addition, some authors have criticized the use of GDP to normalize the level of credit in the economy. For instance, Repullo and Saurina (2011) point out that the credit-to-GDP gap will tend to be negatively correlated with GDP, and its use could exacerbate the procyclicality of macroprudential policy. Similar problems were highlighted by the Basel Committee (BCBS 2010b). Real credit per capita has been proposed as an alternative measure to overcome this potential drawback.

From a conceptual perspective, we agree with these criticisms. But, in the absence of clear theoretical foundations, *any* proposed gap measure is nothing more than an indicator, including when derived with more sophisticated empirical methods.³ What should matter to policymakers is the relative performance of different possible measures, which can be assessed empirically.

In this paper, we therefore run a horse race between different proxies for excessive credit. Given that excessive credit is unobservable, we assess performance based on how well different credit gaps predict systemic banking crises. Performance is judged by the “area under the curve” (AUC), a summary measure of its predictive power. And we focus on (quasi) real-time information, which is the relevant case for policymakers who can only use the information they have available at each point in time to predict a crisis.⁴

²Relatedly, Edge and Meisenzahl (2011) document a large difference between real-time and full-sample estimates of credit-to-GDP gaps due to the endpoint problem of the HP filter.

³For instance, Buncic and Melecky (2014) and Juselius and Drehmann (2020) derive credit gaps based on multivariate VARs.

⁴We shorten “quasi real time” to “real time” for the remainder of the paper. Our real-time estimates use only data up to time t to estimate gaps at time t , with the sample expanding with each observation. But the data we use are those available at the time of estimation, rather than those available at time t , and hence are not truly “real time.”

To keep the analysis concise, we split it into two parts. First, we compare different possible formulations of the linear projection, given the lack of exploratory work elsewhere. We examine a wide range of different combinations of lags in the underlying regression, and also consider projections based on equations estimated both economy-by-economy and in a panel, where the coefficients on the lags are constrained to be the same for all economies.

In a second step, we select the best performing of these linear projections and compare it against two alternative means of deriving “gaps”: the HP trend and 20-quarter changes in credit. For each of these measures, we consider two means of normalizing credit, either by using nominal GDP or by transforming it into real-credit-per-capita terms.

The key finding for the different ways to derive projection gaps is that it is crucial to estimate the underlying linear equation as a panel instead of running economy-by-economy regressions. The panel approach results in a material improvement in performance across many forecast horizons and subsample specifications. The analysis also points to a potential “endpoint” problem of the linear projection gap, especially for small samples. If we compare the forecast performance of the full sample versus the real-time gaps, the performance of the real-time gaps in the economy-by-economy specification is much weaker. The reason is that during a credit boom—for example, in the early 2000s in the United States—the estimated coefficients increase in real time so that the residuals that the projection gap is based on don’t increase sharply, and are hence less likely to signal the impending crisis.

The panel helps to alleviate this endpoint problem. More generally, it points to the benefits of including international data in the assessment of credit gaps for individual economies: perhaps because of the relative rarity of financial crises, there are material gains in using the experiences of other economies to calibrate and assess early-warning indicators.

For practical purposes, it is also interesting to note that different lag structures in the linear projections have a limited effect on crisis prediction performance, provided the included lags are sufficiently long (generally 20 or more quarters).

When we compare the alternative approaches to generate gaps, two findings stand out. First, normalizing by GDP results in superior

forecast performance over normalizing by the population. Second, while the estimated projection gap generally has the highest AUC, differences in performance relative to a gap derived by an HP filter or 20-quarter changes tend to be quantitatively small, albeit in many cases statistically significant. Larger differences are only found for the post-2000 period and for emerging market economies.

But despite the statistical results, differences between different methods to derive gaps (at least when normalized by GDP) are not meaningful from a practical perspective. Dealing with the inherent uncertainty in identifying credit booms is more important by an order of magnitude. Across the different specification, and independent of the gap method, around 30 percent of signals are incorrect. And the higher AUC of the projection gap relative to an HP-filtered gap results in issuing 2–3 percentage points less incorrect signals in normal times. If policymakers would mechanically follow these gaps this would imply that, over a 10-year period, they could expect that the indicators would give wrong signals for around three years, independent of the gap they chose. Over the same period, the 2–3 percentage points difference of fewer wrong calls in normal times for the projection GDP gap relative to the HP GDP gap amounts to making the right call in just one additional quarter.

Addressing the underlying uncertainty about predicting crises, rather than the choice between these indicators, is therefore the key challenge. One possible source of improvement is to take a broader range of indicators into account. We do not do so in this paper, since we wish to focus on the debate about different methods to derive credit gaps as one fundamental component in early-warning indicator models, in light of its importance in the Basel III framework.⁵ Therefore, alternative methods, including those focused on multivariate measures, are beyond the scope of this paper.⁶ But even in

⁵The Basel III framework recognizes that the credit-to-GDP gap can only be a starting point of discussions about countercyclical capital buffers, as authorities should consider all available information (BCBS 2010b).

⁶Multivariate measures have been shown to have the potential to improve forecast performance, starting with Borio and Lowe (2002). Band-pass filters have been used in both univariate (Aikman, Haldane, and Nelson 2015) and multivariate (Drehmann, Borio, and Tsatsaronis 2011) contexts. Galati et al. (2016) extract a financial cycle using a multivariate unobserved-components model on the credit-to-GDP ratio, total credit, and house prices for six economies, and find

these cases, indicators provide incorrect signals, requiring judgment in practice and the recognition that policymaking based on these indicators is fraught with uncertainty.

In the next section, we outline the two challenges to the HP credit gap measure that we examine. Section 3 contains our methodology for comparing the different measures in light of the objective, and section 4 introduces the data. Section 5 provides a detailed analysis of the performance of different linear-projection based gaps, and section 6 compares the best of these against alternative measures. Robustness exercises are discussed in section 7. In section 8, we consider the practical implications of the differences before we conclude.

2. Critiques of the Baseline Credit Gap

Our baseline credit gap was proposed by Borio and Lowe (2002). They suggested measuring the credit gap as deviations of the credit-to-GDP ratio from a one-sided Hodrick-Prescott (HP) filter with a large smoothing parameter (400,000 for quarterly data). This measure has been subject to a number of criticisms. Here we outline two prominent ones: namely that the normalization is problematic, and the HP filter has undesirable properties.

In order to turn the nominal level of credit into a magnitude that is comparable both across time and across countries, it must be normalized in some manner. In our baseline measure, the normalization is to divide nominal credit by nominal GDP. Repullo

that the resulting medium-term cycles vary in terms of length and amplitude across countries and over time. Rünstler and Vlekke (2018) conduct a similar exercise and identify medium-term cycles in credit volumes that are linked to GDP performance at longer frequencies than business cycles. Other recent examples include Schüler, Hiebert, and Peltonen (2015), who find that a financial cycle based on the common frequencies of credit and asset prices outperforms the credit-to-GDP gap in predicting systemic banking crises at horizons of one-to-three years; Aldasoro, Borio, and Drehmann (2018), who show that combining various indicators of excessive debt with property prices can help to improve financial crisis prediction; Alessi and Detken (2018), who use “random forest” machine learning methods based on a number of economic and financial indicators and find that this outperforms a logit model based on the same explanatory variables in terms of out-of-sample performance; and Lang et al. (2019), who develop a combined indicator that captures risks stemming from domestic credit, real estate markets, asset prices, and external imbalances and that outperforms univariate early-warning indicators.

and Saurina (2011) suggest that this could be problematic, since it would suggest reducing capital requirements when GDP growth is high and increasing them when GDP growth is low, hence exacerbating the procyclicality of regulations related to bank capital.⁷ As discussed, this was already identified as a potential problem by the Basel Committee (2010b), which identified it as one of the reasons why policymakers' judgment is necessary when setting the countercyclical capital buffer. Jordà, Schularick, and Taylor (2017) and Richter, Schularick, and Wachtel (2017) use real credit per capita as their measure of normalized credit instead.

The other key component to measuring a credit gap is the definition of the gap—or, equivalently, defining the trend against which credit will be compared.

Following the original work by Borio and Lowe (2002), the long-term trend of the credit-to-GDP ratio is often calculated by means of a one-sided (i.e., real-time) HP filter. The filter is run in quasi real time, i.e., recursively, with an expanding sample each period. Thus, a trend calculated for, say, end-1998 only takes account of information up to 1998 even if this calculation is done in 2018. The HP filter also uses a much larger smoothing parameter—400,000 for quarterly data—than the one employed in the business cycle literature. This choice can be rationalized by the observation that credit cycles are on average about four times longer than standard business cycles and crises tend to occur once every 20–25 years (Drehmann et al. 2010).⁸

Hamilton (2018) points out some serious potential shortcomings with the HP filter in general—in particular, that

- (i) it produces spurious dynamics that are not based on the underlying data-generating process;

⁷Also see the discussion in Jordà (2011).

⁸Hodrick and Prescott (1997) set λ equal to 1,600 for extracting business cycles in quarterly data. Ravn and Uhlig (2002) show that, for series of other frequencies (daily, annual, etc.), it is consistent to set λ equal to 1,600 multiplied by the fourth power of the observation frequency ratio, implying λ equal to 400,000 if credit cycles are four times longer than business cycles. Empirically, λ equal to 400,000 also delivers the credit-to-GDP gap with the best forecasting performance (Drehmann, Borio, and Tsatsaronis 2011).

- (ii) the dynamics at the ends of the sample differ from those in the middle;⁹ and
- (iii) the standard implementation of the HP filter stands at stark odds from its statistical foundations.

To avoid these drawbacks, Hamilton suggests an alternative using a “linear projection” based on estimating the equation:

$$y_{t+h} = \beta_0 + \sum_{j=1}^J \beta_j y_{t+j-1} + \nu_{t+h}. \quad (1)$$

The estimated residual from this equation, ν_{t+h} , is the projection gap that will be assessed as a predictor of financial crises. Richter, Schularick, and Wachtel (2017) implement this method, but with one alteration: they normalize the residuals by their standard deviation, (σ_ν) , to produce the projection gap.

Hamilton suggests that including four lags ($J = 4$) and a value of h corresponding to five years (i.e., $h = 20$ with quarterly data) for applications to debt (or credit) cycles may be appropriate. But given that the baseline HP-filter-based credit-to-GDP gap has already been carefully tested with different assumptions about the smoothing parameter (see discussion above), we first examine a range of possible formulations of the linear projection model, with varying numbers and lengths of lags, to see how sensitive the results are. We also compare the results when we estimate the underlying equation economy-by-economy versus in a panel with the β_j 's constrained to be the same for all economies, while allowing separate fixed effects (β_0 's) for each economy.

An alternative approach that we also examine is to detrend by computing growth rates. Taking the 20-quarter change in credit/GDP or real credit per capita provides a filter-free way of extracting a credit gap measure. This approach has been used, for

⁹The baseline credit gaps measure uses a one-sided filter, with observations added recursively. On the one hand, this means that we are never comparing an observation from the middle of the sample with one from the end, mitigating the second critique. On the other hand, given that the gaps are taken from samples of different sizes, their properties could still vary.

example, in Jordà, Schularick, and Taylor (2011) and Jordà et al. (2017).

In the following section, we outline the methodology to assess predictive performance that we use for the horse race between different measures of the credit gap to see how they compare.

3. Assessing Predictive Performance

As discussed in the introduction, all proposed gaps are intended to be indicators of excessive credit growth. In line with a long research tradition, we judge performance by how well the different measures predict systemic banking crises.

We follow the literature and use the area under the ROC curve (AUC) as a statistical measure to judge forecast performance.¹⁰ It is a very intuitive measure. To fix ideas, assume a very simple economy that is in one of two states: $S = 0$, or $S = 1$. States are not directly observable, but a gap measure, G , carries imperfect information about the current state. In particular, the higher the value of G , the more likely it is that $S = 1$. In an ideal situation, there would be a threshold θ^i such that, if $G > \theta^i$, we would know that $S = 1$ (and $S = 0$ for $G \leq \theta^i$). But, if the signal is noisy, there is a tradeoff between the rate of true positives, $TPR[S(\theta^i)] = P(G > \theta^i | S = 1)$, and the rate of false positives, $FPR[S(\theta^i)] = P(G > \theta^i | S = 0)$.¹¹ For very low values of the threshold, for instance, the TPR will be close to one, but the same will also hold for FPR. We therefore look over all thresholds θ^i . And the mapping from FPR to TPR for all θ^i gives the ROC curve.

The area under this curve, the AUC, can interpreted as the likelihood that the distribution of G when $S = 1$ is stochastically larger

¹⁰ROC stands for receiver operating characteristic. The somewhat awkward name goes back to its original use of trying to differentiate noise from signals of radars during World War II. Since then it has been used in many other sciences (e.g., Swets and Pickett 1982). Over the last 10 years it has become increasingly popular in the context of crises or recession predictions, following in particular the work of Oscar Jordà (e.g., Berge and Jordà 2011, or Jordà, Schularick, and Taylor 2011).

¹¹The FPR and the complement of the TPR correspond to the familiar type II and type I errors.

than when $S = 0$. It is a convenient and interpretable summary measure of the signaling quality. A completely uninformative indicator has an AUC of 0.5. Correspondingly, the AUC for the perfect indicator equals 1. The AUC of an informative indicator falls in between and is statistically different from 0.5. For two competing indicators, G_1 and G_2 , it is also easy to test whether $AUC(G_1)$ is equal to $AUC(G_2)$ by using a Wald test.

We estimate the AUC nonparametrically with Stata. Standard errors are bootstrapped using 1,000 replications. We cluster at the country level. The Wald test for equality of AUCs also uses the joint bootstrap estimated variance-covariance matrix. As such we account for the very high correlation between different gap measures, often in the range of 0.9.¹²

For practical policy proposes, in addition to statistical power to predict crises, the right timing and stability of signals are important (Drehmann and Juselius 2014). EWIs need to signal a crisis early enough so that policy actions can be implemented in time to be effective. Yet, EWIs should not signal crises too early, as there are costs to macroprudential policies, and early adoption could undermine the support for necessary policy measures (e.g., Caruana 2010). EWIs should also be stable, as policymakers tend to base their decisions on trends rather than reacting to changes immediately (e.g., Bernanke 2004). A gradual implementation of policy measures may also allow policymakers to influence market expectations more efficiently, and to deal with uncertainties in the transmission mechanism (Committee on the Global Financial System 2012).

To assess the appropriate timing of a gap measure G_i , we follow Drehmann and Juselius (2014) and compute $AUC(G_{ij})$ for all horizons j within a three-year window before a crisis, i.e., j runs from -12 to -1 quarters.¹³ When we compute $AUC(G_{ij})$, we ignore signals in all other quarters than j in the window. For example, at horizon -6 , the rate of correctly predicted crises is solely determined by signals issued six quarters before crises. False alarms, on

¹²The high correlations are unsurprising given that all gaps are based on credit either normalized by GDP or population.

¹³By looking at each horizon separately, we wish to draw attention to the temporal stability of the EWIs, which is important for policymaking, rather than to the average time pattern.

the other hand, are based on all signals issued outside the three-year window before crises occur. We also do not consider signals issued during a crisis, as binary EWIs become biased if the crises periods are included in the analysis (Bussiere and Fratzscher 2006).

4. Data

Our data cover 41 economies.¹⁴ We use quarterly data with samples from as early as 1970 (depending on data availability) to derive the trend. The sample ends in the third quarter of 2017.

In our baseline specification to test forecast performance we only include gaps for an economy once we have 15 years of quarterly data, leading to an earliest date of 1985:Q1 in the horse race.¹⁵ This is necessary to ensure adequate data for the calculation of trends with the HP filter or regression coefficients with the linear projections.¹⁶ This starting point also approximately coincides with when many countries liberalized their financial systems, which in turn affected the dynamics of financial cycles and their relation with financial crises (Borio 2014). For a small number of economies, we further delay their inclusion in the panel until an end of a crisis: there is little practical point in beginning to test for crises when an economy is already in one.

Our measure of credit is as published in the BIS database of total credit to the private nonfinancial sector (see Dembiermont, Drehmann, and Muksakunratana 2013), capturing total borrowing from all domestic and foreign sources. Our nominal GDP series used

¹⁴The sample includes Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hong Kong SAR, Hungary, India, Indonesia, Ireland, Italy, Japan, Korea, Malaysia, Mexico, the Netherlands, Norway, the Philippines, Poland, Portugal, Russia, Saudi Arabia, Singapore, Spain, South Africa, Sweden, Switzerland, Thailand, Turkey, the United Kingdom, and the United States.

¹⁵For 20 economies in our sample, data are available from 1970:Q1, so they are included from the start in 1985. By 2000:Q1, 28 economies are included in the sample.

¹⁶As a robustness check we also used a run-in period of only 10 years with qualitatively similar results. However, we prefer the 15-year specification, as 10 years of data for country-specific projections are rather limited. On this basis, one could even argue for a longer sample such as 20 years to estimate stable trends. By doing so, the crises of the late 80s would, however, drop out of the horse race, severely limiting the number of observed crises episodes.

to generate credit-to-GDP are drawn from national sources. To generate the capita gaps, we use CPI from national sources and population numbers from the International Monetary Fund and the World Bank.

In total we have 27 crises in our sample. For crisis dating, we rely on the new European Systemic Risk Board crisis data set (Lo Duca et al. 2017) for European countries and on Drehmann et al. (2010) for the rest.¹⁷ As discussed, we drop post-crisis periods as identified in Lo Duca et al. (2017) and Laeven and Valencia (2012) for European and non-European economies, respectively.

5. A Horse Race between Linear Projection Gaps

Our first exercise is to compare different linear projection gaps in order to get a sense of which performs best. We started with a broad set of options, with h ranging from 4 to 36 quarters, and one to eight lags included in the equation. In all cases, we considered two normalizations of credit, namely by GDP (i.e., the credit-to-GDP ratio, with both credit and GDP measured in nominal terms) and per capita (that is, nominal credit divided by the product of the level of the CPI and the population). These different normalizations are indicated by “GDP” and “capita,” respectively.

When using real credit per capita, we face a scaling issue. The reason is that real credit per capita is measured in units of local currency, normalized by the CPI and population. National currencies have, however, very different units, as indicated by simple dollar exchange rates ranging from below one to multiples of thousands. While the growth gap method is invariant to scaling, this is not the case for the HP gap or the projection gap. To overcome the scaling problem for the per capita normalizations, we take natural logs of normalized credit. The gap measure may then be interpreted as the percentage difference between the level and the underlying trend.

We also perform the estimation both economy-by-economy and as a panel with economy fixed effects but with other coefficients constrained to be identical across all economies. We do this recursively,

¹⁷We exclude crises related to transitioning economies or that were imported from abroad based on Lo Duca et al. (2017). In addition, we classify the crisis in 2008 in Switzerland as imported.

adding one quarterly observation at a time to an expanding sample. With each recursion we take the final residual as a measure of the credit gap in that period. This approach is consistent with the idea that we require a measure that is useful in real time; in the same way, our HP-filter results will be based on a one-sided filter.

As well as adding observations with each recursion, we also add economies as data become available. For comparability between panel and economy-specific estimation, we only include an economy in the panel once we have 15 years of data for reasons discussed above.

Consistent with the intuition of Hamilton (2018), linear projections based on low values of h do not perform well.¹⁸ In addition, performance generally drops off with additional lags. We hence report a range of results for $h \in \{20, 24, 28, 32, 36\}$, each with one, two, and four lags. Combined with two different normalizations and both economy-by-economy and panel estimation, we are comparing the AUCs of 60 different formulations of linear projections for each of 12 different horizons.

The key takeaways are summarized in figure 1.¹⁹ For each panel, the solid line in figure 1 represents the AUC at different horizons, up to 12 quarters. Symmetric dotted lines indicate 95 percent confidence bands around the point estimates. In addition, the dot-dash black lines indicate the results for our, ultimately, preferred specification for the projections gaps, based on the panel estimates, normalized by GDP and with lags 28 and 29.

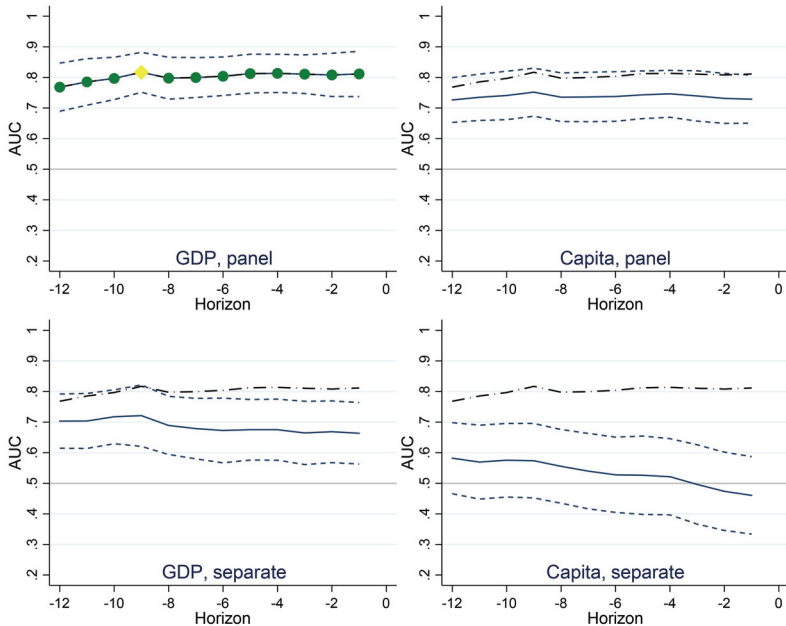
To highlight difference across the specifications, we add yellow diamonds and green dots (see online version of paper for figures in color). They are defined as follows:

- Yellow diamonds: Highest AUC across all of the 60 specifications at that given forecast horizon;
- Green dots: AUC is not statistically different from the highest AUC at a 95 percent confidence level, based on bootstrapped critical values using 1,000 replications.

¹⁸Full results are available on request.

¹⁹The figures and tables of underlying data for all 60 different formulations are shown in the online appendix (figure OA1 and table OA2), available at <http://www.ijcb.org>. Also available are all figures in color.

Figure 1. AUCs for Different Measures of the Linear Projection Gap Based on Lags 28–29



Notes: AUCs for different forecast horizons based on lags 28–29. A dot-dash line indicates the results for panel estimation with GDP normalization on lags 28–29, for ease of comparison. For the full set of graphs for $h \in \{20, 24, 28, 32, 36\}$, each with one, two, and four lags, please see the online appendix figure OA1. Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Yellow diamond: highest AUC across the 60 specifications at that given forecast horizon. Green dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on boot-strapped critical values using 1,000 replications.

Comparing the results, two results are evident from figure 1:

- (i) Normalizing by GDP statistically dominates normalizing by population.
- (ii) The panel estimation dominates estimation by each economy separately.

We therefore focus on cases where credit is normalized by GDP, and the linear model is estimated as a panel.

Reading across all different specifications in the online appendix (figure OA1), it is also clear that:

- (iii) The lag length and choice of h make little difference for the predictive performance of the different projection gaps, at least when h is between five and nine years.

This is not surprising, as the different gaps share very similar cyclical properties, with an average cycle length of 16 years.²⁰ Given that we judge performance by the AUC, we ultimately chose the specification with the highest AUC on average. As such we focus on the model with $h = 28$ and two lags as our preferred linear projection model for the remainder of the paper.²¹ However, as a robustness check, we will also assess the original specification for the projection gap suggested by Hamilton (2018) later.

To further uncover the sensitivity of the linear projection gap's performance to modeling assumptions, figure 2 reports the AUCs for the projection gaps based on real-time information versus over the full sample; estimated economy-by-economy (labeled "separate") versus as a panel; and normalized by GDP versus population. All panels in figure 2 are based on our preferred projection specification.

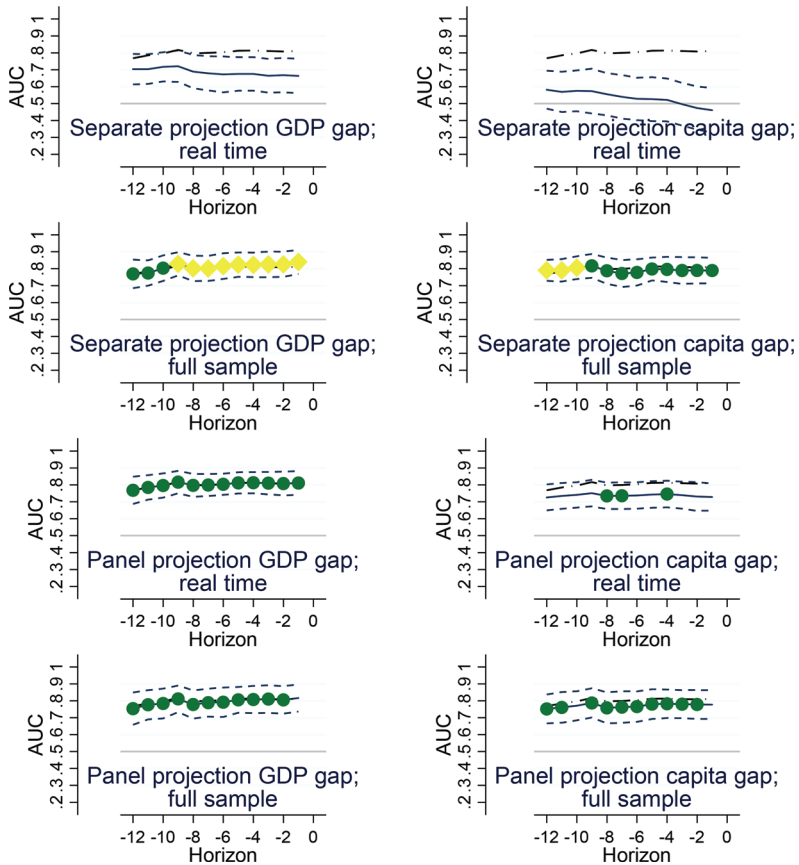
In line with the results above, it is clear that normalizing by GDP generally generates higher AUCs than normalizing by the population, especially when estimating in real time. The improvement based on full sample averages 0.06, whereas for real time it is a larger 0.11.

The figure also highlights that estimating using a panel instead of on each country individually makes little difference when applied to the full sample (fourth row of the figure versus the second row;

²⁰Cyclical properties, based on a turning-point analysis, are presented in table OA1 in the online appendix for all different gaps discussed in the paper. The average cycle length of 16 years is similar to that of the baseline HP-filtered credit-to-GDP gap. It also in line with the financial cycle literature (e.g., Claessens, Kose, and Terrones 2012; Drehmann, Borio, and Tsatsaronis 2012; and Aikman, Haldane, and Nelson 2015).

²¹The average AUC for including only one lag of 28 quarters is marginally higher at the fourth decimal (the average AUC for using lags 28 and 29 or only lag 28 are 0.803 to three decimal places). We prefer two lags, mindful of the original justification of Hamilton (2018) for proposing four lags for the linear projection: d lags should in principle work with any process up to order $I(d)$.

Figure 2. AUCs for Different Measures of the Linear Projection Gap for $h = 28$ with Two Lags



Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for real-time panel estimation with GDP normalization, for ease of comparison. Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Linear projections based on economy-by-economy real-time estimates (top row), economy-by-economy full-sample estimates (second row), panel real-time estimates (third row), and panel full-sample estimates (bottom row). Left column is credit normalized by GDP, and right column is based on real credit per capita. Yellow diamond: highest AUC across the eight specifications at that given forecast horizon. Green dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications.

the average AUC declines by 0.02) but dramatically improves performance when applied to expanding samples (first versus third rows; the average improvement in AUC is 0.16).

This points to a potential “endpoint” problem of the linear projection gap, especially for small samples. The reason is that during a credit boom—for example, in the early 2000s in the United States—the estimated coefficients increase in real time so that the residuals that the projection gap is based on don’t increase sharply, and are hence less likely to signal the impending crisis (see figure OA2 in the online appendix).

Finally, using the full sample rather than expanding sample regressions generally improves AUCs, although this is not a practical option for policymakers seeking to construct an EWI. Comparing analogous panels between rows 1 and 2, and also rows 3 and 4, full-sample estimation dramatically improves the AUC when the normalization is by population (by an average of 0.15) or the estimation is economy-by-economy (by 0.19) or both (by 0.26). By contrast, the difference is trivially negative when panel estimation is applied to credit normalized by GDP (-0.01).

These results suggest caution in interpreting some implementations of the linear projection. For example, Richter, Schularick, and Wachtel (2017) use the linear projection-based gap in country-by-country estimation on the full sample based on credit normalized by population with $h = 20$ and four lags. In the context of our figure 2, their results are closest to the second panel on the right column. However, if the objective is to assess the usefulness of measures of the credit gap to policymakers, the real-time results are the relevant ones to focus on. These are given in the top-right panel. The point AUCs here are less than 0.5 at some horizons and never statistically significantly different from an uninformative indicator, indicating that this implementation of the linear projection has no statistical power for predicting crises in real time in our panel.

6. Widening the Field

Given our preferred linear projection model, we now compare it against alternatives. We consider six gaps, as summarized in table 1. As in the previous section, we focus on two different normalizations

Table 1. Different Credit Gap Measure Labels

		Gap Measure		
		Difference from One-Sided HP Trend	Five-Year Growth	Residual from Linear Projection
Normalization	GDP	HP GDP Gap (Baseline)	Growth GDP Gap	Projection GDP Gap
	Real Credit per Capita ^a	HP Capita Gap	Growth Capita Gap	Projection Capita Gap

Note: ^aTo overcome the scaling problem of real credit per capita, $\ln(\text{real credit per capita})$ is used.

of credit, namely by GDP and per capita. For each ratio, we apply three possible gap measures:

- (i) the difference from a one-sided HP-filtered credit with a smoothing parameter of 400,000 (the HP gap);
- (ii) 20-quarter (five year) growth rates (the growth gap); and
- (iii) the residual from real-time linear projections with $h = 28$ and two lags (the projection gap).

As before, we include the gaps for a country for each of the measures once we have 15 years of underlying credit data for the country.²²

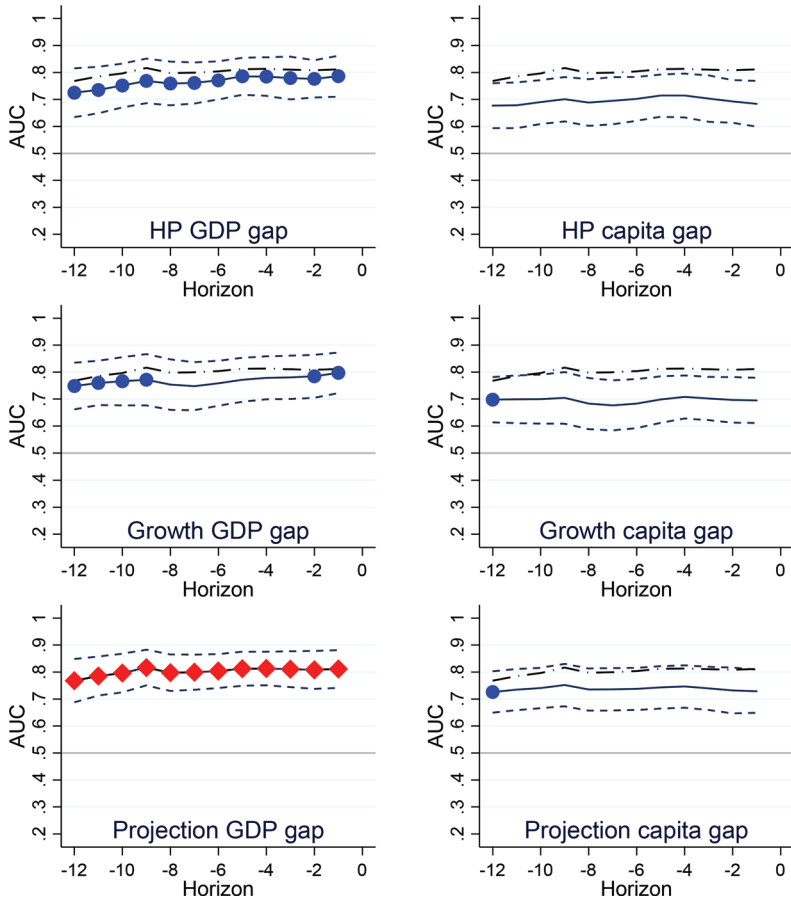
Figure 3 presents the main results (table OA3 in the online appendix shows the underlying statistics). Panels in the left-hand column are based on credit-to-GDP ratios, and the right-hand column on real credit per capita. The top row shows the HP gaps, the middle row the growth gaps, and the bottom row the projection gaps. For each panel, the solid line represents the AUC at different horizons, up to 12 quarters. Symmetric dotted lines indicate 95 percent confidence bands around the point estimates.

The figure summarizes the key takeaways from the horse race: First, normalizing by GDP results in superior forecast performance over normalizing by the population. This holds across all methods to derive the gaps and for most forecast horizons.

Second, the panel projection GDP gap has the highest AUCs of all the different gap measures for all horizons, although the differences vis-à-vis the AUCs of the HP GDP gap are very small and never statistically significant. This is in stark contrast to the comparative performance of the same model when applied economy-by-economy, as we have shown in an earlier version of this paper (Drehmann and Yetman 2018). Then, the HP gap consistently outperforms the projection gap. The forecast performance of the growth GDP gap is also not much worse, albeit with some significant differences to the projection gap.

²²Graphs of the underlying credit gap data, by economy, are available in the online appendix (figure OA3).

Figure 3. AUCs for Different Measures of the Credit Gap



Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP graph, for ease of comparison. Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications. See the online appendix for the underlying statistics (table OA3).

Third, while the differences are sometimes statistically significant, they are generally not large from a policy perspective. The average AUC differences across horizons between the best performer

and the other gaps are less than 0.04 for both the HP GDP gap and the growth GDP gap and 0.06 for the projection capita gap.

7. Robustness Checks

As robustness checks, we consider splitting the sample in three different ways: by time, between advanced and emerging market economies, and between countries that experienced a (domestically driven) crisis during the GFC and those that did not. We also compare the result with those obtained using the original specification for the projection gap suggested by Hamilton (2018) for credit gap calculations, based on lags 20–23 (instead of 28–29). To preserve space, we only show the gaps normalized by GDP; full versions of the graphs and the underlying data are reported in the online appendix.

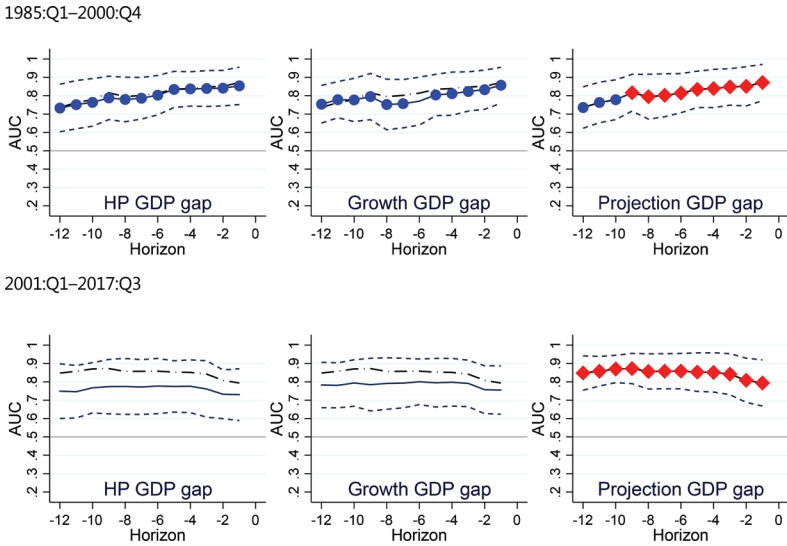
For the first exercise, we split the sample at the end of 2000. The results are reported in figure 4. They illustrate the key role that later periods play in the strong performance of the projection GDP gap. This measure no longer has the highest AUC at the longest horizons for the early sample split, although it is never statistically significantly different from the best performer. However, in the later subsample the projection GDP gap is the best-performing EWI at all horizons, and the difference is always statistically significant.

We next compare advanced and emerging market economies in figure 5. For the advanced economies there is little to choose between any of the measures statistically at most horizons. By contrast, for emerging market economies (EMEs) AUC performance is lower and more dispersed and confidence bands are much wider, suggesting that crisis prediction is inherently more difficult in EMEs.

Results also seem not to be driven by the global financial crisis (GFC): they are very similar for economies that had a domestically driven crisis during the GFC and those that did not (figure 6).

Finally, we compare the results with those based on projection gap parameters originally suggested by Hamilton (2018), using lags 20–23, to see how sensitive our results could be to the risk of overfitting of the projection equation. The results are displayed in figure 7. While there are differences between this specification and the one using lags 28–29, these are quantitatively small and the projection GDP gaps continues to perform best, by a small margin, at all horizons.

Figure 4. AUCs for Different Measures of the Credit Gap: Different Time Periods



Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP gap for the respective time periods. Panels based on normalizing by population are excluded to save space, but are available in the online appendix (figures OA4.1 and OA4.2). Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon for the respective time period. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications (for the respective time period). See the online appendix for the underlying statistics (tables OA4.1 and OA4.2).

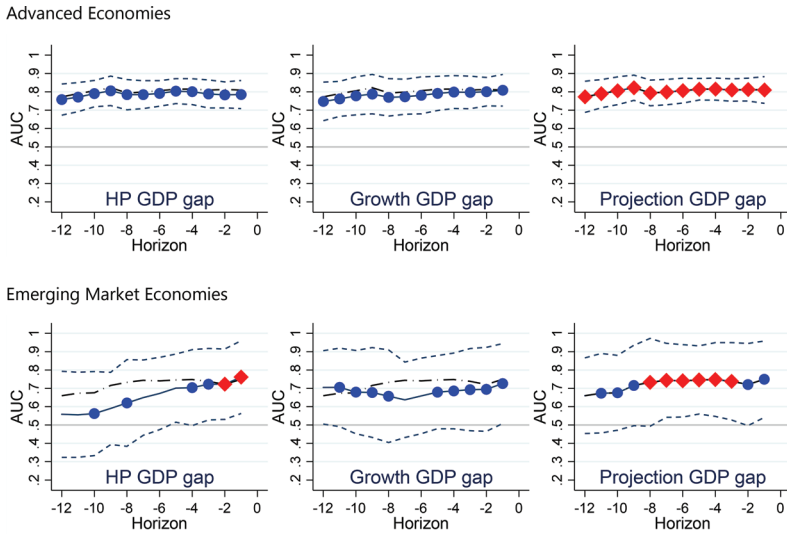
These results support our main takeaways above, that the projection GDP gap is the best-performing EWI overall in our sample, but differences are sometimes small and sample dependent.

8. Practical Implications

The analysis so far has several important practical implications for deriving indicators that signal “excessive” credit growth.

The core takeaway from our first set of results is that, when deriving projection gaps, it is crucial to use a panel approach rather

Figure 5. AUCs for Different Measures of the Credit Gap: Different Country Groups

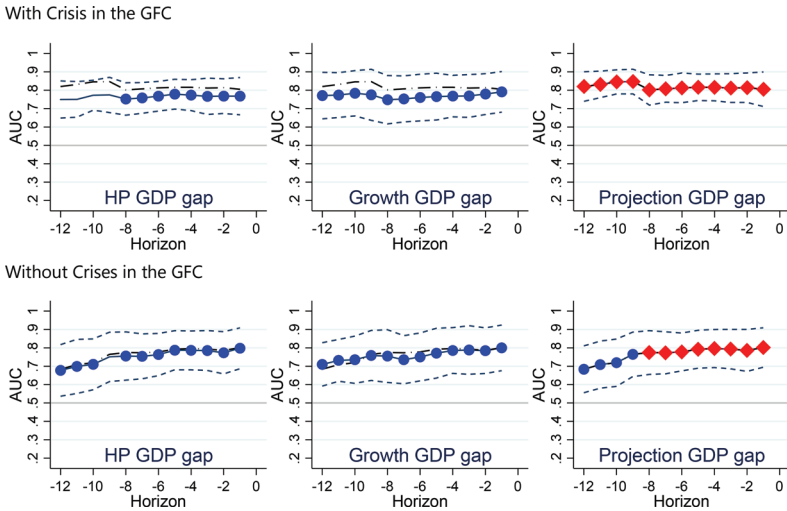


Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP gap for the respective country group. Panels based on normalizing by population are excluded to save space, but are available in the online appendix (figures OA5.1 and OA5.2). Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon for the respective country group. For horizons 9 to 12, the growth capita gap has the highest AUCs. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications (for the respective country group). See the online appendix for the underlying statistics (tables OA5.1 and OA5.2).

than running country-by-country regressions.²³ It is also important to assess predictive performance by using real-time estimates, as this is what policymakers can do in practice and results can differ significantly from a full-sample analysis. The question of lag length, on the other hand, is second order as long as h is between five and nine years.

²³While our analysis is clear that a panel approach is important, the optimal panel of countries may differ for specific practical purposes.

Figure 6. AUCs for Different Measures of the Credit Gap: Countries with or without Crises in the GFC

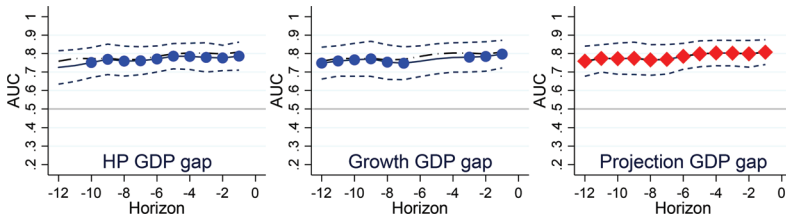


Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP gap for the respective country group. Panels based on normalizing by population are excluded to save space, but are available in the online appendix (figures OA6.1 and OA6.2). Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon for the respective country group. For horizons 9 to 12, the growth capita gap has the highest AUCs for countries without crises in the GFC. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications (for the respective country group). See the online appendix for the underlying statistics (tables OA6.1 and OA6.2).

Across all the results, it also stands out that normalizing credit by GDP results in superior forecast performance than normalizing by the population.

Our analysis is, however, less clear cut on the best approach to derive gaps. The statistical results (figures 3–7) suggest that the projection GDP gap is marginally better than the HP GDP gap, which in turn somewhat outperforms the growth GDP gap. But despite the statistically significant differences in forecast performance between the different gaps, they are not meaningful from a practical perspective. The main uncertainty policymakers face is that indicators give

Figure 7. AUCs for Different Measures of the Credit Gap, Projection GDP Gap Based on Lags 20–23



Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP gap. Panels based on normalizing by population are excluded to save space, but are available in the online appendix (figure OA7). Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications. See the online appendix for the underlying statistics (table OA7).

wrong signals: they may miss crises or may issue wrong crises calls in calm times.

To illustrate this, we take the HP and the projection gaps and undertake a simplified analysis where we do not look at 12 different forecast horizons but instead differentiate between no forthcoming crisis (labeled “normal”) and pre-crisis periods. The pre-crisis periods are the 12 quarters in the run-up to crises. As before, we drop the observations during actual crises. In this analysis, the AUC of the projection GDP gap (0.80) is higher than the AUC of the HP GDP gap (0.77) but the difference is not statistically significant at the 5 percent level. We then pick, for each of the GDP gaps, one particular threshold which, if breached, is seen as a crisis signal. This threshold is the one with the lowest noise-to-signal ratio that signals at least a 66 percent probability of a crisis in the pre-crisis periods.²⁴

²⁴This assumes that policymakers are more worried about missing crises than false alarms, and follows some of our earlier work (e.g., Borio and Drehmann 2009; and Drehmann, Borio, and Tsatsaronis 2011). However, the exact specification is arbitrary and many different approaches are possible, and sophisticated policy analysis often uses a range of different rule for robustness (see, e.g., Alessi and Detken 2018).

The identified thresholds are 6.0 for the HP GDP gap and 14.9 for the projection GDP gap.

To highlight the real-time uncertainty, table 2 shows the fraction of correct and incorrect signals in the normal and pre-crisis (i.e., in the 12 quarters before crisis) periods for the full sample and also the average across the robustness checks run in the previous section. Numbers in italics show the fraction of correct/incorrect signals for the individual indicators, while the other numbers provide the percentage of observations where both signals are giving the same or different messages.

The results have important implications from a policy perspective. First, independent of the indicator, around 30 percent of signals are wrong. Second, there is disagreement between the indicators in around 10 percent of the cases. Third, both indicators perform exactly equal in pre-crisis periods. Fourth, the projection gap performs marginally better in normal times by issuing 2–3 percentage points fewer wrong calls.

If policymakers would mechanically follow this rule, this would imply that, over a 10-year period, they could expect that the indicators would give wrong signals for around 3 years with either of the two gaps. Over the same period, the 2–3 percentage points difference of fewer wrong calls in normal times for the projection GDP gap relative to the HP GDP gap amounts to a single quarter. As such, dealing with the inherent uncertainty in identifying credit booms is an order of magnitude more important in practice than the choice between the different credit-to-GDP gaps.²⁵ Note, however, that despite this inherent uncertainty, all these gaps perform better than simple coin tosses. Thus, using them to calibrate prudential policies improves welfare, the more so if we consider the high typical costs of systemic crisis—100 percent of GDP or more (e.g., BCBS 2010a; Fender and Lewrick 2016).

9. Conclusions

The credit gap, defined as the deviation of the credit-to-GDP ratio from a one-sided HP-filtered trend with a smoothing parameter of

²⁵This also holds true if we add the growth GDP gap into the comparison.

Table 2. The Fraction of Correct and Wrong Signals of the HP and Projection Gap in Percent

		Projection GDP Gap						Total
		Normal			Pre-crisis			
		Correct	Incorrect	Incorrect	Incorrect	Correct	Correct	
<i>Full Sample</i>								
HP GDP Gap	Normal	Correct	66	5			71	
		Incorrect	7	22			29	
	Pre-crisis	Incorrect			25	9	34	
		Correct			9	57	66	
Total			73	27	34	66		
<i>Robustness Checks (Average)</i>								
HP GDP Gap	Normal	Correct	64	5			69	
		Incorrect	8	23			31	
	Pre-crisis	Incorrect			27	7	33	
		Correct			7	60	67	
Total			71	29	33	67		

Notes: Fraction of correct and incorrect signals in normal and pre-crisis periods. A crisis signal is issued if the gap breaches the critical threshold. For the projection gap, the threshold is 14.9. For the HP GDP gap, the threshold is 6.0. If a crisis signal is issued in a pre-crisis period, it is counted as correct. If it is issued in normal times, it is counted as incorrect. Pre-crisis periods are the 12 quarters in the run-up to a crisis. For robustness, we also show the average fraction in each cell for the sample splits shown in figures 4–6 (pre-2001, post-2001, advanced economies, emerging markets, countries that had a crisis during the GFC, countries that did not have a crisis during the GFC).

400,000 (for quarterly data), has been suggested as a useful measure for predicting crises. Two criticisms leveled at this measure are that (i) the normalization may be problematic because of the positive correlation between credit and GDP, and (ii) the HP filter has undesirable properties.

In this paper, we examine alternative measures of the credit gap that have been advocated by others to address these concerns.

We find that credit gaps based on linear projections in real time perform poorly in real time when based on country-by-country estimation. But when we estimate as a panel, and impose the same coefficients on all economies, linear projections perform marginally better than the baseline credit-to-GDP gap, with larger improvements concentrated in the post-2000 period and for emerging market economies, although the differences across the measures are often statistically small. The improvement in performance between the linear projection using a panel instead of applied to individual economies points to the importance of considering international evidence when calculating credit gaps for individual economies. That said, the practical relevance of the improvement is limited. Over a 10-year horizon, policymakers could expect one less wrong call on average.

References

- Aikman, D., A. Haldane, and B. Nelson. 2015. "Curbing the Credit Cycle." *Economic Journal* 125 (585): 1072–1109.
- Aldasoro, I., C. Borio, and M. Drehmann. 2018. "Early Warning Indicators of Banking Crises: Expanding the Family." *BIS Quarterly Review* (March): 29–45.
- Alessi, L., and C. Detken. 2018. "Identifying Excessive Credit Growth and Leverage." *Journal of Financial Stability* 35 (April): 215–25.
- Basel Committee on Banking Supervision (BCBS). 2010a. "An Assessment of the Long-term Economic Impact of Stronger Capital and Liquidity Requirements." Bank for International Settlements (August).
- . 2010b. "Guidance for National Authorities Operating the Countercyclical Capital Buffer." Bank for International Settlements (December).

- Berge, T., and O. Jordà. 2011. "Evaluating the Classification of Economic Activity into Recessions and Expansions." *American Economic Journal: Macroeconomics* 3 (2): 246–77.
- Bernanke, B. 2004. "Gradualism." Remarks at an economics luncheon co-sponsored by the Federal Reserve Bank of San Francisco (Seattle Branch) and the University of Washington, Seattle, May 20.
- Borio, C. 2014. "The Financial Cycle and Macroeconomics: What Have We Learnt?" *Journal of Banking and Finance* 45 (August): 182–98.
- Borio, C., and M. Drehmann. 2009. "Assessing the Risk of Banking Crises — Revisited." *BIS Quarterly Review* (March): 29–46.
- Borio, C., and P. Lowe. 2002. "Asset Prices, Financial and Monetary Stability: Exploring the Nexus." BIS Working Paper No. 114.
- Buncic, D., and M. Melecky. 2014. "Equilibrium Credit: The Reference Point for Macroprudential Supervisors." *Journal of Banking and Finance* 41 (April): 135–54.
- Bussiere, M., and M. Fratzscher. 2006. "Towards a New Early Warning System of Financial Crises." *Journal of International Money and Finance* 25 (6): 953–73.
- Caruana, J. 2010. "The Challenge of Taking Macroprudential Decisions: Who Will Press Which Button(s)?" Speech at the 13th Annual International Banking Conference, Federal Reserve Bank of Chicago, in cooperation with the International Monetary Fund, Chicago, September 24.
- Claessens, S., M. Kose, and M. Terrones. 2012. "How Do Business and Financial Cycles Interact?" *Journal of International Economics* 87 (1): 178–90.
- Committee on the Global Financial System (CGFS). 2012. "Operationalising the Selection and Application of Macroprudential Instruments." CGFS Papers No. 48.
- Dembiermont, C., M. Drehmann, and S. Muksakunratana. 2013. "How Much Does the Private Sector Really Borrow? A New Database for Total Credit to the Private Non-financial Sector." *BIS Quarterly Review* (March): 65–81.
- Drehmann, M., C. Borio, L. Gambacorta, G. Jiménez, and C. Trucharte. 2010. "Countercyclical Capital Buffers: Exploring Options." BIS Working Paper No. 317.

- Drehmann, M., C. Borio, and K. Tsatsaronis. 2011. "Anchoring Countercyclical Capital Buffers: The Role of Credit Aggregates." *International Journal of Central Banking* 7 (4, December): 189–240.
- . 2012. "Characterising the Financial Cycle: Don't Lose Sight of the Medium Term!" BIS Working Paper No. 380.
- Drehmann, M., and M. Juselius. 2014. "Evaluating Early Warning Indicators of Banking Crises: Satisfying Policy Requirements." *International Journal of Forecasting* 30 (3): 759–80.
- Drehmann, M., and K. Tsatsaronis. 2014. "The Credit-to-GDP Gap and Countercyclical Capital Buffers: Questions and Answers." *BIS Quarterly Review* (March): 55–73.
- Drehmann, M., and J. Yetman. 2018. "Why You Should Use the Hodrick-Prescott Filter — At Least to Generate Credit Gaps." BIS Working Paper No. 744.
- Edge, R., and R. Meisenzahl. 2011. "The Unreliability of Credit-to-GDP Ratio Gaps in Real-Time: Implications for Countercyclical Capital Buffers." *International Journal of Central Banking* 7 (4, December): 261–98.
- Fender, I., and U. Lewrick. 2016. "Adding It All Up: The Macroeconomic Impact of Basel III and Outstanding Reform Issues." BIS Working Paper No. 591.
- Galati, G., I. Hindrayanto, S. Koopman, and M. Vlekke. 2016. "Measuring Financial Cycles in a Model-based Analysis: Empirical Evidence for the United States and the Euro Area." *Economics Letters* 145 (August): 83–87.
- Hamilton, J. 2018. "Why You Should Never Use the Hodrick-Prescott Filter." *Review of Economics and Statistics* 100 (5): 831–43.
- Hodrick, R., and E. Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking* 29 (1): 1–16.
- Jordà, O. 2011. "Discussion of 'Anchoring Countercyclical Capital Buffers: The Role of Credit Aggregates'." *International Journal of Central Banking* 7 (4, December): 241–59.
- Jordà, O., B. Richter, M. Schularick, and A. Taylor. 2017. "Bank Capital Redux: Solvency, Liquidity, and Crisis." NBER Working Paper No. 23287.

- Jordà, O., M. Schularick, and A. Taylor. 2011. "Financial Crises, Credit Booms, and External Imbalances: 140 Years of Lessons." *IMF Economic Review* 59 (2): 340–78.
- . 2017. "Macrofinancial History and the New Business Cycle Facts." *NBER Macroeconomics Annual 2016*, Vol. 31, ed. M. Eichenbaum, E. Hurst, and J. A. Parker, 213–63 (chapter 3). Chicago: University of Chicago Press.
- Juselius, M., and M. Drehmann. 2020. "Leverage Dynamics and the Burden of Debt." *Oxford Bulletin of Economics and Statistics* 82 (2): 347–64.
- Kindleberger, C. 2000. *Maniacs, Panics and Crashes*. Cambridge: Cambridge University Press.
- Laeven, L., and F. Valencia. 2012. "Systemic Banking Crises Database: An Update." IMF Working Paper No. 12/163.
- Lang, J., C. Izzo, S. Fahr, and J. Ruzicka. 2019. "Anticipating the Bust: A New Cyclical Systemic Risk Indicator to Assess the Likelihood and Severity of Financial Crises." ECB Occasional Paper No. 219.
- Lo Duca, M., A. Koban, M. Basten, E. Bengtsson, B. Klaus, P. Kusmierczyk, J. Lang, C. Detken, and T. Peltonen. 2017. "A New Database for Financial Crises in European Countries: ECB/ESRB EU Crises Database." European Systemic Risk Board/ECB Occasional Paper No. 194.
- Minsky, H. 1982. "Can It Happen Again? Essays on Instability and Finance." Armonk, NY: M. E. Sharpe.
- Ravn, M., and H. Uhlig. 2002. "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations." *Review of Economics and Statistics* 84 (2): 371–76.
- Repullo, R., and J. Saurina. 2011. "The Countercyclical Capital Buffer of Basel III: A Critical Assessment." CEPR Discussion Paper No. 8304.
- Richter, B., M. Schularick, and P. Wachtel. 2017. "When to Lean against the Wind." CEPR Discussion Paper No. 12188.
- Rünstler, G., and M. Vlekke. 2018. "Business, Housing and Credit Cycles." *Journal of Applied Econometrics* 33 (2): 212–26.
- Schüler, Y., P. Hiebert, and T. Peltonen. 2015. "Characterising the Financial Cycle: A Multivariate and Time-varying Approach." ECB Working Paper No. 1846.
- Swets, J., and R. Picket. 1982. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. New York: Academic Press.

The Impact of Regime Change on the Influence of the Central Bank's Inflation Forecasts: Evidence from Japan's Shift to Inflation Targeting*

Masazumi Hattori,^a Steven Kong,^b Frank Packer,^c and
Toshitaka Sekine^a

^aHitotsubashi University

^bHong Kong Monetary Authority

^cBank for International Settlements

Many central banks release inflation forecasts to reduce uncertainty; at the same time, an increasing number rely on a publicly stated medium-term inflation target to help anchor expectations. We examine how the adoption of an inflation target (IT) by a major central bank, the Bank of Japan (BOJ), influenced the impact of its inflation forecasts on private-sector expectations. We find that the relative accuracy of central bank forecasts versus those of the private sector declined, a deterioration not evident in GDP forecasts. This appears to have been due to a structural (upward) shift in central bank inflation forecasts with the introduction of the IT regime. Regression results suggest that private-sector forecasts discounted the shift in central bank forecasts. The results are consistent with a regime, after the adoption of inflation targeting, in which the

*We would like to thank Ikuko Samikawa, Yohei Yamamoto, and the staff members of the Bank of Japan for their helpful comments, as well as attendees of presentations of the paper at the European Central Bank (ECB), the Bank for International Settlements (BIS), the Bank of Japan, the Bank of Korea, the Federal Reserve Bank of San Francisco, the Reserve Bank of Australia, the Reserve Bank of New Zealand, and Kobe University. The authors would especially like to thank Jens Christiansen, Ippei Fujiwara, Narayana Kocherlakota, Kevin Lansing, Richhild Moessner, Benoit Mojon, and John Williams for their comments and/or advice. The views expressed in this paper are those of the authors and not necessarily the views of the Bank of Japan, the Bank for International Settlements, or the Hong Kong Monetary Authority. We also wish to thank Pierpaolo Benigno and anonymous referees for constructive suggestions. Corresponding author (Packer): frank.packer@bis.org.

private sector viewed the central bank forecasts as upwardly biased. More generally, they confirm the difficulty in raising inflation expectations from below in the presence of an effective lower bound in the nominal policy interest rate.

JEL Codes: E31, E52, E58.

1. Introduction

How central banks should best communicate to the market is an increasingly important topic in the central banking literature. With ever greater frequency, central banks communicate through forecasts of prices and output over both the near and medium term. These forecasts can serve the purpose of reducing errors and uncertainty by private forecasters, with regard to economic fundamentals as well as the future policy actions of the central bank. In so doing, they can improve the effectiveness of other central bank communications and policies as well as economic welfare more generally. This paper contributes to the literature on central bank forecasts, by documenting how the release of the forecasts of one major central bank—the Bank of Japan (BOJ)—has been influencing private-sector expectations of inflation, and asking why the nature of this influence may have shifted over time.

At the same time, central banks of the 21st century generally rely on a publicly stated medium-term inflation target to help anchor expectations of inflation. Inflation targeting (IT) removes uncertainty about at least one of the ultimate objectives of the central bank, however much macroeconomic and global shocks may influence near-term inflation outcomes. The Bank of Japan adopted inflation targeting in early 2013, relatively late in the community of central banks in advanced economies, and more than a decade after they began to release economic forecasts. This paper aims to examine whether the impact of Bank of Japan forecasts on those of the private sector has been influenced by the adoption of an inflation target, which makes this paper unique in the empirical literature.

In contrast to most other advanced economies' experiences with inflation targeting, where IT was introduced in an effort to bring overly high inflation down and stabilize it at low levels, the Bank of Japan moved to IT when existing inflation (and indeed the inflation

of the previous 15 years) was below the new target. In cross-country work, Ehrmann (2015) suggests that central banks may have more difficulty in hitting newly adopted inflation targets from below than from above, as inflation expectations in such cases can be sticky in response to positive inflation surprises. The data set of Ehrmann's paper ends too quickly to lend insight into Japan's experience, however.

The value-added of our paper is as follows. While there is a large literature on the effectiveness of inflation forecasts, as well a separate one on the effectiveness of IT frameworks for monetary policy, our paper is the first, to our knowledge, that empirically examines how inflation forecasts by the central bank might be affected by the introduction of an inflation-targeting regime. The main theoretical reference to date is Dale, Orphanides, and Osterholm (2011), which models the joint presence of private-sector and central bank inflation forecasts, as well as of central bank inflation targets. In the model, if central bank forecasts are imprecise enough, the introduction of inflation targets can crowd out a role for central bank forecasts in communicating imperfect information. Changes to other parameters of the model can do so as well, such as a structural change that makes it difficult for the private sector to assess the quality of the central bank's forecasts.

Another argument is that central bank (CB) forecasts may be discounted in an IT regime, because the CB has the incentive to adjust its forecasts towards the target to communicate its commitment to achieve the inflation target. In other words, with a target to meet, central bank inflation forecasts became more Odyssean in nature rather than Delphic (for discussions of the distinction, see Campbell et al. 2012 and Andrade et al. 2018).¹ Because private forecasters are ex ante aware of the dual nature of the central bank's forecast once there is an inflation target, they will discount the central bank forecasts relative to those undertaken before the target was adopted, if the bank's ability to achieve it is in doubt.

¹In fact, from April 2013, shortly after the adoption of inflation targeting, it was announced that BOJ inflation forecasts would be made assuming the effects of past policy decisions. Since that time, at least during the sample period of this paper, its two-year-ahead inflation forecasts (excluding consumption tax effects) had been close to around 2 percent. Prior to that time, forecasts had been only conditioned on the future path of interest rates (see footnote 15).

Japan introduced an inflation target when its inflation was below the target, which is not the typical situation in which inflation targeting has been introduced historically. But below-target inflation can no longer be viewed as unusual, with inflation levels in advanced as well as many emerging economies persistently weak and well below established targets. For countries that may be considering introducing an inflation-targeting regime in the midst of a secular wave of disinflationary pressure, the experience of Japan poses important lessons. The Japanese experience also allows us to investigate whether the influences of the IT regime that might in theory affect the accuracy of inflation forecasts have in fact been observed in practice.

Historically speaking, Japan introduced an inflation target due to a political shock, which had been largely unpredicted at the time. The introduction of inflation targeting was triggered by the election of the Liberal Democratic Party (LDP) and its leader Shinzo Abe to prime minister in December 2012. Aggressive monetary policy easing was one of his “three arrows” of economic policy, and once he became prime minister, Abe insisted on an inflation-targeting regime to achieve this end. While the nomination and ascension of Haruhiko Kuroda to be governor of the Bank of Japan in April 2013 is often associated with inflation targeting in Japan, it was because of the Abe administration’s pressure that Governor Shirakawa was forced to introduce an inflation-targeting regime in January 2013 well before his term ended.²

Abe’s victory in the election of the LDP leadership the previous September was not widely expected, and in fact the result was quite a close call. Were it not for a last-second endorsement, the head of the party and the eventual position of prime minister could easily have gone to an individual with much more conservative views on monetary policy.³ Thus, when considering the political events

²On November 12, 2012, Shirakawa stated in a public speech the view that it was economic growth supported by increased growth potential that was necessary to overcome deflation (Shirakawa 2012). Moreover, in his memoirs Shirakawa wrote, “I was against strongly adhering to a specific number like ‘2%’ for the target inflation rate (authors’ translation)” (Shirakawa 2018, p. 318).

³There were in fact five candidates up for the LDP’s presidential election in September 2012. A veteran politician, Shigeru Ishiba, won considerably more votes than Abe in the first round of voting—199 votes versus 141 (out of 489).

as they actually occurred, Japan would appear to provide a natural experiment on what would happen to central banks' and the private sector's inflation forecasts after an unanticipated political shock results in the introduction of an inflation-targeting regime.

To preview our results, in the estimations that follow, we find that after the introduction of inflation targeting, the relative accuracy of central bank forecasts versus those of the private sector declined. Such a relative deterioration of central bank forecast performance is not evident in the gross domestic product (GDP) forecasts. This appears to be due to a structural shift in central banks' forecasts starting with the introduction of the IT regime. Regression estimates of monthly changes in private-sector forecasts, which include the deviation of their forecasts from Bank of Japan forecasts as an explanatory variable, then show the best fit to be one that includes a level shift downward in the IT era, which discounts the change in BOJ forecasts. Once again, a similar pattern is not apparent in the case of regressions for GDP forecasts.

The adjustment of central bank forecasts does not appear due to their being crowded out by perfectly credible inflation targets, nor do the regression results suggest that increased uncertainty with regard to the precision of central bank forecasts are the main factor, as theory might suggest (Dale, Orphanides, and Osterholm 2011). Rather, the results are consistent with central bank forecasts having become more Odyssean (Campbell et al. 2012 and Andrade et al. 2018), and private-sector forecasters largely adjust for the resulting bias of the central bank forecast, anticipating the problems of monetary transmission in an era of chronically below-target inflation and the zero lower bound.

Abe was not welcomed by a number of big names, including the head of his own political faction. Abe only became a viable candidate when Taro Aso, a former prime minister, decided to support Abe at the last moment. Because the top candidate did not get the majority of votes, it went to a second round, which is the first time that had happened in more than 40 years. In the second round, Abe won the majority. This in turn was the first time that the candidate in the second place in the first-round voting had won in the final round in more than 70 years. The previous front-runner, Ishiba, had expressed a reserved view about inflation targeting and aggressive monetary easing, expressing more concerns about the risk of high inflation by mentioning the possibility of hyperinflation in past interviews to media in 2010 (LDP Policy Research Council Chairperson's Regular Press Conference, February 17, 2010) and 2012 (*Nikkei* newspaper, December 21, 2012).

The rest of the paper will proceed as follows. In the next section, we review the literature on central bank forecasts as a form of central bank communication, as well as communication in light of the introduction of inflation-targeting regimes. In section 3, we discuss the data and institutional background, as well as outline the empirical strategy behind the tests for the effectiveness of central bank forecasts. Section 4 reviews the performance of central bank and private-sector forecasts both prior to and subsequent to the introduction of inflation targeting, and tests for structural breaks in the forecast series. In section 5, we present the main results, based first on monthly, and then quarterly, data. Section 6 concludes.

2. Review of the Literature: The Impact of Central Bank Inflation Forecasts and Targets

The literature on the role of central bank communication in monetary policymaking exploded in the late 1990s and the early 2000s, and this early literature is summarized comprehensively in Blinder et al. (2008). To quote its assessment, central bank communication “has the ability to move financial markets, to improve the predictability of monetary policy, and the potential to help monetary authorities achieve macroeconomic objectives.” At the same time, there was not yet a consensus on best practice across central banks, since communication strategies clearly differed significantly.

An increasingly important strand of the literature focuses on how central bank communication affects private-sector forecasts of inflation. Since private-sector expectations of inflation determine ex ante real interest rates, by influencing these expectations central bank communication can in turn determine monetary conditions. Romer and Romer (2000) show that the Federal Reserve had, at least during their period of investigation, superior information to the private sector when it came to inflation forecasts, and the private sector indirectly inferred this information from the policy changes undertaken by the Federal Reserve. A number of other papers have since shown that the release of information by the central bank can increase the predictive precision of private interest rate forecasts.

An early look at the influence of the publication of the central bank’s own inflation forecasts in clarifying future economic developments was provided by Fujiwara (2005), who showed that central

bank forecasts have a significant effect on private-sector forecasts as well as diminishing uncertainty. The more recent strands of the literature document the impact of central bank forecasts on the actual level of private-sector inflation expectations. Hubert (2014) found that central bank forecasts in the case of the United States became a focal point for private-sector expectations, while Pedersen (2015) showed that the forecasts published by the central bank in the case of Chile influenced the short-run inflation forecasts of the private sector. Hubert's (2015) study of five advanced economies again found that central bank inflation forecasts indeed influence the level of private forecasts in all cases. More recently, de Mendonca and de Deus (2019) find that higher central bank forecasts in three emerging market economies result in upwardly revised private-sector forecasts, but more in the case of GDP growth than inflation forecasts.

Though also a subject of the central bank communication literature, the announcement of medium- to long-term inflation targets differs from those of inflation forecasts. The introduction of inflation targeting has been shown to reduce the dispersion of inflation forecasts generally (Crowe 2010), which is what theory would predict if targets are credible enough to provide an anchor to expectations. However, the finding does not apply when only developed countries alone are examined (Cecchetti and Hakkio 2009, Capistran and Ramos-Francia 2010).⁴ Likely reasons for this finding include the pre-existing relative stability of inflation in developed countries and already homogenous views about future developments.

Inflation-targeting regimes became widespread in an era when countries viewed them as a tool to rein in high inflation by anchoring expectations at the target. However, over the past decade weak inflation has meant that inflation has been persistently below levels considered optimal across a wide range of countries, not least the United States. Ehrmann (2015) suggests that at low levels of inflation, inflationary expectations are less likely to be anchored

⁴The results are not yet clear-cut in cross-sectional empirical work either. While Ehrmann, Eijffinger, and Fratzscher (2012) find that transparency—in which having an inflation objective is one component—can reduce the dispersion of inflation forecasts, by contrast, Siklos (2013), in a study covering nine economies, finds that transparency of the central bank is associated with an increase in disagreement of inflation forecasts, a finding which holds regardless of IT regime.

by a target, and are more sensitive to lower-than-expected inflation shocks than higher-than-expected inflation shocks. The author concludes there may be unique difficulties in managing inflationary expectations when the central bank is targeting inflation from below, perhaps due to the difficulties of operating monetary policy at the effective zero lower bound.⁵

How might the impact of central bank inflation forecasts on private-sector expectations change with the adoption of an inflation target? Morris and Shin (2002) make the point that public information has potentially a dual role: it both conveys the status of fundamentals and serves as a focal point for beliefs. In the latter role, there are conditions under which it can crowd out the incentive of the private sector to produce high-quality forecasts.⁶ Demertzis and Viegi (2008) apply the Morris-Shin model explicitly to the announcement of an inflation target and show that inflation targets may indeed serve as focal points for coordinating private expectations. But they note that anchoring is improved only if large shocks are not anticipated and all other public information is unclear.

As mentioned in the introduction, in the theoretical article by Dale, Orphanides, and Osterholm (2011), the private sector and the central bank both produce inflation forecasts, using their own forecasting models, and the central bank also has the ability to announce an inflation target. The private sector takes the central bank's forecast into account when forming its forecast: the private-sector forecast is the weighted average of forecasts solely based on its own model and one published by the central bank, and if the recent relative performance of the central bank forecast declines, the weight on

⁵Christensen and Spiegel (2019) also provide evidence that inflation targets are difficult to achieve from below.

⁶Morris, Shin, and Tong (2006) specified further the conditions under which the crowding out of the incentive to provide accurate forecasts might occur. Demertzis and Hoerberichts (2007) and Kool, Middeldorp, and Rosenkranz (2011) present related models in which increased transparency of central bank communication can also crowd out private information. An empirical study that relates an inflation target to the level impact of central bank inflation forecasts is Pedersen (2015). When private forecasters believe that inflation will be over the central bank's target in the medium and long term, the short-run inflation forecasts are then higher than otherwise. However, as an inflation target is in place throughout the sample period, the paper does not assess whether the existence of the target itself affects the influence of central bank forecasts.

the central bank forecast in forming the private-sector forecast will also decline accordingly. The information value of the central bank's forecast is effectively discounted.

As for interaction between inflation forecasts and targets in the paper's model, while inflation forecasts are of variable precision (as in Morris and Shin 2002) and thus have "the potential to mislead and distract," inflation targets, by contrast, are assumed to be credible and thus can make central bank forecasts redundant and less distracting to the private sector (see Dale, Orphanides, and Osterholm 2011, p. 24ff). Within the framework of the model, channels through which central bank inflation forecasts can lose explanatory power with the introduction of an inflation target include (i) the inflation target anchors expectations such that the noisy central bank forecast now adds less net information to the market; (ii) the introduction of the inflation target raises uncertainty about the central bank's model of the inflation and the precision of their forecasts.⁷

Though not covered by the model in Dale, Orphanides, and Osterholm (2011), there is a further explanation of why central bank inflation forecasts can lose explanatory power under inflation targeting: the forecasts may become more Odyssean in nature to communicate the central bank's intent to achieve the target (Campbell et al. 2012 and Andrade et al. 2018), while private-sector forecasters may be skeptical about the central banks' ability to achieve the adopted inflation target. This skepticism can become particularly ingrained when attempting to reach inflation targets from below, due to the effective zero lower bound of the nominal policy interest rate. In this case, even if the central bank's target has credibility of intent, the lack of credibility of action may further feed skepticism (See Bomfim and Rudebusch 2000 for further discussion of this distinction).

⁷The above summary is based both on the model setup in Dale, Orphanides, and Osterholm (2011) and footnote 11 in the same work. In footnote 11, the authors note that the gain parameter k^f which represents an ability to assess the quality of the central bank's forecasts, "could also be seen as partly reflecting the extent to which the central bank makes and communicates changes in its analytical framework." So while clarity of objectives of inflation targets may encourage more aggressive easing (Orphanides 2018), we interpret the model as implying that when accompanied by untested actions, the parameter reflecting the ability of the private sector to assess the quality of the central bank forecast could be affected by the change of monetary policy regime.

In sum, the literature, despite clarifying in many respects how central bank forecasts might affect private forecasts, still has open questions with regard to how that impact might be affected by the introduction of an inflation target. Further, the empirical forecasting literature suggests that the properties of central bank inflation forecasts under an inflation-targeting regime might differ from those without inflation targets, particularly when the central bank has difficulty targeting inflation from below. Our paper, by focusing on the case of Japan, in which the central bank has provided inflation forecasts since 2000 but only since 2013 introduced an inflation-targeting regime, is well placed to shed light on the issue.⁸

3. Data and Empirical Strategy

3.1 Data

3.1.1 Private-Sector Forecasts

The main objective of the empirical analysis is to assess the impact of the forecasts of the Bank of Japan on private-sector inflationary expectations. As the main proxy measure of private inflationary expectations, we take the inflation forecasts from the so-called ESP survey of professional forecasters surveyed by the Japan Center for Economic Research (JCER). The survey started in 2004, which thus determines the beginning of the sample period for our regression analysis (2004–16).⁹ Around 40 economists and market analysts from the private sector and independent research institutes are asked

⁸There is also a literature that investigates how individual forecasters' incentives in the private sector can pose tradeoffs with the objective of minimizing forecast errors. For example, some forecasts are biased towards outcomes that favor the forecaster's employer (Ito 1990), while others can be influenced by the incentives of less able forecasters to mimic more capable ones (Ehrbeck and Waldmann 1996), or the incentives to benefit from the publicity that results from sharp differences from the consensus (Laster, Bennett, and Geoum 1999; Ottaviani and Sorensen 2006).

⁹The ESP forecasts were originally collected by the Economic Planning Association, an organization affiliated with the Cabinet Office, which published a periodic journal titled *Economy, Society, Policy* (which is where the acronym "ESP" came from). In April 2012, the Japan Center for Economic Research took over the survey.

their forecasts for the change in annual average level of consumer price index (CPI) excluding fresh food (“core inflation”) over the current and next fiscal years (from April to March of the following calendar year) along with other major macroeconomic variables including GDP growth. Private forecasters are surveyed monthly, with the survey period spanning the last few days of a month and the first few days of the following month, and the mean of the forecasts is published about a week after the close of the survey. For the purposes of this study, medians have also been made available to us. We focus on the median of these forecasts as the principal summary statistic: the choice is based on the fact that the Bank of Japan forecasts are also summarized by the median of forecasts of policy board members. Medians are also less susceptible to the influence of outlier forecasts.

3.1.2 Bank of Japan Inflation Forecasts

As mentioned above, our objective is to analyze the effect of the inflation forecasts of Japan’s central bank, the BOJ, on inflationary expectations of the private sector. In October 2000, the BOJ began to publish summary statistics of the internal forecasts made by individual members of its policy board for inflation, or the change in annual average level of CPI excluding fresh food (“core inflation”) over the current fiscal year. In 2001, the bank also began to release next-fiscal-year forecasts. Initially the Bank of Japan only announced ranges of forecasts, but from 2003 also included the medians of these forecasts. For the purposes of this paper, we focus on the median of the inflation forecasts of the Policy Board.

The frequency with which the forecasts have been provided has changed over time. Next-fiscal-year forecasts were first published annually and then, starting in 2005, on a semiannual basis every April and October. From mid-2008, the forecasts were released in January and July as well, thus increasing the frequency to a quarterly basis. We have collected the historical figures from a number of BOJ publications, including the “Outlook for Economic Activity and Prices” and “Statement on Monetary Policy.”

The focus of this paper is on the impact of next-year forecasts—in particular, how changes in BOJ forecasts for the next fiscal

Table 1. Bank of Japan's Forecasts and ESP Forecasts

	Bank of Japan's Forecasts	Private Sector's Forecasts
Source	BOJ Publications (e.g., "Outlook for Economic Activity and Prices," "Statement on Monetary Policy")	Japan Center for Economic Research ("ESP Forecast")
Frequency	October 2000–April 2008: Semiannually. July 2008–Now: Quarterly	May 2004–Now: Monthly
Forecast Variable	Annual Core Inflation (i.e., Headline Inflation Excluding Fresh Food)	Annual Core Inflation
Forecast Horizon	Current and Next Fiscal Years; Two-Year-Ahead Forecasts from October 2008	Current and Next Fiscal Years; Two-Year-Ahead Forecasts are available from time to time
Data Level	Range and Median of Individual Forecasts	Mean and Median of Individual Forecasts; Individual Forecasts are also available
Sources: Bank of Japan; Japan Center for Economic Research.		

year influence the private sector's forecasts for the same periods.¹⁰ Current-year forecasts are also available, but their movements reflect changes in realized inflation outcomes as much as changes in the outlook. Further, central banks usually are concerned with medium- to long-term inflation expectations, for which the next-year forecasts are a much better proxy. The features of the BOJ and the forecasts from the JCER survey are summarized in table 1.

¹⁰Two-year-ahead inflation forecasts have been regularly provided by the JCER from July 2013 and by the BOJ since October 2008 but are not used in this study due to the limited sample size.

3.1.3 Control Variables

We include monthly control variables in regression analyses that, in addition to the Bank of Japan forecasts, should also regularly shape private-sector inflation expectations. Particularly when assessing the impact of BOJ forecasts, it is important to control for significant changes to macroeconomic and financial market conditions that might affect inflationary expectations.

The main control variables that we include in this study are as follows:

Inflation “Surprises” from the Monthly CPI Releases ($InfSurp_t$). An inflation surprise is defined as the currently realized year-on-year quarterly core inflation minus the latest mean inflation forecast for that quarter from the ESP survey. Realized quarterly core inflation is calculated as the year-on-year change in the average core CPI level for the months of that quarter. When the core CPI level is only available for the first month or first two months of a quarter, realized inflation is the year-on-year change in the average core CPI level for which realized data are available. A positive surprise may lead the private sector to upgrade its inflation outlook. Pedersen (2015) shows that surprises in monthly released data affect current-year inflation expectations of private forecasters but not their next-year inflation expectations.

Changes in the Expected Yen Exchange Rate ($\Delta e_{t,ny}^{esp}$, *ny* for Next Year). We measure the log change in the expected yen–dollar rate between two consecutive ESP surveys for the next fiscal year. Expected depreciation of the Japanese yen might exert some upward pressure on inflation in Japan via exchange rate pass-through, while appreciation could work in the opposite direction.

Changes in the Spot Oil Prices (Δoil_t^{spot}) and Average Futures Oil Prices for the Next Fiscal Year (Δoil_t^{ny}). We measure the log changes in the spot prices as well as in the average prices of future contracts with delivery in the next fiscal years for West Texas Intermediate (WTI) crude oil.¹¹ Both the inflation forecasts made by the BOJ and by the private sector incorporate expected movements in energy prices. Changes in spot oil prices,

¹¹See appendix table A.1 for the full description of variables, including details on how the average prices are calculated.

as well as changes in oil price expectations, as reflected in futures prices, could shape the private sector's inflationary expectations.

We also include the lag of the change in inflationary expectations to control for persistence in the movement of inflationary expectations. A delayed response by the forecasts of professional forecasters to macroeconomic shocks, consistent with information rigidities and rejecting the null hypothesis of full information, has been documented by Coibion and Gorodnichenko (2012).¹²

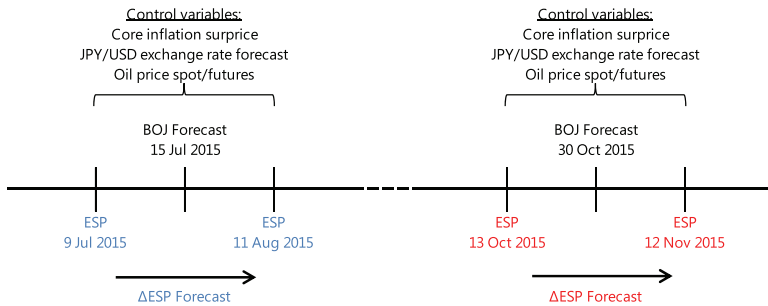
The Introduction of Inflation Targeting (IT). The full sample goes from 2004 (when the ESP survey began) to end-2016; the BOJ's adoption of inflation targeting covers only the final part of the full sample period. On January 22, 2013, the BOJ set an inflation target of 2 percent, and within a few months had introduced a regime of quantitative and qualitative easing measures (QQE) with the explicit objective of achieving that target in two years.¹³ By including simple and interactive dummies, our empirical model will take into account the adoption of inflation-targeting policy during the sample period, with a view towards shedding light on the effect it may have had on the relationship between central bank and private-sector forecasts.

The Lehman Brothers Default Shock. While we include many variables in the specification, we do not want to rule out the possibility that during certain extreme events, changed forecasts by the Bank of Japan and private-sector forecasts may show some spurious relationship due to factors outside the model. One plausible example of this is the Lehman Brothers default of September

¹²Townsend (1983) also discusses how learning mechanisms can convert serially uncorrelated shocks into serially correlated movements in economic decision variables.

¹³Since March 2006, the Bank had adopted a numerical reference (1 percent CPI inflation) as "understanding of price stability"; in February 2012, the Bank had switched that understanding to "inflation goal"; in January 2013, to "inflation target"; and the explicit time commitment of two years was only announced in April 2013. See appendix I of Nishizaki, Sekine, and Ueno (2014) and Hattori and Yetman (2017) for changes in exact wordings of these numerical reference points. Among them, the introduction of the 2 percent inflation "target" stood out as the most significant change in the monetary policy framework compared with the 1 percent inflation "understanding" or "goal." The (unreported) recursive breakpoint Chow test indicates that this is the timing when the structural break occurred in the BOJ inflation forecast.

Figure 1. Illustration of Matching Procedure and Methodology



Note: The dates under BOJ forecast and ESP indicate the date when the forecasts were published.

2008, after which business and consumer sentiment plunged dramatically. For this reason, we also report a regression model for a sample that excludes the two monthly observations immediately after the Lehmann shock.

Tax Delay Dummies. All monthly specifications include period dummies for December 2014 as well as June 2016, since very large ESP forecast changes in those months reflected announced delays of the consumption tax hike not yet reflected in the lower-frequency BOJ forecasts.

3.2 Empirical Strategy

The empirical approach is as follows. To ensure the data are aligned correctly, we match each publication of BOJ forecasts with two sets of ESP forecasts: one that comes from the survey date right before the release date of the BOJ forecast and one that comes from the survey date right after the release of BOJ forecast. The matching procedure for two successive dates is illustrated in figure 1. Combined with the intervening months for which there are no BOJ forecasts, the overall result is 150 monthly observations of ESP forecast changes, 42 of which are matched with 42 releases of BOJ forecasts between 2004 and 2016.

We take the monthly change in the median of ESP forecasts for the next fiscal year, $\Delta\pi_{t,ny}^{esp}$, as the dependent variable in our

main regression model. The key explanatory variable is the difference between the median of the BOJ forecasts and the ESP forecasts in the survey right before the release of the BOJ forecasts ($\pi_{t-1,ny}^{boj} - \pi_{t-1,ny}^{esp}$). During the intervening months when there are no BOJ forecasts, this variable is set to zero to reflect the view that in the months without a forecast the information content in the difference should be nil.¹⁴ Using this explanatory variable in a regression allows us to assess the degree to which private analysts adjust their expectations in response to the deviation of the biannual or quarterly BOJ forecasts from their own forecasts. If the degree of adjustment is significant, even after controlling for other factors, then this is consistent with the hypothesis that the private sector believes that the BOJ forecasts contain some valuable information about the economy beyond changes to the private sector's existing information set (as captured by the control variables in figure 1).

We examine the bilateral relation (without controlling for other factors) between the previous difference of the BOJ and the ESP forecasts (horizontal axis) and the change in the ESP forecasts (vertical axis) for the subset of months in which there is a BOJ forecast in figure 2. Indeed, a positive relation is apparent, which suggests that private forecasters may in fact have changed their forecasts in response to the newly released BOJ forecasts. Of course, this relationship needs to be examined more carefully in the monthly frequency multivariate regression model to follow, which controls for other determinants of inflation expectations.

4. Forecast Performance

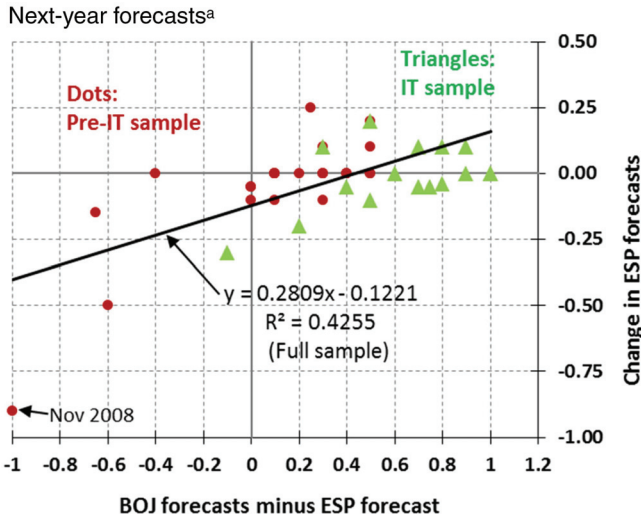
4.1 *The Relative Accuracy of BOJ Forecasts*

Before going to the regression analysis, we examine the performance of Bank of Japan and private-sector forecasts for CPI inflation and, for comparative reference, GDP growth.

As referred to above, extant research shows that Bank of Japan forecasts influence private-sector forecasts (e.g., Fujiwara 2005 and subsequently Hubert 2015). This influence could have been due to

¹⁴An alternative treatment of the variable, where the difference is set to the difference between the last available BOJ forecast and the latest ESP forecast, yields qualitatively very similar results.

Figure 2. Responsiveness of ESP Forecasts to the Difference between BOJ Forecasts and ESP Forecasts in the Previous Survey^a (in percentage points)



^aChanges in ESP forecasts refer to the changes in the median of forecasts of core inflation by private forecasters responding to the ESP surveys—one before the BOJ forecasts release and one after that. BOJ forecasts refer to the median of forecasts of core inflation by BOJ policy board members. BOJ forecasts minus ESP forecasts refer to the differences between BOJ forecasts and the ESP forecasts in the survey prior to the release of BOJ forecasts.

Sources: Bank of Japan; Japan Center for Economic Research; authors' calculations.

a prevailing view that the Bank of Japan forecasts were superior to private-sector forecasts, and in some sense based on a superior information set. Such a superior information set could of course include inside knowledge about the future direction of policy, though it is worth noting that officially Bank of Japan forecasts are made with reference to the view of market participants regarding the future course of policy. However, shortly after the adoption of the inflation-targeting regime, the Bank of Japan changed its forecast assumptions to include judgments of the Bank about the effects of past policy decisions.¹⁵

¹⁵From October 2000 through October 2005, Bank of Japan forecasts were based on the assumption that there will be no change in monetary policy; from

That said, in the pre-IT era, Bank of Japan forecast accuracy appears to be roughly similar to that of private-sector economists. Table 2, top panel, summarizes the mean errors and root-mean squared errors (RMSE) of the private-sector forecasts and the Bank of Japan forecasts for inflation during both the 2004–12 (pre-IT) and the 2013–16 (IT) periods. During the pre-IT period, the private-sector forecasts have lower mean error and RMSE than the BOJ's, but in both cases the differences are statistically insignificant.

Given the results in the literature that Bank of Japan forecasts influence those of the private sector, what the above findings confirm is that the impact of the Bank of Japan forecasts need not be due to a strictly superior information set or forecasting technology than that of the private sector. Rather, information that the Bank of Japan conveyed via its forecasts could be viewed as complementary to that of the private sector, and thus have an impact on the margin.

What about after the implementation of the inflation-targeting policy? The private-sector forecasts now have consistently lower mean error and RMSE than those of the Bank of Japan. Further, the differences in mean error and RMSE are statistically significant at the 5 percent level. The errors in the Bank of Japan's forecasts for inflation in the IT era—which now explicitly incorporated the Bank's assessment of the impact of past policy decisions—were invariably due to their being too high relative to realized inflation.

There is a striking asymmetry in forecast accuracy results when we examine forecasts of GDP instead of forecasts of inflation (table 2, bottom panel). Unlike the case of the CPI forecast, the BOJ's GDP forecast did not deteriorate after the introduction of inflation targeting; rather, it actually improves, as does that of the private sector. Further, the BOJ's GDP forecast performance is statistically indistinguishable from the ESP's GDP forecast, both in terms of the mean forecast error and RMSE. This is in stark contrast to the relative accuracy of the CPI measures.

April 2006 through January 2013, forecasts were in reference to the view of market participants regarding the future course of the policy rates, as incorporated in market interest rates. From April 2013 to the present, the forecasts were made assuming the effects of past policy decisions and with reference to views incorporated in financial markets regarding future policy.

Table 2. Accuracy of Next-Year Forecast^a (in percentage points)

	Of Which:						
	2004–16			Before Inflation Targeting		Inflation-Targeting Period ^c	
	BOJ	ESP		BOJ	ESP	BOJ	ESP
Core Inflation							
Mean Forecast Error ^b	0.6409	0.3099*		0.3401	0.1940	1.2296	0.4983**
Root Mean Squared Error ^b	1.0788	0.8394		0.8843	0.8515	1.3357	0.8193**
Real GDP Growth							
Mean Forecast Error ^b	0.8976	0.7214		1.0038	0.9077	0.7250	0.4188
Root Mean Squared Error ^b	1.8676	1.8009		2.1431	2.1498	1.3010	1.0016

^aComparison of the BOJ forecasts and matching ESP forecasts, taken in the survey right after the release of BOJ forecasts. An alternative comparison which matches BOJ forecasts instead with the ESP forecasts immediately before the BOJ forecast release does not significantly affect the results.

^bForecast errors are calculated by subtracting realized inflation (real GDP growth) rate from forecasts, i.e., a positive forecast error indicates the realized inflation (real GDP growth) rate is smaller than the forecast. Sample includes forecasts made in 2004–2016.

^cAs IT was announced on 22 January 2013, the IT period sample for matched BOJ and ESP forecasts starts from the January 2013 BOJ forecasts and the subsequent ESP forecasts.

** and * indicate the difference between BOJ forecast and ESP forecast is significant at 5 percent and 10 percent levels, respectively (t-test).

Sources: Bank of Japan; authors' calculations.

4.2 Evidence of a Structural Break

To investigate further the connection between the introduction of the inflation target and the poor performance of official forecasts, we test for a structural break in the Bank of Japan's inflation forecast series.¹⁶ The break is posited to be when the BOJ introduced the 2 percent inflation-targeting regime in January 2013. Based on the breakpoint Chow test, the null hypothesis of no break in the BOJ forecast series for CPI at that time is rejected at the 1 percent significance level (p-value, 0.0055). By contrast, BOJ forecasts for GDP show no evidence of a structural shift (p-value, 0.9736).

At the same time, private-sector forecasts for CPI also show evidence of a structural break, not shared by their forecasts for GDP (p-values, 0.0020 and 0.2363, respectively).

In line with these results, the coefficient on an inflation-targeting dummy, which takes on the value of one since January 2013, is positive and significant at the 1 percent level for both the BOJ and private-sector forecasts in the following simple regression:

$$\pi_{t,ny}^{boj} \text{ or } \pi_{t,ny}^{esp} = \text{Constant} + \text{IT Dummy} + u_t. \quad (1)$$

However, the shift of Bank of Japan inflation forecasts is larger than that of private forecasts: the obtained coefficients on the IT dummy are 1.27 for the BOJ and 0.75 for the private sector. This implies that the wedge between BOJ and ESP inflation forecasts increased by around 0.5 percentage point on average after the adoption of inflation targeting.

This pattern of significant structural change for the Bank of Japan inflation forecasts, not replicated in their GDP forecasts, is consistent with the view, alluded to earlier, that the adoption of IT in early 2013 was the result of an exogenous political event, which then appears to have caused a change in the inflation forecasts by the BOJ (but not similarly for the GDP forecast). As a result, the private sector also adjusted its inflation forecasts, but less so than

¹⁶For core CPI forecasts, those without consumption tax effects are used to avoid detecting spurious structural change. For ESP, April to September 2013 where those excluding consumption tax effects were not surveyed, the series is adjusted by another time dummy for the corresponding period. For real GDP, the outliers after the Great Financial Crisis (February and March 2009) and the China shock (April and May 2016) are adjusted by time dummies.

the Bank of Japan. This larger shift of the Bank of Japan inflation forecast resulted in its relatively poor forecast performance.

5. Regression Analysis

In this section, we examine how the private sector corrected for the incremental increase in the BOJ inflation forecast with the advent of inflation targeting.

5.1 Baseline Specification

As noted above, the principal regression equation takes as the dependent variable the monthly change in the median of ESP inflation forecasts for the next fiscal year $\Delta\pi_{t,ny}^{esp}$. For the explanatory variables, the key explanatory variable of interest is the difference between a fresh BOJ median forecast for the next year (available on a biannual or, later in the sample, quarterly basis) and the median ESP forecast, or $(\pi_{t-1,ny}^{boj} - \pi_{t-1,ny}^{esp})$. As explained above, for months when a fresh BOJ forecast is not available, this variable is set at zero to reflect the notion that there should be no additional information content. As previously mentioned, we also include a number of control variables for monthly changes in the economy and financial markets: inflation “surprises”; changes in the expected yen exchange rate; and changes in oil prices, both spot and future.

$$\begin{aligned} \Delta\pi_{t,ny}^{esp} = & Constant + \beta_1\Delta\pi_{t-1,ny}^{esp} + \beta_2InfSurpt + \beta_3\Delta e_{t,ny}^{esp} \\ & + \beta_4\Delta oil_t^{ny} \text{ (or } \Delta oil_t^{spot}) + \beta_5 \left(\pi_{t-1,ny}^{boj} - \pi_{t-1,ny}^{esp} \right) + u_t. \end{aligned} \quad (2)$$

The estimation results for the inflation forecasts are reported in table 3. We first report models for inflationary expectations without considering BOJ forecasts. The change in oil prices—whether via the spot (column 1) or futures (column 2) channel—has the right sign in that a positive change leads to an upward adjustment of the private sector’s forecasts of inflation. Since the coefficient on the oil futures prices variable is statistically significant while that on the spot oil price is not, for the rest of the paper we mainly rely on the oil futures price as a factor shaping inflationary expectations. The inflation surprise coefficient also has the right sign but is not quite statistically

Table 3. Baseline Specification: Next-Year Forecasts

	(1)	(2)	(3)	(4)	(5)
Constant	-0.0188* (0.0096)	-0.0196** (0.0098)	-0.0337*** (0.0113)	-0.0290*** (0.0098)	-0.0217** (0.0092)
Lagged Dep. Variable	0.2523** (0.1226)	0.2353* (0.1217)	0.2265** (0.1093)	0.1958** (0.0988)	0.2573*** (0.0939)
Inflation Surprise	0.1117 (0.0784)	0.1169 (.0763)	0.1227* (0.0740)	0.1058 (0.0706)	0.1031 (0.0677)
Change in USD Forecast	0.0241*** (0.0086)	0.0239*** (0.0082)	0.0199*** (0.0070)	0.0211*** (0.0065)	0.0166** (0.0064)
Change in Spot Oil Price	0.0032 (0.0021)				
Change in Oil Price Forecast		0.0044* (0.0026)	0.0041** (0.0020)	0.0032** (0.0015)	0.0017 (0.0014)
Diff (BOJ-ESP)			0.1545*** (0.0480)	0.3140*** (0.0879)	0.1923*** (0.0553)
Diff (BOJ-ESP)*Dummy IT				-0.1186 (0.1082)	0.0120 (0.0858)
Dummy IT				-0.1002** (0.0468)	-0.1100** (0.0531)
Obs.	150	150	150	150	148
R ²	0.5556	0.5630	0.6200	0.6628	0.6393
Adj. R ²	0.5369	0.5446	0.6013	0.6412	0.6158

Notes: Equation (5) excludes the first two observations after Lehman Brothers' bankruptcy. Tax dummies for December 2014 and June 2016 to reflect announced delays in consumption tax hike included in regression though coefficients not reported. Figures in parentheses indicate standard errors. ***, **, and * indicate significance levels at 1 percent, 5 percent, and 10 percent, respectively.

significant. On the other hand, changes in the expected yen exchange rate do significantly affect inflation expectations: the coefficient suggests that a 10 percent depreciation of the yen exchange rate would be associated with a 0.24 percentage point increase in expected inflation. The lagged dependent variable is statistically significant as well, consistent with a partially delayed response of professional forecasters to new information. The adjusted R-squared for the expectations models without Bank of Japan forecasts approximate to 54 percent in both cases.¹⁷

In column 3, we include the main explanatory variable ($\pi_{t-1,ny}^{boj} - \pi_{t-1,ny}^{esp}$) and find it is statistically significant at the 1 percent level. Even after controlling for other information that might have influenced expectations between the two ESP forecasts, the private-sector forecasters do indeed appear to take into account the degree to which recent Bank of Japan forecasts differ from their own previous forecasts when updating their own forecasts. The size of the coefficient on the variable suggests that on average for every 1 percentage point increase in the differential between BOJ and ESP forecasts in the month of the BOJ forecast, the private-sector forecasters would raise their own forecast by around 0.15 percentage point. The adjusted R-squared increases from 54 percent to 60 percent when consideration is made of the Bank of Japan forecasts, as shown in column 3.

As discussed above, it is likely that the specification is incomplete due to shifts in the monetary policy regime. We thus extend the

¹⁷In unreported specifications, we also included expected real GDP growth and the forecast long-term interest rate, but they were not consistently significant, nor did they change the main results. We also tried and found to be statistically insignificant the level of the forecasted variable (inflation), a measure of economic slack (the unemployment rate), a policy rate instrument (the call rate), actual inflation volatility, forecasted stock prices (TOPIX), and forecasted money supply (M2). Statistical tests reject significant (first-order) autocorrelation of the residuals in the major specifications. Decomposition of the differenced explanatory variable into separate private-sector and central bank forecasts resulted in small and statistically insignificant differences in the absolute value of the coefficients. We also ran robustness checks that confirmed that adjusting for the consumption tax hike (both ESP and the BOJ release forecasts net of the expected impact of the consumption tax hike of 2014), or including inflation volatility or a dummy for the inflation goal period did not change the main conclusions. The results are available upon request.

main regression equation by allowing for the impact of the central bank forecasts to change after the BOJ adopted inflation targeting.

$$\begin{aligned} \Delta\pi_{t,ny}^{esp} = & Constant + \beta_1\Delta\pi_{t-1,ny}^{esp} + \beta_2InfSurp_t + \beta_3\Delta e_{t,ny}^{esp} \\ & + \beta_4\Delta oil_t^{ny} + \beta_5\left(\pi_{t-1,ny}^{boj} - \pi_{t-1,ny}^{esp}\right) \\ & + \beta_6\left(\pi_{t-1,ny}^{boj} - \pi_{t-1,ny}^{esp}\right) * Dummy\ IT \\ & + \beta_7Dummy\ IT + u_t \end{aligned} \quad (3)$$

Column 4 reports the estimation results for the regression equation which adds both a period dummy which is one when inflation targeting was in effect (i.e., starting from the ESP survey in February 2013), and an interaction term which is the product of this dummy and the main explanatory variable. The two additional variables are intended to capture the fact that introduction of the IT regime could have affected the impact of the BOJ forecasts in two ways: it could have led the private sector to view the BOJ forecasts as consistently biased (shift in the constant), or it could have reduced the impact of the changes in BOJ forecast (the slope).

The economic significance of the main explanatory variable increases, as the coefficient on the variable rises from around 0.15 to 0.31. Namely, the current specification suggests that private forecasters increase their next-year forecast by 0.31 percentage point in response to a 1 percentage point increase in the difference between the BOJ forecast and ESP forecast.

At the same time, the sign of the coefficients for the added terms suggests that the impact of the Bank of Japan forecasts has been transformed since the introduction of the inflation-targeting policy. The interaction term in column 4 is negative, as is the coefficient for the IT dummy, statistically significantly so in the case of the IT dummy. This latter coefficient is robust to the deletion of the first two months' observations after the Lehman failure from the sample (column 5).¹⁸

¹⁸We also ran a separate set of (unreported) regressions using similar specifications for the BOJ and ESP GDP forecasts. In contrast to the effect on inflation forecasts, the impact on GDP forecasts and their determinants from the introduction of an inflation-targeting regime was minimal. The results are available upon request.

Numerical impacts are calculated as follows: *ceteris paribus*, a 1 percentage point increase in the central bank forecasts corresponds to around a 0.30 percentage point increase in those of the private sector (column 4). As discussed earlier with regard to equation (1), the wedge between BOJ and ESP inflation forecasts ($\pi_{t,ny}^{boj} - \pi_{t,ny}^{esp}$) increased by around 0.5 percentage point on average after the adoption of inflation targeting, which would imply a boost to ESP forecasts by 0.15 percentage point during the IT regime. At the same time, however, the coefficient on the IT dummy implies that the private sector is discounting the central bank forecasts by 0.10 percentage point.¹⁹ The calculation suggests that an increase in BOJ inflation forecasts after the adoption of inflation targeting likely raised ESP inflation forecasts by only a small margin (0.05 percentage point). Similar calculations using the coefficients when controlling for the Lehman episode (column 5) result in no margin left, i.e., private-sector forecasters completely discounted the increase in BOJ inflation forecasts from the start of the IT era.

Japan's limited experience with inflation targeting has for the most part coincided with quantitative and qualitative easing policies. A factor to keep in mind is that the private sector's forecasts for long-term inflation rates in Japan had been well below 2 percent for many years. The negative sign on the IT dummy coefficient likely reflected more pessimistic views among private-sector forecasters on the ability of measures to achieve the 2 percent inflation target from below—efforts which were in many respects unprecedented—while the Bank of Japan was focused on communication consistent with achieving its target, or so-called Odyssean forward guidance. These competing incentives may have made forecasting more difficult and hence led to a decline in accuracy of forecasts and lower confidence in BOJ forecasts.

¹⁹ An alternative interpretation of the result is that the introduction of IT may have influenced other variables, which account for the negative coefficient on the IT dummy beyond the change in central bank forecasts. However, tests do not support structural change in any of the other explanatory variables, nor is the null hypothesis of no change in the coefficients on the other explanatory variables in the regression rejected.

Another possibility is that the central banks' forecasting models for the overall macroeconomy simply deteriorated in 2013 with the introduction of variety of unprecedented monetary policy measures whose transmission mechanisms were untested, and there was less confidence in the precision of central bank's economic forecasts in general rather than inflation forecasts in particular. However, as discussed above, we do not find evidence for a structural break in GDP forecasts, nor any change in the influence of BOJ forecasts for GDP, at the start of the IT era.

5.2 *Alternative Specification*

As a robustness check, we report the results from running the alternative regressions using quarterly ESP forecasts instead of at the monthly frequency. As the Bank of Japan forecasts are mostly at the quarterly frequency, this allows for a differenced specification where the change in the Bank of Japan forecast is one of the explanatory variables.²⁰

Since the Bank of Japan forecast observations are only available at a lower frequency than the rest of the sample, the estimate of the impact of the Bank of Japan forecasts can be subject to noise using monthly data.²¹ On the other hand, by using quarterly data in a small sample, the researcher may lose some precision in the estimates of the determinants of the change in private-sector forecasts. With this caveat in mind, we examine the results for next-year forecasts in table 4, but using only those months for which the BOJ forecasts are available. We estimate the regressions in differences, where the dependent variable is the change in the private-sector forecast medians over the period, and the main explanatory variable

²⁰The alternative quarterly specification is estimated only over the time period during which the BOJ was issuing forecasts at a quarterly frequency (July 2008 onwards).

²¹This would be particularly the case if one expected the impact of the control variables to be different in periods with and without BOJ forecasts. However, statistical tests reject the hypothesis that the coefficients of the control variables differ in the periods when there are Bank of Japan forecast announcements.

**Table 4. Alternative Specification:
Next-Year Forecasts, Quarterly**

	(1)	(2)	(3)
Constant	-0.0039 (0.0109)	-0.0219 (0.0160)	-0.0215 (0.0161)
Inflation Surprise	0.2139* (0.1082)	0.2584** (0.1230)	0.2618** (0.1237)
Change in USD Forecast	0.0119** (0.0054)	0.0071 (0.0072)	0.0071 (0.0072)
Change in Oil Price Forecast	0.0010 (0.0021)	0.0010 (0.0020)	0.0007 (0.0026)
Change in BOJ Forecast	0.7564*** (0.0989)	0.7116*** (0.0883)	0.7043*** (0.0916)
Change in BOJ Forecast*Dummy IT		0.1317 (0.9482)	0.1446 (0.1573)
Dummy IT		0.0413* (0.0220)	0.0409* (0.0220)
Obs.	34	34	33
R ²	0.8553	0.8718	0.8520
Adj. R ²	0.8353	0.8433	0.8178

Notes: Equation (3) excludes the first observation after Lehman Brothers' bankruptcy. Changes in the USD forecast and oil price forecast are computed for periods corresponding to the quarterly changes in the ESP forecast. Figures in parentheses indicate standard errors. ***, **, and * indicate significance levels at 1 percent, 5 percent, and 10 percent, respectively.

is now the change in the Bank of Japan forecasts over the period between forecasts.²²

$$\Delta\pi_{t,ny}^{esp} = Constant + \gamma_1 InfSurp_t + \gamma_2 \Delta e_{t,ny}^{esp} + \gamma_3 \Delta oil_t^{ny} + \gamma_4 \Delta \pi_{t,ny}^{boj} + u_t \tag{4}$$

As in the baseline regressions, a dummy for the IT regime, as well as a variable interacting this regime dummy with the main explanatory variable of interest—in this case, the change in the BOJ inflation forecasts—are included in latter specifications.

²²While the breakpoint Chow test detects a structural shift in both $\pi_{t,ny}^{boj}$ and $\pi_{t,ny}^{esp}$ as previously discussed, the same Chow test does not find a shift in their first difference ($\Delta\pi_{t,ny}^{boj}$ and $\Delta\pi_{t,ny}^{esp}$). This can happen if $\pi_{t,ny}^{boj}$ and $\pi_{t,ny}^{esp}$ have a one-time stepwise shift. Thus, the differenced explanatory variable would lead us to expect a different impact from the IT dummy in table 4 than in table 3.

Dependent-variable own lag is not included in the quarterly specification, as private forecasters do not appear to be adjusting their forecasts at such long lags; further, Durbin-Watson statistics close to 2 for the key specifications of table 4's regressions provide no evidence that the error terms are positively autocorrelated.

The impact of the change in Bank of Japan forecasts is statistically significant, with coefficients of around 0.70–0.76, suggesting that more than two-thirds of changes in the BOJ forecasts are passed through to changes in the ESP forecasts (table 4, columns 1–3). The adjusted R-squared of over 0.8 in all specifications suggests high degrees of explanatory power. The signs of the control variable coefficients are unchanged, and generally the same control variables that are statistically significant in the earlier regressions are also significant in the quarterly difference regressions.

Important points to notice are (i) the coefficient on the variable interacting the BOJ forecast change with the IT dummy is not statistically significant, and (ii) the IT dummy on its own is positive, but at 0.04 (columns 2–3) corresponds to a miniscule 0.01 percentage point on a monthly frequency. The first point is consistent with the observation that the interactive variable is also not significant in the earlier baseline specification (table 3, columns 4–5). The second point is also consistent with the baseline result that private-sector inflation forecasts were not raised meaningfully even after the Bank of Japan raised its inflation forecasts upon the adoption of inflation targeting. That is, in the baseline specification, the impact of the wider wedge between BOJ and ESP inflation forecasts was effectively cancelled by the level shift. For these reasons, this alternative specification is consistent with the baseline specification's result that private-sector forecasters discounted the increase in BOJ inflation forecasts in the IT era. However, that aspect is not so clearly seen in this alternative specification, as its explanatory variable, the first difference in the BOJ's inflation forecast $\Delta\pi_{t,ny}^{boj}$, largely conceals the shift in its level $\pi_{t,ny}^{boj}$ (see footnote 22).

6. Conclusion

The impact of central bank inflation forecasts on those of the private sector can be influenced by the introduction of an inflation-targeting

regime in numerous ways. If the target is particularly credible, the usefulness of the central bank forecasts might be reduced due to their diminished information value. But if the target is not viewed as achievable, and central bank forecasts are viewed as influenced by the target, once again the usefulness of the forecasts might be affected.

We argue that our results are more consistent with the latter channel: there was a structural upward shift in BOJ inflation forecasts following the adoption of inflation targeting in 2013—reflecting the incentive of the central bank to communicate its intent to achieve the target—that affected their use by the private sector. The fact that forecast assumptions were changed at the time to include the central bank’s judgment of “the effects of past policy decisions” was yet another aspect of the IT regime that could have diminished their value to the private sector. The decline in the accuracy of central bank forecasts in the IT era versus those of the private sector is consistent with such a structural shift. And the systematic *downward* discounting of the central bank forecasts that followed suggests that private-sector forecasters likely viewed the BOJ forecasts as upwardly biased. By contrast, the inability of private-sector expectations of inflation to rise beyond 1.5 percent for any extended period after the announcement of the 2 percent inflation target is *prima facie* evidence that it wasn’t the introduction of a *credible* target that could have been responsible for any change in the influence of central bank forecasts.

We view Japan’s situation as increasingly relevant and the results as generally useful. Since the global financial crisis, inflation levels in both advanced and many emerging economies have been persistently weak and below established targets. One after another, advanced economies adopted unconventional monetary policies whose effectiveness was untested. Further, the inflation forecasts of many monetary authorities, including the U.S. Federal Reserve, have repeatedly been higher than both observed inflation and the forecasts of the market. One renowned scholar and Fed watcher has even suggested that market participants might see the Federal Reserve forecasts “as a disconnect from reality” (Summers 2016). The undershooting of inflation outcomes from the forecasts and targets laid out by central banks is by now a widespread phenomenon, which can hardly be viewed as unique to Japan.

Thus, this case study gives us general insights into the relationship between inflationary expectations and central bank and private-sector forecasts, as well as the impact of different monetary policy regimes, especially when the targeted inflation rate is higher than the expected inflation rate and the nominal policy rate is close to an effective lower bound. We hope our findings here will stimulate further research on the impact of central bank forecasts under different policy regimes, as well as the tradeoffs that monetary authority may face when issuing the forecasts.

Appendix

Table A.1. Variable Description

Variable	Variable Description	Sources
$\pi_{t,ny}^{esp}$	ESP inflation forecast at time t for next year, in percent.	JCER
$\pi_{t-1,ny}^{boj}$	The latest BOJ inflation forecast for next year known to ESP survey respondents when they make forecasts at time t , in percent.	BOJ
$\Delta\pi_{t,ny}^{esp}$	Change in ESP inflation forecast between time $t - 1$ and t for next year, in percentage points.	JCER; authors' calculations
$\Delta\pi_{t,ny}^{boj}$	Change in BOJ inflation forecast for next year (quarterly in the alternative specification), in percentage points.	BOJ; authors' calculations
$\pi_{t-1,ny}^{boj} - \pi_{t-1,ny}^{esp}$	The latest BOJ inflation forecast for next year known to ESP survey respondents when they make forecasts at time t minus ESP inflation forecast for next year at time $t-1$.	JCER; BOJ; authors' calculations
$\Delta e_{t,ny}^{esp}$	Log change in ESP JPY/USD exchange rate forecast between time $t - 1$ and t for next year, in percent. A positive change indicates depreciation of JPY is expected.	JCER; authors' calculations

(continued)

Table A.1. (Continued)

Variable	Variable Description	Sources
Δoil_t^{spot}	Log change in spot WTI oil price between time $t - 1$ and t , in percent.	Bloomberg; authors' calculations
Δoil_t^{ny}	Log change in the average of prices of WTI oil futures with deliveries in next fiscal year, between time $t-1$ and t , in percent. Namely, the log change in the average of future prices of contracts to be delivered in each month of the next fiscal year. The average of future prices is calculated as $[F(\text{Apr})+F(\text{May})+\dots+F(\text{Feb})+F(\text{Mar})]/12$, where $F(\cdot)$ represents the future price of contract to be delivered in a particular month.	Bloomberg; authors' calculations
$InfSurp_t$	Core inflation surprise known at time t , defined as realized quarterly inflation at time t minus quarterly inflation forecasted prior to the release of realized figures, in percent.	Statistics Bureau of Japan; JCER; authors' calculations
$DumIT$	Dummy variable for inflation target period, equal to 1 for ESP surveys from February 2013 onwards, and 0 otherwise.	
$DumTaxDelay$	Two separate dummy variables are included to control for delays of consumption tax hike, one equal to 1 for ESP survey of December 2014, and 0 otherwise, the other equal to 1 for the ESP survey of June 2016, and 0 otherwise.	

References

- Andrade, P., G. Gaballo, E. Mengus, and B. Mojon. 2018. "Forward Guidance and Heterogeneous Beliefs." BIS Working Paper No. 750 (October).
- Blinder, A. S., M. Ehrmann, M. Fratzscher, J. De Haan, and D. Jansen. 2008. "Central Bank Communication and Monetary

- Policy: A Survey of Theory and Evidence.” *Journal of Economic Literature* 46 (4): 910–45.
- Bomfim, A. N., and G. D. Rudebusch. 2000. “Opportunistic and Deliberate Disinflation under Imperfect Credibility.” *Journal of Money, Credit and Banking* 32 (4, Part 1): 707–21.
- Campbell, J. R., C. L. Evans, J. D. M. Fisher, and A. Justiniano. 2012. “Macroeconomic Effects of Federal Reserve Forward Guidance.” *Brooking Papers on Economic Activity* (Spring): 1–80.
- Capistran, C., and M. Ramos-Francia. 2010. “Does Inflation Targeting Affect the Dispersion of Inflation Expectations?” *Journal of Money, Credit and Banking* 42 (1): 113–34.
- Cecchetti, S. G., and C. Hakkio. 2009. “Inflation Targeting and Private Sector Forecasts.” NBER Working Paper No. 15424.
- Christensen, J., and M. Spiegel. 2019. “Negative Interest Rates and Inflation Expectations in Japan.” FRBSF Economics Letter No. 2019-22, August 26.
- Coibion, O., and Y. Gorodnichenko. 2012. “What Can Survey Forecasts Tell Us about Information Rigidities?” *Journal of Political Economy* 120 (1): 116–59.
- Crowe, C. 2010. “Testing the Transparency Benefits of Inflation Targeting: Evidence from Private Sector Forecasts.” *Journal of Monetary Economics* 57 (2): 226–32.
- Dale, S., A. Orphanides, and P. Osterholm. 2011. “Imperfect Central Bank Communication — Information vs. Distraction.” *International Journal of Central Banking* 7 (2, June): 3–39.
- de Mendonca, H. F., and J. D. B. V. de Deus. 2019. “Central Bank Forecasts and Private Expectations: An Empirical Assessment from Three Emerging Economies.” *Economic Modelling* 83 (December): 234–44.
- Demertzis, M., and N. Hoeberichts. 2007. “The Costs of Increasing Transparency.” *Open Economies Review* 18 (3): 263–80.
- Demertzis, M., and N. Viegi. 2008. “Inflation Targets as Focal Points.” *International Journal of Central Banking* 4 (1, March): 55–87.
- Ehrbeck, T., and R. Waldmann. 1996. “Why Are Professional Forecasters Biased? Agency versus Behavioral Explanations.” *Quarterly Journal of Economics* 111 (1): 21–40.
- Ehrmann, M. 2015. “Targeting Inflation from Below: How Do Inflation Expectations Behave?” *International Journal of Central Banking* 11 (Supplement 1, September): 213–49.

- Ehrmann, M., S. Eijffinger, and M. Fratzscher. 2012. "The Role of Central Bank Transparency for Guiding Private Sector Forecasts." *Scandinavian Journal of Economics* 114 (3): 1018–52.
- Fujiwara, I. 2005. "Is the Central Bank's Publication of Economic Forecasts Influential?" *Economics Letters* 89 (3): 255–61.
- Hattori, M., and J. Yetman. 2017. "The Evolution of Inflation Expectations in Japan." *Journal of the Japanese and International Economies* 46 (December): 53–68.
- Hubert, P. 2014. "FOMC Forecasts as a Focal Point for Private Expectations." *Journal of Money, Credit and Banking* 46 (7): 1381–1420.
- . 2015. "Do Central Bank Forecasts Influence Private Agents? Forecasting Performance versus Signals." *Journal of Money, Credit and Banking* 47 (4): 771–89.
- Ito, T. 1990. "Foreign Exchange Rate Expectations: Micro Survey Data." *American Economic Review* 80 (3): 434–49.
- Kool, C., M. Middeldorp, and S. Rosenkranz. 2011. "Central Bank Transparency and the Crowding Out of Private Information in Financial Markets." *Journal of Money, Credit and Banking* 43 (4): 765–74.
- Laster, D., P. Bennett, and I. S. Geoum. 1999. "Rational Bias in Macroeconomic Forecasts." *Quarterly Journal of Economics* 114 (1): 293–318.
- Morris, S., and H. S. Shin. 2002. "Social Value of Public Information." *American Economic Review* 92 (5): 1521–34.
- Morris, S., H. S. Shin, and H. Tong. 2006. "Social Value of Public Information: Morris and Shin (2002) Is Actually Pro-Transparency, Not Con: Reply." *American Economic Review* 96 (1): 453–55.
- Nishizaki, K., T. Sekine, and Y. Ueno. 2014. "Chronic Deflation in Japan." *Asian Economic Policy Review* 9 (1): 20–39.
- Orphanides, A. 2018. "The Boundaries of Central Bank Independence: Lessons from Unconventional Times." Discussion Paper No. 2018-E-10, Institute for Monetary and Economic Studies, Bank of Japan.
- Ottaviani, M., and P. N. Sorensen. 2006. "The Strategy of Professional Forecasting." *Journal of Financial Economics* 81 (2): 441–66.

- Pedersen, M. 2015. "What Affects the Predictions of Private Forecasters? The Role of Central Bank Forecasts in Chile." *International Journal of Forecasting* 31 (4): 1043–55.
- Romer, C., and D. Romer. 2000. "Federal Reserve Information and the Behavior of Interest Rates." *American Economic Review* 90 (3): 429–57.
- Shirakawa, M. 2012. "Toward Sustainable Growth with Price Stability." Speech at the Kisaragi-kai Meeting in Tokyo, November 12.
- . 2018. *Central Bank — My Insider Recollections of 39 Years* (in Japanese) (authors' translation). Toyo Keizai Inc.
- Siklos, P. 2013. "Sources of Disagreement in Inflation Forecasts: An International Empirical Investigation." *Journal of International Economics* 90 (1): 218–31.
- Summers, L. 2016. "Jackson Hole: The Fed Has Not Inspired Faith in Its Policy Framework." Blog Entry, *Financial Times*, August 29.
- Townsend, R. M. 1983. "Forecasting the Forecasts of Others." *Journal of Political Economy* 91 (4): 546–88.

On the Optimal Labor Income Share*

Jakub Growiec,^a Peter McAdam,^b and Jakub Mućk^a
^aSGH Warsaw School of Economics and Narodowy Bank Polski
^bEuropean Central Bank

Labor’s income share has attracted interest reflecting its decline. But, from an efficiency standpoint, can we say what share would hold in the social optimum? We address this question using a microfounded endogenous growth model calibrated on U.S. data. In our baseline case the socially optimal labor share is 17 percent (11 percentage points) above the decentralized (historical) equilibrium. This wedge reflects the presence of externalities in R&D in the decentralized equilibrium, whose importance is conditioned by the degree of factor substitutability. We also study the dependence of both long-run growth equilibriums on different model parameterizations and relate our results to Piketty’s “Laws of Capitalism.”

JEL Codes: O33, O41.

1. Introduction

Although interest in labor’s share of income has a long tradition in economics, current interest has crystallized around its apparent fall in recent decades across many countries.¹ Much of the public discussion suggests that the primary reason to be concerned with

*We thank two anonymous referees and the editor Keith Kuester for valuable comments, as well as those of numerous seminar and workshop participants. We also gratefully acknowledge financial support from the Polish National Science Center (Narodowe Centrum Nauki) under the grant Opus 14 No. 2017/27/B/HS4/00189. Jakub Mućk also gives thanks for financial support from the Foundation for Polish Science within the START fellowship. The opinions expressed in this paper are those of the authors and do not necessarily reflect views of the European Central Bank, Narodowy Bank Polski, or the Eurosystem. Corresponding author (McAdam) e-mail: peter.mcadam@ecb.europa.eu.

¹The scope of economic interest in the labor share is extremely diverse—aside from the conventional political economy, inequality, and growth literatures, the labor share is an important consideration in inflation modeling (Galí and Gertler 1999; McAdam and Willman 2004) and the consequence of firm market power, de Loecker, Eeckhout, and Unger (2020).

a low labor income share is wealth inequality. That may be a valid concern. The current paper however provides another rationale for interest in the labor share. We demonstrate that even in a representative household model (in which wealth inequality is absent by definition), the level of the labor share determines whether the economy is allocating its resources efficiently. Thus, we can say that the labor share may be too low (or high)—even without distributional concerns.

Remarkably, there appears to be no investigation of this issue in the literature. This contrasts with equivalent discussions in the growth literature: since Ramsey (1928), the question of whether a decentralized economy saves “too little” is fundamental (e.g., de La Grandville 2012). Likewise, in terms of production, modern endogenous growth theory typically suggests that the presence of various distortions implies that the economy produces less output and less research and development (R&D) relative to the first-best allocation (the “social optimum”) (e.g., Jones and Williams 2000; Alvarez-Pelaez and Groth 2005). These *distortions* include monopoly power and markups plus the existence of technological externalities. Their presence can mean that individual firms have weak incentives to work to their fullest capacity, or indeed to invest and innovate, if not all of the benefits accrue to them.

But what of the labor share of income? Does, for example, “too little” output in the decentralized economy translate into a labor share that is also somehow too low? Ex ante, it is by no means obvious. Given widespread interest in the labor share, this constitutes an important gap in our knowledge, which we seek to address.² In the context of a microfounded endogenous growth model calibrated on U.S. data, with labor- and capital-augmenting technical change in the aggregate production function, we find that in our baseline case the socially optimal labor share is markedly above the decentralized equilibrium. The decentralized labor share, in other words, is too low.

The key channels underlying that result are the following. First, the social planner saves more and thus has more physical capital in

²Note, we not only study the implications for the labor share but also the growth rate, employment in the research sector, consumption, capital accumulation, etc.

the long run. This is because the social planner takes into account the *social* rather than purely private return on production capital, and because the planner is able to internalize the positive returns to capital used in R&D. This abundance of capital makes labor relatively scarce. If, and only if, both factors have a substitution elasticity in production below one (i.e., are gross complements), this capital abundance pushes the labor share up.

Second, in comparison with the decentralized allocation, the social planner tends to allocate relatively more labor to the R&D sectors and less to the (final good) production sector. Because labor-augmenting R&D is the ultimate source of per capita growth, this increases the long-run growth rate. Yet, even though the social planner increases labor-augmenting R&D relative to the decentralized allocation, it increases capital-augmenting R&D even more. The ratio of capital- to labor-augmenting R&D is always higher in the social planner allocation. This effect increases unit productivity of capital in the steady state, augmenting both production and R&D sectors and feeding back once again to the steady-state growth rate and the labor share.

Another appealing aspect of our framework is that it can provide new convincing results on one important strand of the literature on the labor share, namely Piketty's "Laws of Capitalism" (Piketty 2014). These predict that the capital-output ratio and the capital income share should increase whenever the pace of economic growth declines. We say "convincing results" because in our setup all relevant variables (factor accumulation and factor intensity, growth, technological progress, marginal product of capital and thus the interest rate, etc.) are endogenous and modeled in a sufficiently flexible manner. Our analysis underscores that Piketty's laws should not be interpreted causally, but rather as correlations generated by changes in deeper characteristics of the economy. Crucially, though, in fact these correlations are not guaranteed to hold at all. For example, we find that increases in factor substitutability are able to raise the capital-output ratio, the capital income share, and the economic growth rate, thus simultaneously *violating* both of Piketty's laws.

Finally we demonstrate that, through the lens of our model, a fluctuating labor share would not in itself necessarily be a sign of inefficiency. Rather, the planner's solution shows that volatility in

Figure 1. Historical Labor Share: United States (1899–2010) and France (1897–2010)



Notes: The U.S. data are taken from Piketty and Zucman (2014) over the sample 1929–2010; prior to 1929 the labor share is extrapolated using the database by Groth and Madsen (2016), which provides compensation of employees and value-added data starting in 1898 based on historical source provided by Liesner (1989). The French data are also taken from Piketty and Zucman (2014). The dashed line is the level of the labor income share and the solid line is a simple moving-average process approximating its trend characteristics: $1/10 \sum_{j=-5}^4 ls_{t-j}$ where ls_t is the labor share. See also Charpe, Bridji, and McAdam (2020) for a discussion and analysis of the properties of historical labor share measures.

this share is a natural outcome. Indeed, we know from historical sources such as Piketty and Zucman (2014) that labor income shares, even over long horizons (e.g., above 100 years), can fluctuate considerably (see figure 1 for the United States and France; see also Charpe, Bridji, and McAdam 2020).³ By comparison, can we describe the decentralized labor share as being characterized by excessive volatility? The reasons an optimal allocation would produce oscillations, too, relate to the fact that there is an entrenched tension between capital- and labor-augmenting technologies. Labor-augmenting developments generate economic growth but also make capital relatively scarce, necessitating a reallocation of resources towards capital to overcome this scarcity. By the same token,

³Both aspects matter for any normative discussion on the labor share. For instance, if the labor share is falling yet still above its “optimal” level (or fluctuating around it), then, arguably, this might be interpreted passively, as a manifestation of recognized fluctuations in factor shares (e.g., Mućk, McAdam, and Growiec 2018). Indeed, given that long and persistent fluctuations in the labor share are observed in practice, we might also wonder whether such fluctuations are socially optimal.

capital-augmenting developments make labor relatively scarce and trigger the opposite reallocation.

The paper is organized as follows. Section 2 describes the model (also contained in Growiec, McAdam, and Mućk 2018). This is a non-scale model of endogenous growth with two R&D sectors, giving rise to capital- as well as labor-augmenting innovations, drawing from the seminal contributions of Romer (1990) and Acemoglu (2003).⁴ The model economy uses the Dixit-Stiglitz monopolistic competition setup and the increasing variety framework of the R&D sectors. Two R&D sectors are included to enable an endogenous determination of factor shares. Both the social planner and decentralized allocations are solved for and compared. We see that the presence of markups arising from imperfect competition (and market power) and technological and R&D externalities, are the key reasons why the decentralized allocation produces relatively lower output growth and labor share.

Section 3 calibrates the model to U.S. data. We assume that a range of long-run averages from U.S. data (evaluated over 1929–2015) correspond to the decentralized balanced growth path (BGP) of the model. Around this central calibration, though, we extensively examine robustness of our result to alternative parameterizations.

Thereafter, in section 4 we solve the BGP of each allocation (i.e., decentralized and first best) and compare them. We list the channels and assumptions underlying the differences between both allocations. We find that—assuming that factors are gross complements in production—the decentralized labor share is indeed socially suboptimal. The difference, moreover, is large: about 17 percent (11 percentage points). We describe the mechanisms which underlie this wedge. For robustness, we also consider production characterized by Cobb-Douglas as well as gross substitutes. In the latter case, and almost only in that case, the socially optimal labor share falls below the decentralized one. However, already for $\sigma = 1.25$, which constitutes a mild degree of gross substitutability, its value is counterfactually

⁴The term “scale effect” states that an increase in an economy’s labor endowment leads to a higher real growth rate. This relation arises from the (counterfactual) assumption that growth is proportional to the number of R&D workers (Jones 1995).

low (at around 0.5) and also associated with counterfactually high per capita growth rates.⁵

In section 5 we also study the dependence of both long-run growth equilibriums on model parameters and relate our results to Piketty's "Laws of Capitalism." We also consider the dynamic properties of the model around the balanced growth path (both in the decentralized and optimal allocation) in terms of oscillatory dynamics. Section 6 concludes. Additional material is found in the appendixes.

Finally, note that while making a first attempt at a new research question, we abstract from several issues. First, to repeat a remark made earlier, our concern is not about inequality among heterogeneous agents; there are many papers on this topic. Indeed, although the labor share and inequality are clearly related, they are by no means interchangeable (Atkinson, Piketty, and Saez 2011); an economy may well exhibit a socially optimal factor income division yet still be characterized by considerable inequality—as, for example, if there are different skill characteristics in the labor force (and thus appreciable wage dispersion), asymmetric corporate or union insider power, or if there is financial repression and rent seeking, etc. Indeed, one can draw an analogy with Ramsey (1928)—whose concern lay with the level of the socially optimal aggregate savings rate, not how savings behavior is distributed across economic agents (such as by wealth, age percentiles, etc.). Our concern therefore is somewhat more straightforward—namely, how would a social planner choose functional income shares. And, would that share be realistic⁶ in terms of its central value (relative to the decentralized optimum) and its volatility (again compared with the decentralized optimum and historical averages). Second, we do not discuss policy designs able to alleviate the discrepancy between the decentralized allocation and the social optimum, nor the dynamics with which

⁵It may be, as Piketty and Zucman (2014) argue, that one might expect a higher elasticity of substitution in "high-tech" economies where there are lots of alternative uses and forms for capital.

⁶By way of realism, consider another "optimal" rule in growth theory: namely the golden rule savings ratio which in standard form equates the optimal savings rate with the capital income share (which is usually around 30 percent); see de la Grandville (2012). With the exception of some Asian economies and for some particular periods, such values are highly counterfactual.

they could be introduced.⁷ Our results are obtained by comparing long-run equilibriums of two entirely separate model economies (decentralized and first-best allocation). Therefore we are silent on the possible evolution of the labor share along the transition path following the introduction of policy measures able to shift the decentralized allocation towards the first best. In consequence, we cannot say (i) if the labor share should rise or fall in the short to medium run, and (ii) how long the transition to the optimal labor share should take.

2. Model

The framework is a generalization of Acemoglu (2003) with capital- and labor-augmenting R&D, building on the earlier induced innovation literature from Kennedy (1964) onwards as well as general innovation in monopolistic competition and growth literatures (e.g., Dixit and Stiglitz 1977, Romer 1990, and Jones 1999).

By “generalization” we mean that we relax a number of features to make our conclusions more applicable to the studied question, as well as to correct for some counterfactual features (such as the aforementioned scale effects). Formally, (i) our model is non-scale: both R&D functions are specified in terms of percentages of population employed in either R&D sector; (ii) we also assume R&D workers are drawn from the same pool as production workers;⁸ (iii) we assume more general R&D technologies which allow for mutual spillovers between both R&D sectors (cf. Li 2000) and for concavity in capital-augmenting technical change; (iv) in contrast to Acemoglu (2003), the BGP growth rate in our model depends on preferences via employment in production and R&D—the tradeoff is due to drawing researchers from the same employment pool as production

⁷Interestingly, Atkinson (2015) lists a number of proposals for reducing inequality trends, the first of which is that “the direction of technical change should be an explicit concern of policy-makers.”

⁸Acemoglu (2003) assumes that labor supply in the production sector is inelastic and R&D is carried out by a separate group of “scientists” who cannot engage in production labor. Our assumption affects the tension between both R&D sectors by providing R&D workers with a third option, the production sector.

workers (a tradeoff not present in his model); and (v) we use *normalized* constant elasticity of substitution (CES) production functions⁹ which, importantly, ensures valid comparative static comparisons in the elasticity of factor substitution. To start matters off, we consider the simpler case of the social planner allocation.¹⁰

2.1 The Social Planner's Problem

The social planner maximizes the representative household's utility from discounted consumption, c , given standard constant relative risk aversion (CRRA) preferences, (1).

$$\max \int_0^{\infty} \frac{c^{1-\gamma} - 1}{1-\gamma} e^{-(\rho-n)t} dt, \quad (1)$$

where $\gamma > 0$ is the inverse of the intertemporal elasticity of substitution, $\rho > 0$ is the rate of time preference, and $n > 0$ is the (exogenous) growth rate of the labor supply.

The maximization is subject to the budget constraint (2) (i.e., the equation of motion of the aggregate per capita capital stock k), the "normalized" production function (3), the two R&D technologies (4)–(5), and the labor market clearing condition (6):¹¹

$$\dot{k} = y - c - (\delta + n)k - \zeta \dot{a}, \quad (2)$$

$$y = y_0 \left(\pi_0 \left(\lambda_b \frac{k}{k_0} \right)^\xi + (1 - \pi_0) \left(\frac{\lambda_a}{\lambda_{a0}} \frac{\ell_Y}{\ell_{Y0}} \right)^\xi \right)^{1/\xi}, \quad (3)$$

⁹Normalization essentially implies representing the production relations in consistent index number form. Its parameters then have a direct economic interpretation. Otherwise, the parameters can be shown to be scale dependent (i.e., a circular function of σ itself, as well as a function of the implicit normalization points). Subscript 0's denote the specific normalization points: geometric (arithmetic) averages for non-stationary (stationary) variables. See de la Grandville (1989), Klump and de la Grandville (2000), and Klump and Preissler (2000) for the seminal theoretical contributions. In our case, normalization is essentially important, since comparative statics on production function parameters are a key concern.

¹⁰It is simpler because, solving under the social optimum, we can impose symmetry directly and deal in terms of aggregates; see Bénassy (1998).

¹¹There are three control (c, ℓ_a, ℓ_b) and three state (k, λ_a, λ_b) variables in this optimization problem.

$$\dot{\lambda}_a = A \left(\lambda_a \lambda_b^\phi x^{\eta_a} \ell_a^{\nu_a} \right), \tag{4}$$

$$\dot{\lambda}_b = B \left(\lambda_b^{1-\omega} x^{\eta_b} \ell_b^{\nu_b} \right) - d\lambda_b, \tag{5}$$

$$1 = \ell_a + \ell_b + \ell_Y. \tag{6}$$

In (2) and (3), $y = Y/L$ and $k = K/L$ (i.e., output and capital per capita), where L is total employment and ℓ_a and ℓ_b are the shares (or “research intensity”) employed in labor- and capital-augmenting R&D, respectively (and, respectively, generating increases in λ_a and λ_b). The remaining fraction of population ℓ_Y is employed in production. We assume that capital augmentation is subject to gradual decay at rate $d > 0$, which mirrors susceptibility to obsolescence and embodied character of capital-augmenting technologies; Solow (1960). This assumption is critical for the asymptotic constancy of unit capital productivity λ_b in the model, and thus for the existence of a BGP with purely labor-augmenting technical change.

The term π denotes the capital income share, and $\xi = \frac{\sigma-1}{\sigma}$, where $\sigma \in [0, \infty)$ is the elasticity of substitution between capital and labor. This parameter, important in many contexts,¹² turns out also to be critical in our analysis with the distinction as to whether factors are gross complements, i.e., $\sigma < 1$, or gross substitutes, $\sigma > 1$, in production.

Factor-augmenting innovations are created endogenously by the respective R&D sectors (Acemoglu 2003), increasing the underlying parameters λ_a, λ_b , as in (4) and (5). Parameters A and B capture the unit productivity of the labor- and capital-augmenting R&D process, respectively; ϕ captures the spillover from capital- to labor-augmenting R&D;¹³ and ω measures the degree of decreasing returns

¹²CES function (3) nests the linear, Cobb-Douglas, and Leontief forms, respectively, when $\xi = 1, 0, -\infty$. The value of the elasticity of factor substitution has been shown to be a key parameter in many economic fields: e.g., the gains from trade (Saam 2008); the strength of extensive growth (de La Grandville 2016); multiple growth equilibriums, development traps, and indeterminacy (Azariadis 1996; Klump 2002; Kaas and von Thadden 2003; Guo and Lansing 2009); the response of investment and labor demand to various policy changes and shocks (Rowthorn 1999); etc.

¹³We assume $\phi > 0$, indicating that more efficient use of physical capital also increases the productivity of labor-augmenting R&D. Observe, there are *mutual* spillovers between both R&D sectors, with no prior restriction on their strength: $\dot{\lambda}_a = A\lambda_a^{1-\eta_a} \lambda_b^{\phi+\eta_a} k^{\eta_a} \ell_a^{\nu_a}$ and $\dot{\lambda}_b = B\lambda_a^{-\eta_b} \lambda_b^{1-\omega+\eta_b} k^{\eta_b} \ell_b^{\nu_b} - d\lambda_b$.

to scale in capital-augmenting R&D. By assuming $\omega \in (0, 1)$ we allow for the “standing on shoulders” effect in capital-augmenting R&D, albeit we limit its scope insofar as it is less than proportional to the existing technology stock (Jones 1995).

The term $x \equiv \frac{\lambda_b k}{\lambda_a}$ captures the technology-corrected degree of capital augmentation of the workplace. This term represents positive spillovers from capital intensity in the R&D sector and will be constant along the BGP. The long-term endogenous growth engine is located in the linear labor-augmenting R&D equation. To fulfill the requirement of the existence of a BGP along which the growth rates of λ_a and λ_b are constant, we assume that $\eta_b \phi + \eta_a \omega \neq 0$.¹⁴ Note that the above parameterization of R&D equations, with six free parameters in equations (4)–(5), is the most general one possible under the requirement of existence of a BGP with purely labor-augmenting technical change (Uzawa 1961; Jones 1999; Acemoglu 2003; Growiec 2007).

The last term in (2) captures a negative externality that arises from implementing new labor-augmenting technologies, with $\zeta \geq 0$. Motivated by León-Ledesma and Satchi (2019), we allow for a non-negative cost of adopting new labor-augmenting technologies: since workers (as opposed to machines) need to develop skills compatible with each new technology, it is assumed that there is a capital cost of such technology shifts (potentially representing training costs, learning-by-doing, etc.). We posit that new capital investments are diminished by $\zeta \dot{a}$, where $\dot{a} = g \lambda_a \left(\frac{\pi}{\pi_0} \right)^{1/\alpha}$, g being the economic growth rate (Growiec, McAdam, and Mućk 2018). For analytical simplicity we consider these costs exogenous to the firms.

Finally, R&D activity may be subject to duplication externalities; the greater the number of researchers searching for new ideas, the more likely is duplication. Thus research effort may be characterized by diminishing returns; Kortum (1993). This is captured by parameters $\nu_a, \nu_b \in (0, 1]$: the higher is ν , the lower the extent of duplication.¹⁵

¹⁴All our qualitative results also go through for the special case $\eta_a = \eta_b = 0$, which fully excludes capital spillovers in R&D. The current inequality condition is not required in such cases.

¹⁵Observe that switching off all externalities and spillovers in (4)–(5) by setting $d = \omega = \eta_a = \eta_b = 0$ and $\nu_a = \nu_b = 1$ retrieves the original specification of R&D

Variables with subscript 0 ($\pi_0, y_0, k_0, \lambda_{a0}, \ell_{Y0}$) are CES normalization constants.

2.2 *Decentralized Allocation*

The construction of the decentralized allocation draws from Romer (1990), Acemoglu (2003), and Jones (2005). It has been also presented in Growiec, McAdam, and Mućk (2018). We use the Dixit and Stiglitz (1977) monopolistic competition setup and the increasing variety framework of the R&D sector. The general equilibrium is obtained as an outcome of the interplay between households; final goods producers; aggregators of bundles of capital- and labor-intensive intermediate goods; monopolistically competitive producers of differentiated capital- and labor-intensive intermediate goods; and competitive capital- and labor-augmenting R&D firms.

2.2.1 *Households*

Analogous to the social planner's allocation, we again assume that the representative household maximizes discounted CRRA utility:

$$\max \int_0^{\infty} \frac{c^{1-\gamma} - 1}{1-\gamma} e^{-(\rho-n)t} dt$$

subject to the budget constraint:

$$\dot{v} = (r - \delta - n)v + w - c, \quad (7)$$

where $v = V/L$ is the household's per capita holding of assets, $V = K + p_a \lambda_a + p_b \lambda_b$. The representative household is the owner of all capital and also holds the shares of monopolistic producers of differentiated capital- and labor-intensive intermediate goods (priced p_a and p_b , respectively). Capital is rented at a net market rental rate

in Acemoglu (2003). Moreover, compared with models which use Cobb-Douglas production, equation (5) is akin to Jones's (1995) formulation of the R&D sector, generalized by adding obsolescence and positive spillovers from capital intensity. Thus, setting $d = \eta_b = 0$ retrieves Jones's original specification. And (4) is the same as in Romer (1990) but scale free (it features ℓ_b instead of $\ell_b \cdot L$), with a positive spillover from capital intensity and a direct spillover from λ_b ; setting $\phi = \eta_a = 0$ retrieves the scale-free version of Romer (1990), cf. Jones (1999).

equal to the gross rental rate after depreciation: $r - \delta$. In turn, w is the market wage rate. Solving the household's optimization problem yields the familiar Euler equation:

$$\hat{c} = \frac{r - \delta - \rho}{\gamma}, \quad (8)$$

where $\hat{c} = \dot{c}/c$ is the per capita growth rate ("hats" denote growth rates).

2.2.2 Final Goods Producers

The role of final goods producers is to generate the output of final goods (which are then either consumed by the representative household or saved and invested, leading to physical capital accumulation), taking bundles of capital- and labor-intensive intermediate goods (Y_K, Y_L) as inputs. They operate in a perfectly competitive environment, where both bundles are remunerated at market rates p_K and p_L , respectively.

The final goods producers operate a normalized CES technology:

$$Y = Y_0 \left(\pi_0 \left(\frac{Y_K}{Y_{K0}} \right)^\xi + (1 - \pi_0) \left(\frac{Y_L}{Y_{L0}} \right)^\xi \right)^{\frac{1}{\xi}}. \quad (9)$$

The optimality condition implies that final goods producers' demand for capital- and labor-intensive intermediate goods bundles satisfies

$$\frac{p_K}{p_L} = \frac{\pi}{1 - \pi} \frac{Y_L}{Y_K}, \quad (10)$$

where $\pi = \pi_0 \left(\frac{Y_K}{Y_{K0}} \frac{Y_0}{Y} \right)^\xi$ is the elasticity of final output with respect to Y_K (in equilibrium it will be equal to the labor share).

2.2.3 Aggregators of Capital- and Labor-Intensive Intermediate Goods

There are two symmetric sectors whose role is to aggregate the differentiated (capital- or labor-intensive) goods into the bundles Y_K and Y_L demanded by final goods producers. It is assumed that the

differentiated goods are imperfectly substitutable (albeit gross substitutes). The degree of substitutability is captured by parameter $\varepsilon \in (0, 1)$:

$$Y_K = \left(\int_0^{N_K} X_{K_i}^\varepsilon di \right)^{\frac{1}{\varepsilon}}. \tag{11}$$

Aggregators operate in a perfectly competitive environment and decide upon their demand for intermediate goods, the price of which will be set by the respective monopolistic producers (discussed in the following subsection).

For capital-intensive bundles, the aggregators maximize

$$\max_{X_{K_i}} \left\{ p_K \left(\int_0^{N_K} X_{K_i}^\varepsilon di \right)^{\frac{1}{\varepsilon}} - \int_0^{N_K} p_{K_i} X_{K_i} di \right\}. \tag{12}$$

There is a continuum of measure N_K of capital-intensive intermediate goods producers. Optimization implies the following demand curve:

$$X_{K_i} = x_K(p_{K_i}) = \left(\frac{p_{K_i}}{p_K} \right)^{\frac{1}{\varepsilon-1}} Y_K^{\frac{1}{\varepsilon}}. \tag{13}$$

Equivalent terms follow for labor-intensive intermediate goods producers.

2.2.4 Producers of Differentiated Intermediate Goods

It is assumed that each of the differentiated capital- or labor-intensive intermediate goods producers, indexed by $i \in [0, N_K]$ or $i \in [0, N_L]$, respectively, has monopoly over its specific variety. It is therefore free to choose its preferred price p_{K_i} or p_{L_i} . These firms operate a simple linear technology, employing either only capital or only labor.

For the case of *capital-intensive* intermediate goods producers, the production function is $X_{K_i} = K_i$. Capital is rented at the gross rental rate r . The optimization problem is

$$\max_{p_{K_i}} (p_{K_i} X_{K_i} - r K_i) = \max_{p_{K_i}} (p_{K_i} - r) x_K(p_{K_i}). \tag{14}$$

The optimal solution implies $p_{Ki} = r/\varepsilon$ for all $i \in [0, N_K]$. This implies symmetry across all differentiated goods: they are sold at equal prices, thus their supply is also identical, $X_{Ki} = \bar{X}_K$ for all i . Market clearing implies

$$K = \int_0^{N_K} K_i di = \int_0^{N_K} X_{Ki} di = N_K \bar{X}_K, \quad Y_K = N_K^{\frac{1-\varepsilon}{\varepsilon}} K. \tag{15}$$

The demand curve implies that the price of intermediate goods is linked to the price of the capital-intensive bundle as in $p_K = p_{Ki} N_K^{\frac{\varepsilon-1}{\varepsilon}} = \frac{r}{\varepsilon} N_K^{\frac{\varepsilon-1}{\varepsilon}}$.

The *labor-intensive* sector follows symmetrically: $X_{Li} = L_{Yi}$, $L_Y = \ell_Y L = \int_0^{N_L} L_{Yi} di$, and $p_{Li} = w/\varepsilon$, $p_L = p_{Li} N_L^{\frac{\varepsilon-1}{\varepsilon}} = \frac{w}{\varepsilon} N_L^{\frac{\varepsilon-1}{\varepsilon}}$, where w is the market wage rate.

Aggregating across all intermediate goods producers, we obtain that their total profits are equal to $\Pi_K N_K = rK \left(\frac{1-\varepsilon}{\varepsilon}\right)$ and $\Pi_L N_L = wL_Y \left(\frac{1-\varepsilon}{\varepsilon}\right)$ for capital- and labor-intensive goods, respectively. Streams of profits per person in the representative household are thus $\pi_K = \Pi_K/L$ and $\pi_L = \Pi_L/L$, respectively. Hence, the total remuneration channeled to the capital-intensive sector equals $p_K Y_K = \frac{r}{\varepsilon} K = rK + \Pi_K N_K$, whereas the total remuneration channeled to the labor-intensive sector equals $p_L Y_L = \frac{w}{\varepsilon} L_Y = wL_Y + \Pi_L N_L$.

In equilibrium, factor shares then amount to

$$\pi = \pi_0 \left(\frac{KY_0}{YK_0}\right)^\xi \left(\frac{N_K}{N_{K0}}\right)^{\xi\left(\frac{1-\varepsilon}{\varepsilon}\right)}, \tag{16}$$

$$1 - \pi = (1 - \pi_0) \left(\frac{Y_0 L_Y}{Y L_{Y0}}\right)^\xi \left(\frac{N_L}{N_{L0}}\right)^{\xi\left(\frac{1-\varepsilon}{\varepsilon}\right)}. \tag{17}$$

Incorporating all these choices into (9), and using the definitions $\lambda_b = N_K^{\frac{1-\varepsilon}{\varepsilon}}$ and $\lambda_a = N_L^{\frac{1-\varepsilon}{\varepsilon}}$ retrieves production function (3).

2.2.5 Capital- and Labor-Augmenting R&D Firms

The role of capital- and labor-augmenting R&D firms is to produce innovations which increase the variety of available differentiated intermediate goods, either N_K or N_L , and thus indirectly also

λ_b and λ_a . Patents never expire, and patent protection is perfect. R&D firms sell these patents to the representative household, which sets up a monopoly for each new variety. Patent price, p_b or p_a , which reflects the discounted stream of future monopoly profits, is set at the competitive market. There is free entry to R&D.

R&D firms employ labor only: $L_a = \ell_a L$ and $L_b = \ell_b L$ workers are employed in the labor- and capital-augmenting R&D sector, respectively. There is also an externality from the physical capital stock per worker, working through the capital spillover term in the R&D production function. Furthermore, the R&D firms perceive their production technology as linear in labor, while in fact it is concave due to duplication externalities (Jones 1995).

Incorporating these assumptions and using the notion $x \equiv \lambda_b k / \lambda_a$, capital-augmenting R&D firms maximize

$$\max_{\ell_b} (p_b \dot{\lambda}_b - w \ell_b) = \max_{\ell_b} ((p_b Q_K - w) \ell_b), \tag{18}$$

where $Q_K = B (\lambda_b^{1-\omega} x^{\eta_b} \ell_b^{\nu_b-1})$ is treated by firms as an exogenously given constant (Romer 1990; Jones 2005). Analogously, labor-augmenting R&D firms maximize

$$\max_{\ell_a} (p_a \dot{\lambda}_a - w \ell_a) = \max_{\ell_a} ((p_a Q_L - w) \ell_a), \tag{19}$$

where $Q_L = A (\lambda_a \lambda_b^\phi x^{\eta_a} \ell_a^{\nu_a-1})$ is treated as exogenous.

Free entry into both R&D sectors implies $w = p_b Q_K = p_a Q_L$. Purchase of a patent entitles the holders to a per capita stream of profits equal to π_K and π_L , respectively. While the production of any labor-augmenting varieties lasts forever, there is a constant rate d at which production of capital-intensive varieties becomes *obsolete*. This effect is external to patent holders and thus is not strategically taken into account when accumulating the patent stock.¹⁶

2.2.6 *Equilibrium*

We define the *decentralized equilibrium* as the collection of time paths of all the respective quantities: $c, \ell_a, \ell_b, k, \lambda_b, \lambda_a, Y_K, Y_L$,

¹⁶In other words, by solving a static optimization problem, capital-augmenting R&D firms do not take the dynamic (external) obsolescence effect into account.

$\{X_{Ki}\}, \{X_{Li}\}$ and prices $r, w, p_K, p_L, \{p_{Ki}\}, \{p_{Li}\}, p_a, p_b$ such that (i) households maximize discounted utility subject to their budget constraint; (ii) profit maximization is followed by final goods producers, aggregators and producers of capital- and labor-intensive intermediate goods, and capital- and labor-augmenting R&D firms; (iii) the labor market clears: $L_a + L_b + L_Y = (\ell_a + \ell_b + \ell_Y)L = L$; (iv) the asset market clears: $V = vL = K + p_a\lambda_a + p_b\lambda_b$, where assets have equal returns: $r - \delta = \frac{\pi_L}{p_a} + \frac{\dot{p}_a}{p_a} = \frac{\pi_K}{p_b} + \frac{\dot{p}_b}{p_b} - d$; and, finally, (v), such that the aggregate capital stock satisfies $\dot{K} = Y - C - \delta K - \zeta \dot{L}$.

2.3 Solving for the Social Planner Allocation

In this section, we first solve analytically for the BGP of the social planner (SP) allocation of our endogenous growth model and then linearize the implied dynamical system around the BGP.

2.3.1 Balanced Growth Path

Any neoclassical growth model can exhibit balanced growth only if technical change is purely labor augmenting or if production is Cobb-Douglas; Uzawa (1961). That condition holds here too. Hence, once we presume a CES production function, the analysis of dynamic consequences of any technical change which is not purely labor augmenting must be done outside the BGP.

Along the BGP, we obtain the following growth rate of key model variables:

$$g = \hat{\lambda}_a = \hat{k} = \hat{c} = \hat{y} = A(\lambda_b^*)^\phi (x^*)^{\eta_a} (\ell_a^*)^{\nu_a}, \tag{20}$$

where stars denote steady-state values. Hence, ultimately long-run growth is driven by labor-augmenting R&D. This can be explained by the fact that labor is the only non-accumulable factor in the model, it is complementary to capital along the aggregate production function, and the labor-augmenting R&D equation is linear with respect to λ_a . The following variables are constant along the BGP: $y/k, c/k, \ell_a, \ell_b$, and λ_b (i.e., asymptotically there is no capital-augmenting technical change).

2.3.2 Euler Equations

Having set up the Hamiltonian (with co-state variables μ_k, μ_a, μ_b),

$$\begin{aligned} \mathcal{H}(c, \ell_a, \ell_b, k, \lambda_a, \lambda_b; \mu_k, \mu_a, \mu_b) = & \frac{c^{1-\gamma} - 1}{1 - \gamma} e^{-(\rho-n)t} \\ & + \mu_k(y - c - (\delta + n)k - \zeta \dot{a}) \\ & + \mu_a A \left(\lambda_a \lambda_b^\phi x^{\eta_a} \ell_a^{\nu_a} \right) \\ & + \mu_b \left(B \left(\lambda_b^{1-\omega} x^{\eta_b} \ell_b^{\nu_b} \right) - d\lambda_b \right), \end{aligned} \tag{21}$$

where

$$y = y_0 \left(\pi_0 \left(\lambda_b \frac{k}{k_0} \right)^\xi + (1 - \pi_0) \left(\frac{\lambda_a}{\lambda_{a0}} \frac{1 - \ell_a - \ell_b}{\ell_{Y0}} \right)^\xi \right)^{1/\xi}, \tag{22}$$

computed its derivatives, and eliminated the co-state variables, after tedious algebra the following Euler equations are obtained for the SP:¹⁷

$$\hat{c} = \frac{1}{\gamma} \left(\frac{y}{k} \left(\pi + \frac{1 - \pi}{\ell_Y} \left(\frac{\eta_a \ell_a}{\nu_a} + \frac{\eta_b \ell_b}{\nu_b} \right) \right) - \delta - \rho \right), \tag{23}$$

$$\varphi_1 \hat{\ell}_a + \varphi_2 \hat{\ell}_b = Q_1, \tag{24}$$

$$\varphi_3 \hat{\ell}_a + \varphi_4 \hat{\ell}_b = Q_2, \tag{25}$$

where

$$\varphi_1 = \nu_a - 1 - (1 - \xi) \pi \frac{\ell_a}{\ell_Y}, \tag{26}$$

$$\varphi_2 = -(1 - \xi) \pi \frac{\ell_b}{\ell_Y}, \tag{27}$$

¹⁷A sufficient condition for all transversality conditions to be satisfied in the social optimum (as well as in the decentralized equilibrium) is that $(1 - \gamma)g + n < \rho$.

$$\varphi_3 = -(1 - \xi)\pi \frac{\ell_a}{\ell_Y}, \quad (28)$$

$$\varphi_4 = \nu_b - 1 - (1 - \xi)\pi \frac{\ell_b}{\ell_Y}, \quad (29)$$

and

$$Q_1 = -\gamma\hat{c} - \rho + n + \hat{\lambda}_a \left(\frac{\ell_Y \nu_a}{\ell_a} + 1 - \eta_a - \eta_b \frac{\ell_b \nu_a}{\ell_a \nu_b} \right) - \phi \hat{\lambda}_b + ((1 - \xi)\pi - \eta_a)\hat{x}, \quad (30)$$

$$Q_2 = -\gamma\hat{c} - \rho + n + \hat{\lambda}_a + \hat{\lambda}_b \left(\frac{\pi}{1 - \pi} \frac{\ell_Y \nu_b}{\ell_b} + (\phi + \eta_a) \frac{\nu_b \ell_a}{\nu_a \ell_b} + \eta_b \right) + ((1 - \xi)\pi - \eta_b)\hat{x} + d \left(\frac{\pi}{1 - \pi} \frac{\ell_Y \nu_b}{\ell_b} + (\phi + \eta_a) \frac{\nu_b \ell_a}{\nu_a \ell_b} - \omega + \eta_b \right). \quad (31)$$

2.3.3 Steady State and Linearization of the Transformed System

The above Euler equations and dynamics of state variables are then rewritten in terms of *stationary* variables which are constant along the BGP, i.e., in coordinates: $u = (c/k)$, ℓ_a , ℓ_b , x , λ_b , and with auxiliary variables $z = (y/k)$, π , g . The full steady state of the transformed system is listed in appendix A.1. This nonlinear system of equations is solved numerically, yielding a steady state of the detrended system and thus a BGP of the model in original variables.¹⁸ All further analysis of the social planner allocation is based on the (numerical) linearization of the five-dimensional dynamical system of equations (23)–(25), (2), and (5), taking the BGP equality (20) as given.

2.4 Solving for the Decentralized Allocation

When solving for the decentralized allocation (DA), we broadly follow the steps carried out in the case of the social planner (SP)

¹⁸We do not have a formal proof of BGP uniqueness, but the large number of numerical checks we have performed (e.g., varying initial conditions of the numerical algorithm, modifying values of model parameters), is suggestive that the BGP is indeed unique and depends smoothly on model parameters.

allocation. We first solve analytically for the BGP of our endogenous growth model and then linearize the implied dynamical system around the BGP.

2.4.1 *Balanced Growth Path*

Along the BGP, we obtain the following growth rate of the key model variables:

$$g = \hat{k} = \hat{c} = \hat{y} = \hat{w} = \hat{p}_b = \hat{p}_{Li} = \hat{\lambda}_a = A(\lambda_b^*)^\phi (x^*)^{\eta_a} (\ell_a^*)^{\nu_a}. \quad (32)$$

The following quantities are constant along the BGP: $y/k, c/k, \ell_a, \ell_b, Y_K/Y, Y_L/Y$, and λ_b (again, note, asymptotically, the absence of capital-augmenting technical change). The following prices are also constant along the BGP: $r, p_a, p_K, p_L, \{p_{K_i}\}$.

2.4.2 *Euler Equations*

The decentralized equilibrium is associated with the following Euler equations describing the first-order conditions:

$$\hat{c} = \frac{\varepsilon\pi \frac{y}{k} - \delta - \rho}{\gamma}, \quad (23')$$

$$\varphi_1 \hat{\ell}_a + \varphi_2 \hat{\ell}_b = \tilde{Q}_1, \quad (24')$$

$$\varphi_3 \hat{\ell}_a + \varphi_4 \hat{\ell}_b = \tilde{Q}_2, \quad (25')$$

where

$$\tilde{Q}_1 = -\varepsilon\pi \frac{y}{k} + \delta + \hat{\lambda}_a \frac{\ell_Y}{\ell_a} - \phi \hat{\lambda}_b + ((1 - \xi)\pi - \eta_a)\hat{x} \quad (30')$$

$$\begin{aligned} \tilde{Q}_2 = & -\varepsilon\pi \frac{y}{k} + \delta + \hat{\lambda}_a + (\hat{\lambda}_b + d) \left(\frac{\pi}{1 - \pi} \frac{\ell_Y}{\ell_b} \right) \\ & - \hat{\lambda}_b(1 - \omega) - d + ((1 - \xi)\pi - \eta_b)\hat{x} \end{aligned} \quad (31')$$

and φ_1 through φ_4 are defined as in (26)–(29). The full steady state of the transformed system is listed in appendix A.2. All further analysis of the decentralized allocation is based on the (numerical)

linearization of the five-dimensional dynamical system of equations (23')–(25'), (2), and (5), taking the BGP equality (32) as given.

2.4.3 *Departures from the Social Optimum*

Departures of the decentralized allocation from the optimal one can be tracked back to specific assumptions regarding the information structure of the decentralized allocation. In the following we try to compare one-to-one all the differences in the equations related to the decentralized and to social planner outcome. As we shall show, those differences follow from the presence of wedges (such as markups from imperfect competition) and externalities (such as duplication externalities from R&D investment). These differences also show in different costs and returns that exist in the social planner and decentralized allocation.

Specifically, the points of comparison are as follows:

1. In the consumption Euler equation, comparing equations (23) with (23'), the term $\frac{y}{k} \left(\pi + \frac{1-\pi}{\ell_Y} \left(\frac{\eta_a \ell_a}{\nu_a} + \frac{\eta_b \ell_b}{\nu_b} \right) \right)$ is replaced by $\varepsilon \pi \frac{y}{k}$. This is due to two effects:
 - (a) in contrast to the social planner, markets fail to account for the external effects of physical capital on R&D activity via the capital spillover terms (with respective elasticities η_b and η_a);
 - (b) ε appears in the decentralized allocation due to imperfect competition in the labor- and capital-augmenting intermediate goods sectors.

Both effects work in the same direction and buy the social planner much more capital in the steady state. The savings rate is much higher in the SP allocation, as two effects add up: (i) accounting for social instead of private returns on production capital, (ii) internalizing the positive returns to capital used in R&D. Therefore in the social planner's steady state, there is relatively lower consumption and output per unit of capital, and greater positive capital spillover in R&D, $x = \frac{\lambda_b k}{\lambda_a}$.

This abundance of capital makes labor relatively scarce, which—if and only if both factors are gross complements

($\sigma < 1$)—increases the labor share in the social planner allocation relative to the decentralized equilibrium.

2. In the Euler equation for ℓ_a , the term $\left(\frac{\ell_Y \nu_a}{\ell_a} + 1 - \eta_a - \eta_b \frac{\ell_b \nu_a}{\ell_a \nu_b}\right)$, is replaced by $\frac{\ell_Y}{\ell_a}$. Analogously, in the Euler equation for ℓ_b , the term given by $\frac{\ell_Y \nu_b}{\ell_b} \frac{\pi}{1-\pi} + (1 - \omega) + \eta_b + (\phi + \eta_a) \frac{\ell_a \nu_b}{\ell_b \nu_a}$ is replaced by $\frac{\ell_Y}{\ell_b} \frac{\pi}{1-\pi}$. This is due to two effects:

- (a) ν_a and ν_b are missing in the respective first components because markets fail to internalize the detrimental R&D duplication effects;
- b) the latter two components are missing because markets fail to account for the positive external effects of accumulating knowledge on future R&D productivity. These effects are included in the shadow prices of λ_a and λ_b in the social planner allocation but not in their respective market prices.

The effect (a) reduces SP's investment in R&D, whereas the effect (b) increases it. In our baseline calibration and its robustness checks, on balance the latter effect robustly prevails and the social planner allocates less labor to production and more to R&D (greater ℓ_a and ℓ_b). Hence, this mechanism causes the social planner to accumulate more investment in both R&D sectors. This, given that labor-augmenting technological progress is the ultimate source of growth in the model, increases the long-run growth rate.

3. In the Euler equations for ℓ_a and ℓ_b (equations (24), (25), (24'), and (25')) the shadow price of physical capital $\hat{c} - \rho + n$ is replaced by its market price $r - \delta = \varepsilon \pi \frac{y}{k} - \delta$, which is lower because it accounts for markups arising from imperfect competition.

This mechanism causes the social planner to accumulate more capital-augmenting R&D ℓ_b relative to labor-augmenting R&D ℓ_a . Therefore the ratio of capital- to labor-augmenting R&D employment is always higher in the social optimum than in the decentralized allocation. Hence, unit capital productivity (λ_b) is higher in the SP steady state. This further adds to the capital spillover term x which, in

turn, augments both R&D sectors, accelerates growth and—by making labor relatively scarce—increases the labor share.

3. Calibration of the Model

3.1 Empirical Calibration Components

The parameter calibration for the decentralized model based on magnitudes from historical data or empirical studies is listed in table 1. We assume that a range of long-run averages from U.S. data (1929–2015) correspond to the decentralized BGP of the model. Doing so allows us to calibrate the rates of economic and population growth, capital productivity and income share, and the consumption-to-capital ratio. Likewise, we assign CES normalization parameters to match U.S. long-run averages for factor income shares (we adjust the payroll share by proprietors' income, as in Mućk, McAdam, and Growiec 2018). This implies an average labor share of 0.67.¹⁹

Next, we turn to the elasticity of substitution between labor and capital (σ) which is the fundamental economic parameter in our analysis. We calibrate factors to be gross complements, i.e., $\sigma < 1$. This choice stems from a fact that the bulk of empirical studies for the U.S. aggregated production function document that the σ is systematically below unity (Klump, McAdam, and Willman 2012).²⁰ Most of the empirical evidence exploiting time-series variation for other countries also implies $\sigma < 1$ (McAdam and Willman 2013; Mućk 2017; Knoblach and Stöckl 2020).

However, the literature based predominantly on *cross-country* variation is rather inconclusive about the magnitude of σ . On one hand, several papers (Piketty and Zucman 2014; Karabarbounis and Neiman 2014) employ gross substitutes; however, the former paper

¹⁹Note that in the model, due to the inelastic labor supply and the firm profits being rebated to the household, markups do not directly affect factor shares.

²⁰For instance, Arrow et al. (1961) found an aggregate elasticity over 1909–49 of 0.57 (similar to that of the more recent Antràs 2004). More recently, Klump, McAdam, and Willman (2007) reported $\hat{\sigma} \approx 0.7$. The tendency towards gross complementarity between factors is also confirmed at the industry level (Herrendorf, Herrington, and Valentinyi 2015; Laeven, McAdam, and Popov 2018) and firm level (Oberfield and Raval 2018). Importantly, the elasticity uncovered is found systematically below unity even if more flexible functional forms of aggregate production function are considered (Growiec and Mućk 2020).

Table 1. Calibrated Parameters (decentralized allocation)

Parameter	Symbol	Value	Target/Source
<i>Income and Production</i>			
GDP per capita Growth	g	0.0171	Geometric Average
Population Growth Rate	n	0.0153	Geometric Average
Capital Productivity	z_0, z^*	0.3442	Geometric Average
Consumption-to-Capital	u^*	0.2199	Geometric Average
Capital Income Share	π_0, π^*	0.3260	Arithmetic Average
Depreciation	δ	0.0600	Caselli (2005)
Factor Substitution Parameter	ξ	-0.4286	$\Rightarrow \sigma = 0.7$, Klump, McAdam, and Willman (2007)
<i>Preferences</i>			
Inverse Intertemporal Elasticity of Substitution	γ	1.7500	Barro and Sala-i-Martin (2003)
Time Preference	ρ	0.0200	Barro and Sala-i-Martin (2003)
<p>Notes: This table shows the parameter values that are used in the central calibration of the model. The values are taken from historical averages (1929-2015) of the relevant U.S. time series.</p>			

calibrates the σ value and the latter estimates it in a cross-country panel context.²¹ On the other hand, recent studies exploiting macro panels and allowing for factor augmentation in the supply-side system approach strongly conclude in favor of gross complementarity in production (Mućk 2017). Given this, we consider $\sigma < 1$ as the benchmark, but we do examine the $\sigma > 1$ case in our robustness exercises.

Finally, the table includes preference parameters (intertemporal elasticity of substitution, time preference) which are difficult to retrieve from historical data. For that reason, in our central calibration we rely on values typically found in the literature.

3.2 Model-Consistent Calibration Components

Next, conditional on the values in table 1, four identities included in the system (see appendix equations (A.9)–(A.17)) drive the calibration of other parameters in a model-consistent manner: r^* , λ_b^* , x^* , and ε . Employment in final production ℓ_Y^* is also set in a model-consistent manner (table 2).

In the absence of any other information, we agnostically assume that the share of population $1 - \ell_Y^*$ is split equally between employment in both (i.e., capital and labor) R&D sectors in the decentralized allocation—although notice these employment shares are endogenous in the social planner solution. For the model-consistent value of ℓ_Y^* , the relevant formula leads to values close to those typically considered for the non-routine cognitive occupational group (e.g., Jaimovich and Siu 2020, using Bureau of Labor Statistics data, show this ratio to be between 29 percent and 38 percent, over 1982–2012).

For the duplication externalities, we assume $\nu_a = \nu_b = 0.75$ following the (albeit single R&D sector) value in Jones and Williams (2000) (although, note again, we conduct extensive robustness checks on these values). The steady-state level of unit capital productivity λ_b^* is normalized to unity, and so are CES normalization parameters λ_{a0} and λ_{b0} .

²¹The latter paper moreover was estimated on a single-equation non-normalized basis which is known to have poor estimation properties in this context (León-Ledesma, McAdam, and Willman 2010; Klump, McAdam, and Willman 2012).

Table 2. Parameter and Steady-State Variable Calibration Conditional on Historical Average Calibration (decentralized allocation)

Parameter	Symbol	Value	Target/Source
<i>Income and Production</i>			
Net Real Rate of Return	$r^* - \delta$	0.0499	$r^* - \delta = \gamma g + \rho$
Substitutability between Intermediate Goods	ε	0.9793	$\varepsilon = \frac{\pi^* z^*}{\pi^* z^*}$
<i>R&D Sectors</i>			
R&D Duplication Parameters	$v_a = v_b$	0.7500	Jones and Williams (2000) Normalized to Unity
Technology Augmenting Terms	$\lambda_{a0}, \lambda_{b0}$	1	
Unit Capital Productivity	λ_b^*	1	$\lambda_b^* = \lambda_{b0} \frac{z^*}{z_0} \left(\frac{\pi^*}{\pi_0} \right)^{\frac{1}{\varepsilon}}$
Employment Share in R&D Sectors	ℓ_a^*, ℓ_b^*	0.2033	$\ell_a^* = \ell_b^*$ for $\ell_a^* + \ell_b^* = 1 - \ell_Y^*$
Capital-Labor Ratio in Efficiency Units [†]	x_0, x^*	61.7900	$x^* = x_0 \frac{\ell_a^*}{\ell_Y^*} \left(\frac{1}{1-\pi_0} \left(\frac{z^* \lambda_{b0}}{z_0 \lambda_b^*} \right) \frac{\xi}{\xi} - \frac{\pi_0}{1-\pi_0} \right)^{-1/\xi}$
<p>Notes: This table shows the parameter and steady-state variable values that are used in the central calibration of the model. The values are taken from values consistent with the structure of the model or from other relevant studies in the literature.</p> <p>[†]$x_0 = \frac{\lambda_{b0} k_0}{\lambda_{a0}} = 61.79$.</p>			

The final step is to assign values to the remaining parameters, in particular the technological parameters of the R&D equations. We do this by solving the four remaining equations in system (A.9)–(A.17) with respect to the remaining parameters; see table B.1. Given this benchmark calibration, the steady state is a saddle point.

4. Is the Decentralized Labor Share Socially Optimal?

Given the model setup and its benchmark calibration, we can now come to our central question: is the decentralized labor share socially optimal? In table 3, columns 1 and 2 show the decentralized allocation (DA) and social planner (SP) outcomes for our benchmark calibration; columns 3 and 4, considered later, alternatively impose Cobb-Douglas and gross substitutes.²²

The BGP of the DA solution features less physical capital, lower growth, and lower R&D activity, but a higher consumption rate (u is higher) than the SP. Moreover, with less capital and lower growth, the net real rate of return of capital is higher, and capital productivity is accordingly higher. Under gross complementarity of capital and labor, the relative scarcity of capital implies that also the labor share is lower. The theoretical underpinnings of these discrepancies have been discussed in section 2.4.3. But the magnitude of the labor share difference is perhaps less obvious. In fact, we see the striking result that the labor share in the social optimum is around 17 percent (11 percentage points) above the decentralized allocation. This means that looking at efficiency considerations only, the labor share not just is empirically too low today, but probably was too low even in the 1980s, before it embarked on a secular downward trend.

²²We made a large number of numerical checks for existence, uniqueness, and stability of the steady state (e.g., varying initial conditions of the numerical algorithm, performing an eigenvalue analysis of the detrended system around the steady state, and modifying values of model parameters). Our results confirm that in the baseline calibration as well as across a large parameter space around it, the steady state of the model is unique, saddle-path stable, and depends smoothly on model parameters. Results of this analysis, beyond the ones reported in figures in our appendixes, are available on request.

Table 3. BGP Comparison under the Baseline Calibration

Variable	Symbol	(1)	(2)	(3)	(4)
		DA	Baseline Gross Comp. $\sigma = 0.7$ $\xi = -0.43$	SP	
				CD	Piketty Gross Sub. $\sigma = 1.25$ $\xi = 0.20$
Output Growth Rate	g	0.0171	0.0339	0.0425	0.0581
Consumption-to-Capital Ratio	u^*	0.2199	0.1628	0.1180	0.0856
Capital Productivity	z^*	0.3442	0.3071	0.2832	0.2743
Employment in Production	ℓ_Y^*	0.5934	0.4385	0.4160	0.3854
Employment in Labor-Augmenting R&D	ℓ_a^*	0.2033	0.2575	0.2447	0.2240
Employment in Capital-Augmenting R&D	ℓ_b^*	0.2033	0.3040	0.3393	0.3906
Relative Share	ℓ_a^*/ℓ_b^*	1	0.8470	0.7212	0.5735
Labor Income Share	$1 - \pi^*$	0.6739	0.7854	0.6739	0.5243
Relative to DA (%)	$\frac{1 - \pi^{*DA}}{1 - \pi^*} - 1$	0	0.1655	0	-0.2220
Capital Income Share	π^*	0.3261	0.2146	0.3261	0.4757
Net Real Rate of Return	$r^* - \delta$	0.0499	0.0059	0.0323	0.0704
Capital-Augmenting Technology	λ_b^*	1.0000	2.3696	3.3162	5.2600
Capital-Labor Ratio in Efficiency Units	x^*	61.7900	173.3363	342.7082	928.9625

Notes: This table shows the outcomes for the endogenous variables for the decentralized and social planner outcomes for different values of the aggregate elasticity of substitution.

To further understand why this discrepancy is so high, let us decompose the capital income share, π , in the following two ways (recalling that $1 - \pi$ is the labor income share):

$$\frac{\pi}{\pi_0} = \left(\frac{\lambda_b k}{k_0} \right)^\xi \left(\frac{y}{y_0} \right)^{-\xi} \Rightarrow \hat{\pi} = \xi(\hat{\lambda}_b + \hat{k} - \hat{y}), \quad (33)$$

$$\frac{\pi}{1 - \pi} = \frac{\pi_0}{1 - \pi_0} \left(\frac{x \ell_{Y0}}{x_0 \ell_Y} \right)^\xi \Rightarrow \hat{\pi} = \xi(1 - \pi)(\hat{x} - \hat{\ell}_Y). \quad (34)$$

Equation (33) shows that under gross complementarity ($\sigma < 1$, or equivalently $\xi < 0$), the capital share increases with capital productivity and decreases with capital augmentation (i.e., the capital-augmenting technology improvements are “labor biased”).

Equation (34), in turn, follows from the definition of the aggregate production function and the effective capital–labor ratio x . Given $\hat{\ell}_Y \equiv - \left(\frac{\ell_a}{\ell_Y} \hat{\ell}_a + \frac{\ell_b}{\ell_Y} \hat{\ell}_b \right)$, the dynamics of employment in the goods sector are equal to the inverse of the dynamics of total R&D employment. It then follows that dynamics of the labor share are *uniquely determined* by the sum of the dynamics of the capital spillover term x and R&D employment. As before, the sign of this relationship depends upon the substitution elasticity: if $\xi < 0$, then increases in R&D intensity reduce π , and thus increase the labor share, and vice versa.

Comparing the decentralized and the social planner’s allocation through the lens of (33), we observe that the large difference in factor shares at the BGP is driven almost exclusively by the difference in the level of capital augmentation λ_b^* . This result suggests that technical change is quantitatively more important for explaining labor share developments than shifts in the capital-output ratio.

Equivalently, by (34), this large difference in the degree of capital augmentation shows up in the capital spillover term x^* . It is also strengthened by the discrepancy in employment in final production ℓ_Y^* , which is higher in the decentralized allocation because the planner devotes more resources to (both types of) R&D. Thanks to this, coupled with relatively more saving, the social planner achieves faster growth at the BGP but with a lower consumption-to-capital

ratio and a lower rate of return to capital. All of these make for a higher labor share in the optimal allocation.

4.1 *Impact of Parameter Variation on the Equilibrium Labor Share*

The results just discussed hold for the benchmark calibration. Accordingly, we now consider sensitivity to deviations from that calibration. Figure 2 presents the impact of varying selected model parameters, holding others constant, on the BGP level of the labor share.

Essentially, all panels can be interpreted through the lens of equations (33) and (34). As agents become less patient (higher ρ), R&D intensity falls, as does the labor share. Similar reasoning pertains to the inverse intertemporal elasticity of substitution γ . That $\frac{\partial(1-\pi)}{\partial\eta_b} > 0$ arises from the usual property that, under our gross complements benchmark, improvements in capital-augmenting technical change are labor biased; analogously, $\frac{\partial(1-\pi)}{\partial\eta_a} < 0$. Likewise, we have under gross complements $\frac{\partial(1-\pi)}{\partial\nu_a} > 0$, $\frac{\partial(1-\pi)}{\partial\nu_b} < 0$. If capital depreciates faster, the capital (labor) share rises (falls).

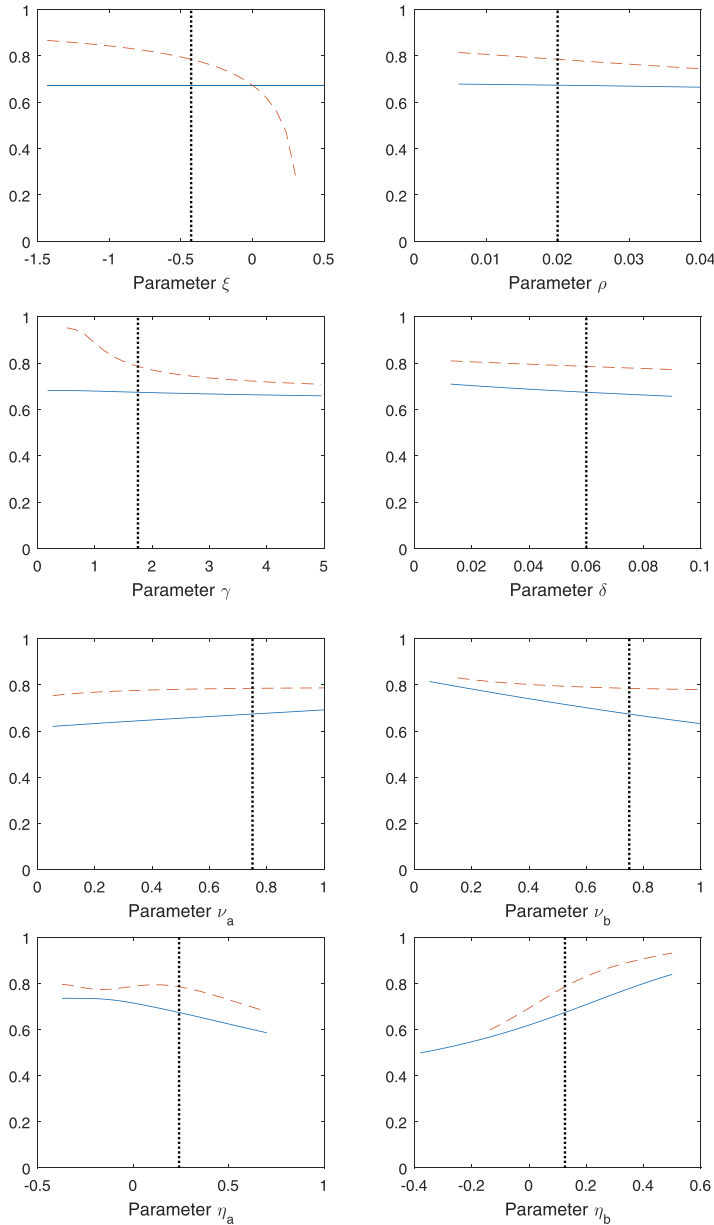
Finally, we see that under gross substitutes, $\sigma > 1$ ($\xi > 0$), the DA labor share *exceeds* that of the SP. We discuss this case further below, but it is straightforward to motivate, since the previously discussed mechanisms go into reverse; capital-augmenting technical improvements tend to be capital biased, as output is now directed towards the relatively abundant, not the scarce factor of production.²³

A more extensive study of the dependence of both BGPs on key model parameters ($\rho, \gamma, \nu_b, \eta_b$) is included in figures D.1–D.3; the equivalent figure for the gross substitutes case is given in figure D.4. They are essentially a mirror image of our benchmark gross complements case.

Moreover, appendix C shows the impact of parameter variations on the equilibrium growth rate.

²³Note that the lack of dependence of the BGP on ξ in the decentralized allocation follows from CES normalization (Klump and de La Grandville 2000), coupled with the fact that we have calibrated the normalization constants to the BGP of the decentralized allocation.

Figure 2. Dependence of Equilibrium Labor Share on Model Parameters



Notes: $1 - \pi$ on vertical axis; corresponding parameter support on the horizontal axis. Social planner allocation (dashed lines), decentralized equilibrium (solid lines). The vertical dotted line in each graph represents the baseline calibrated parameter value.

4.2 *Impact of Elasticity of Substitution Variation on the BGP*

Although we regard the gross complements case to be the more empirically relevant (at least for the aggregate economy), we also investigate the Cobb-Douglas and gross substitutes case. Accordingly, the SP is solved anew and presented in columns 3 and 4 of table 3, respectively.²⁴

Both alternative parameterizations are markedly more growth friendly. Per capita output grows at the counterfactual rate of around 4–6 percent, exceeding both the previous SP and DA by a large margin, with an inflection point at $\xi \approx 0.25$ ($\sigma \approx 1.33$), after which it shoots through the roof. The fact that steady-state per capita growth is an increasing function of the substitution elasticity, though, is to be expected. Intuitively, easier factor substitution—by staving off diminishing returns—can prolong extensive growth (i.e., scarce factors can be substituted by abundant ones). The formal proof of this can be related through the properties of the normalized CES function as a general mean function.²⁵

The consequences for labor's share of income, though, are dire. With gross factor substitutability,²⁶ the arguments of the previous section shift into reverse. Capital improvements are capital biased, and the incentives for capital accumulation are accordingly far higher in this regime. Hence the labor share declines with σ (or equivalently ξ).

It should also be emphasized that a balanced growth path does not exist in our model under sufficiently strong factor substitutability. Gross substitutability, as such, implies that Inada conditions at infinity are violated: the marginal product of per capita capital (MPK) remains bounded *above* zero as the capital stock goes to infinity. But then there is still the question whether the lower bound of MPK, multiplied by the savings rate, is high enough to exceed the capital depreciation rate. If so, and this happens only when

²⁴The effect of a continuous variation in the substitution elasticity is graphed in figure D.5.

²⁵See the discussion in Pitchford (1960) and the subsequent discussions in de La Grandville (1989); Klump and de La Grandville (2000); Klump and Preissler (2000), and Palivos and Karagiannis (2010).

²⁶In the Cobb-Douglas case of $\xi = 0$, factor shares are constant and at their predetermined sample average. Thus $\pi|_{\xi=0} = \pi_0$.

σ exceeds a certain threshold $\bar{\sigma} > 1$, endogenous growth driven by capital accumulation appears (Jones and Manuelli 1990; Palivos and Karagiannis 2010). Combined with the existing growth engine of our model—labor-augmenting R&D—both sources of growth then lead to super-exponential, explosive growth. Then, even with diminishing returns to factors, capital intensity grows without bounds, labor becomes inessential in production, and hence the capital income share tends to unity. We rule such cases out of our analysis.

5. Additional Results

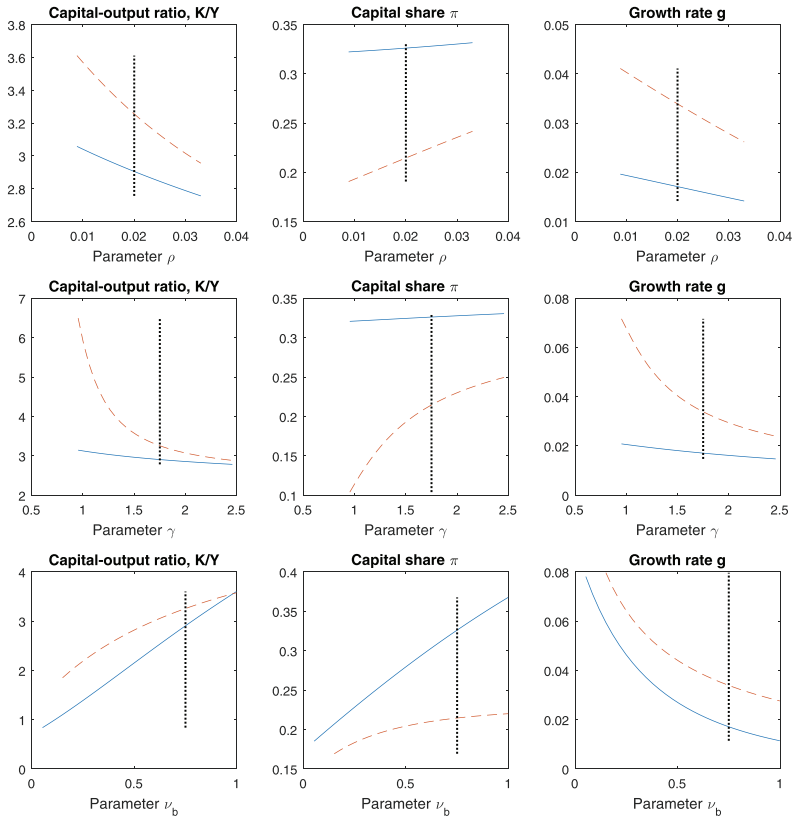
5.1 *Comparing the Model with Piketty's Laws*

As our model endogenizes both economic growth and factor shares, it constitutes an appropriate framework for studying the two “Fundamental Laws of Capitalism” formulated by Piketty (2014), i.e., (i) that the capital–output ratio K/Y rises whenever the economic growth rate g falls, and (ii) that the capital share π rises whenever the growth rate g falls. Our setup has the advantage over Piketty’s that all three variables are endogenous, and hence one can legitimately observe whether changing some parameters implies co-movements that are or are not in line with Piketty’s claims (i) and (ii). In addressing Piketty’s laws with an R&D-based endogenous growth model, we follow the footsteps of Irmen and Tabakovic (2020). In contrast to their contribution, though, our setup departs from Cobb–Douglas technology.²⁷

First, taking Piketty’s claims (i) and (ii) together logically implies that K/Y and the capital share π are positively correlated, suggesting that capital and labor should be gross substitutes ($\sigma > 1$); see equation (33). This is a widely recognized issue with

²⁷In Irmen and Tabakovic (2020), due to Cobb–Douglas technology, factor shares of capital, labor, and ideas in final output are always constant (their proposition 1). Factor shares in GDP, however, may vary because—foremost—GDP includes also new patented technological knowledge, and the proportion of final output to new technological knowledge within GDP is endogenous. By contrast, in our framework already factor shares in final output are variable. Therefore our setup is arguably better suited to identifying first-order effects of technical change on factor shares.

Figure 3. Dependence of BGP on $\rho, \gamma,$ and ν_b : DA vs. SP



Notes: Social planner allocation (dashed lines), decentralized equilibrium (solid lines). The vertical dotted line in each graph represents the baseline calibrated parameter value.

Piketty’s claims (see, e.g., Oberfield and Raval 2018). In our baseline parameterization, we assume gross complements instead.

Second, inspection of figure 3 reveals that under the baseline calibration, both in the decentralized equilibrium and the social planner allocation:

- when households become more patient (ρ goes down) or more willing to substitute consumption intertemporally (γ goes down), only law (ii) holds: the growth rate g goes up, the K/Y ratio goes up, and the capital share π goes down;

- when the capital spillover exponent ν_b in capital-augmenting R&D goes up, both laws are verified: the growth rate g goes down, the K/Y ratio goes up, and the capital share π goes up.

Third, we find (figure D.5) that as the elasticity of substitution goes up, the optimal growth rate g goes up hand in hand with the capital share π and the K/Y ratio. In such case, *both* of Piketty's laws are violated.

5.2 *Is the Decentralized Economy Characterized by Excessive Volatility?*

In the data, we know that—irrespective of the concept utilized—labor shares are highly persistent and variable.²⁸ Although bounded within the unit interval and theoretically stationary, in the data labor income shares often appear to be characterized by marked volatility and long swings. In particular, around 80 percent of total labor share volatility in the United States (1929–2015) has been due to fluctuations in medium- to long-run frequencies (beyond the eight-year mark). As opposed to the short-run component of the labor share, its medium- to long-run component has also been procyclical (Growiec, McAdam, and Mućk 2018).

Other than undermining the case for aggregate Cobb-Douglas production, this also raises the question of whether our framework can generate and rationalize these long cycles. Growiec, McAdam, and Mućk (2018) have confirmed this conjecture for the decentralized allocation of the current model. The question is however equally interesting for the social planner case. Are cycles in factor income shares socially optimal? If so, (stabilization) policies to mitigate labor share or real volatility might be appraised differently.²⁹

Table 4 makes the relevant comparisons across our maintained cases. It shows that the decentralized allocation features relatively shorter cycles but also faster convergence to the BGP. Hence, it

²⁸For international evidence, see Jalava et al. (2006); Bengtsson (2014); and Mućk, McAdam, and Growiec (2018).

²⁹By design, our analysis focuses only on endogenous long swings in factor shares. The deterministic character of the model precludes any conclusions regarding the magnitude and persistence of short-run fluctuations.

Table 4. Dynamics around the BGP

Allocation	Baseline		CD		Piketty	
	DA	SP	DA	SP	DA	SP
Pace of Convergence* (% per year)	6.3%	4.2%	5.8%	3.7%	5.2%	2.9%
Length of Full Cycle† (Years)	52.6	76.7	79.8	83.2	144.0	100.3
Labor Share Cyclicity	+	+	0	0	-	-
Amplitude of $1 - \pi$ Relative to y/k	62.0%	48.0%	NA	NA	28.0%	44.0%

Notes: *Computed as $1 - e^{rr}$ where $rr < 0$ is the real part of the largest stable root; †Computed as $2\pi/ir$ where $ir > 0$ is the imaginary part of two conjugate stable roots (if they exist). "NA" denotes not available/applicable. See table 3 for the definitions of "Baseline," "CD," and "Piketty."

cannot be claimed directly that the decentralized equilibrium has *excessive volatility* of the labor share. If both allocations were to start from the same initial point outside of the BGP, then the decentralized allocation would exhibit a greater frequency but smaller amplitude of cyclical variation.

Having scrutinized the robustness of this dynamic result by extensively altering the parameterization of the model, we conclude that while the decentralized equilibrium generally exhibits shorter cycles, the ordering of both allocations in terms of the pace of convergence can sometimes be reversed. This finding lends partial support to the claim that the decentralized equilibrium is perhaps likely to feature greater labor share volatility compared with the social optimum. However, it is worthwhile to point out that oscillations in the labor income share can still be socially optimal in this model.

Moreover, we also obtain quantitative predictions on the cyclical co-movement of the original model variables (including the economic growth rate g and the labor share $1 - \pi$).³⁰ It turns out, both for the decentralized and optimal allocation, that all variables except for the consumption-capital ratio $u = c/k$ oscillate when converging to the steady state, with the same frequency of oscillations. The level of capital-augmenting technology λ_b , the capital spillover term x , and labor-augmenting R&D employment ℓ_a are always countercyclical, employment in production ℓ_Y is always procyclical, whereas the cyclicality of capital-augmenting R&D ℓ_b is ambiguous (in the baseline calibration, ℓ_b is procyclical in the decentralized allocation but countercyclical in the optimal one). Furthermore, as long as capital and labor are gross complements, the labor income share $1 - \pi$ is unambiguously procyclical as well. These features of cyclical co-movement align well with the empirical evidence for the U.S. medium-term cycle. In particular, the U.S. labor share is indeed procyclical over the medium-to-long run—despite its countercyclicity along the business cycle (Growiec, McAdam, and Mućk 2018; Mućk, McAdam, and Growiec 2018).

³⁰This is done by inspecting the eigenvector associated with the largest stable root of the Jacobian of the system at the steady state.

6. Conclusions

Modern endogenous growth theory tends to suggest that the socially optimal level of economic activity dominates (i.e., exceeds) the decentralized outcome. The decentralized outcome produces too little output because of monopoly behavior, markups, and externalities related to reaping the private returns to innovation. In this paper, we have confirmed this conclusion using a microfounded, calibrated two-sector R&D endogenous growth model. Due to externalities between the two R&D sectors, in our model the decentralized allocation produces also a socially suboptimal level of R&D and, particularly, too little capital-augmenting R&D. This, in addition to a suboptimal level of capital accumulation, translates into too low equilibrium growth.

But what of the labor share? Despite its importance, the conclusions for this variable have perhaps surprisingly not yet been drawn in the literature. Our objective was to bridge that knowledge gap. We found that if the elasticity of factor substitution σ is below unity (as the bulk of evidence suggests for the aggregate U.S. economy), then the decentralized labor share is indeed socially suboptimal. The difference, moreover, is large, around 17 percent in our baseline calibration.

Effectively, the only parameter which can reverse this ordering is the elasticity of substitution. However in the gross substitutes case ($\sigma > 1$) it tends to yield counterfactual outcomes. For example, an elasticity of $\sigma = 1.25$, only slightly above Cobb-Douglas, produces a decentralized labor share above the social planner one, but then the latter is as low as 0.52; as a simple point of comparison, according to the International Labor Organization (ILO) definition of the labor share (using annual data from 1960 to the present), no G-7 country has fallen below a labor share of 0.5. Moreover, such a mild perturbation away from Cobb-Douglas already produces equilibrium per capita growth rates of around 6 percent per annum.

In the future, our results should be contrasted with findings from a highly needed prospective study of optimal factor shares under inequality in factor ownership. Such a study could uncover the associated efficiency versus inequality tradeoff. We expect that the discrepancy between the optimal and decentralized labor share would then be even larger than 17 percent because the social planner might

increase the labor share not just to improve efficiency of production under gross complementarity ($\sigma < 1$) but also to reduce income inequality (given that capital incomes tend to be relatively more concentrated).

Appendix A. Steady State of the Transformed System

A.1 Social Planner Allocation

The steady state of the transformed dynamical system implied by the social planner solution satisfies

$$g = \hat{\lambda}_a = \hat{k} = \hat{c} = \hat{y} = A(\lambda_b^*)^\phi (x^*)^{\eta_a} (\ell_a^*)^{\nu_a} \quad (\text{A.1})$$

$$\gamma g + \delta + \rho = z \left(\pi + \frac{1 - \pi}{\ell_Y} \left(\frac{\eta_a \ell_a}{\nu_a} + \frac{\eta_b \ell_b}{\nu_b} \right) \right) \quad (\text{A.2})$$

$$g = z - \zeta \frac{\dot{a}}{k} - u - (\delta + n) \quad (\text{A.3})$$

$$d = B (\lambda_b^{-\omega} x^{\eta_b} \ell_b^{\nu_b}) \quad (\text{A.4})$$

$$(1 - \gamma)g + n - \rho = d \left(\frac{\pi}{1 - \pi} \frac{\ell_Y \nu_b}{\ell_b} + (\phi + \eta_a) \frac{\nu_b \ell_a}{\nu_a \ell_b} - \omega + \eta_b \right) \quad (\text{A.5})$$

$$(1 - \gamma)g + n - \rho = -g \left(\frac{\ell_Y \nu_a}{\ell_a} - \eta_a - \eta_b \frac{\ell_b \nu_a}{\ell_a \nu_b} \right) \quad (\text{A.6})$$

$$\frac{\pi}{\pi_0} = \left(\frac{\lambda_b}{\lambda_{b0}} \right)^\xi \left(\frac{z}{z_0} \right)^{-\xi} \quad (\text{A.7})$$

$$\frac{z}{z_0} = \frac{\lambda_b}{\lambda_{b0}} \left(\pi_0 + (1 - \pi_0) \left(\frac{x_0}{x} \frac{\ell_Y}{\ell_{Y0}} \right)^\xi \right)^{1/\xi} \quad (\text{A.8})$$

This nonlinear system of equations is solved numerically, yielding the steady state of the detrended system, and thus the BGP of the model in original variables. All further analysis of the social planner allocation is based on the (numerical) linearization of the five-dimensional dynamical system of equations (23)–(25), (2), and (5), taking the BGP equality (20) as given.

A.2 The Decentralized Allocation

As in the case of the social planner, the Euler equations and dynamics of state variables are rewritten in terms of *stationary* variables. The steady state of the transformed system satisfies

$$g = \hat{\lambda}_a = \hat{k} = \hat{c} = \hat{y} = A(\lambda_b^*)^\phi (x^*)^{\eta_a} (\ell_a^*)^{\nu_a} \tag{A.9}$$

$$\gamma g + \rho = r - \delta \tag{A.10}$$

$$g = z - \zeta \frac{\dot{a}}{k} - u - (\delta + n) \tag{A.11}$$

$$d = B (\lambda_b^{-\omega} x^{\eta_b} \ell_b^{\nu_b}) \tag{A.12}$$

$$g \frac{\ell_Y}{\ell_a} = r - \delta \tag{A.13}$$

$$g = r - \delta + d \left(1 - \frac{\pi}{1 - \pi} \frac{\ell_Y}{\ell_b} \right) \tag{A.14}$$

$$r = \varepsilon \pi z \tag{A.15}$$

$$\frac{\pi}{\pi_0} = \left(\frac{\lambda_b}{\lambda_{b0}} \right)^\xi \left(\frac{z}{z_0} \right)^{-\xi} \tag{A.16}$$

$$\frac{z}{z_0} = \frac{\lambda_b}{\lambda_{b0}} \left(\pi_0 + (1 - \pi_0) \left(\frac{x_0}{x} \frac{\ell_Y}{\ell_{Y0}} \right)^\xi \right)^{1/\xi} . \tag{A.17}$$

This nonlinear system of equations is solved numerically, yielding the steady state of the detrended system, and thus the BGP of the model in original variables. All further analysis of the decentralized allocation is based on the (numerical) linearization of the five-dimensional dynamical system of equations (23')–(25'), (2), and (5), taking the BGP equality (32) as given.

Appendix B. Additional Parameters

We solve the four remaining equations in system (A.9)–(A.17) with respect to the remaining parameters; see table B.1. All these parameters are within admissible ranges. For instance, Pessoa (2005)

Table B.1. Baseline Calibration: Additional Parameters

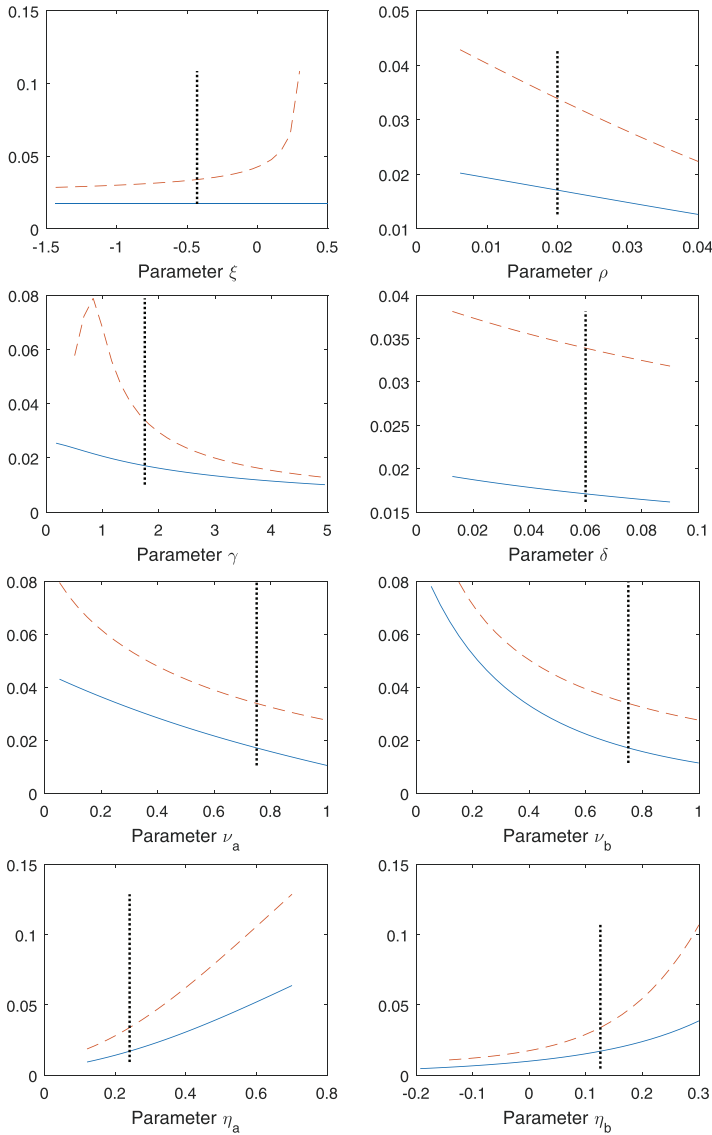
Parameter		Value
<i>Labor-Augmenting R&D</i>		
Unit Productivity	A	0.02
Capital Spillover Exponent	η_a	0.24
<i>Capital-Augmenting R&D</i>		
Unit Productivity	B	0.16
Capital Spillover Exponent	η_b	0.13
Degree of Decreasing Returns	ω	0.50
Obsolescence Rate	d	0.08
Spillover from Capital- to Labor-Augmenting Tech. Change	ϕ	0.30
Technology Choice Externality	ζ	115.28

estimates values for the obsolescence parameter d between 0 and 15 percent; our endogenously determined value is thus centered in that range. Comparing $\eta_a = 0.24$ with $\eta_b = 0.13$ signifies that, first of all, positive spillovers of capital intensity in R&D (effective capital augmentation of the R&D process) assuredly matter for R&D productivity, and second, that they are relatively more important for inventing new labor-augmenting technologies than capital-augmenting ones. Moreover, with $\phi = 0.3$, labor-augmenting R&D—the ultimate engine of long-run growth—is substantially reinforced by spillovers coming from the capital-augmenting R&D sector. On the other hand, $\omega = 0.5$ means that the scope for capital-augmenting R&D is quite strongly limited by decreasing returns. Given this benchmark calibration, as we said in the main text, the steady state is a saddle point.

Appendix C. Robustness Exercises: Impact of Parameter Variation on the Equilibrium Growth Rate

So far we have confirmed the received wisdom that the growth rate in the DA, g^{DA} , is socially suboptimal. This appears to be generally true in our model, regardless of its parameterization.

Figure C.1. Dependence of Equilibrium Growth on Model Parameters



Notes: The real economic growth rate g on vertical axis; corresponding parameter support on the horizontal axis. Social planner allocation (dashed lines), decentralized equilibrium (solid lines). The vertical dotted line in each graph represents the baseline calibrated parameter value.

Focusing on a reasonable parameter support, and assuming gross complements (see figure C.1), we can however identify a few credible cases where the difference between the two growth rates becomes small:

- A higher ρ (i.e., more impatience for current consumption), implies less capital and R&D accumulation and lower equilibrium growth than otherwise. If ρ is sufficiently large, then $g^{SP} \rightarrow_+ g^{DA}$.³¹
- If the consumption smoothing motive is sufficiently weak (γ high), then $g^{SP} \rightarrow_+ g^{DA}$.
- If the capital spillover exponents are weak, $\nu_a \rightarrow 0$ or $\nu_b \rightarrow 0$, then they attenuate the engine of long-run growth in the R&D equations and thus pull both g^{SP} and g^{DA} down.

Finally, note, departing from gross substitutes, we see the dramatic result that as σ (or equivalently ξ) increases, the gap $g^{SP} - g^{DA}$ hyperbolically widens; conversely, it narrows as substitution possibilities tend to zero (the Leontief case). We explore this in the next appendix.

³¹See also figure D.1.

Appendix D. Additional Figures

Figure D.1. Comparing Balanced Growth Paths, DA vs. SP: Dependence on the Time Preference

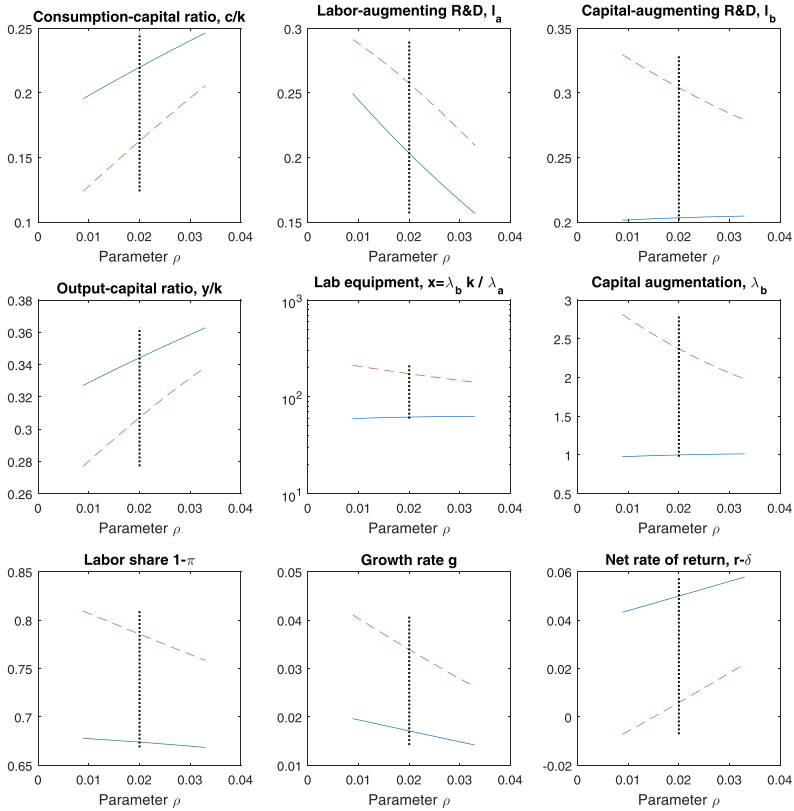


Figure D.2. Comparing Balanced Growth Paths, DA vs. SP: Dependence on the Intertemporal Elasticity of Substitution in Consumption

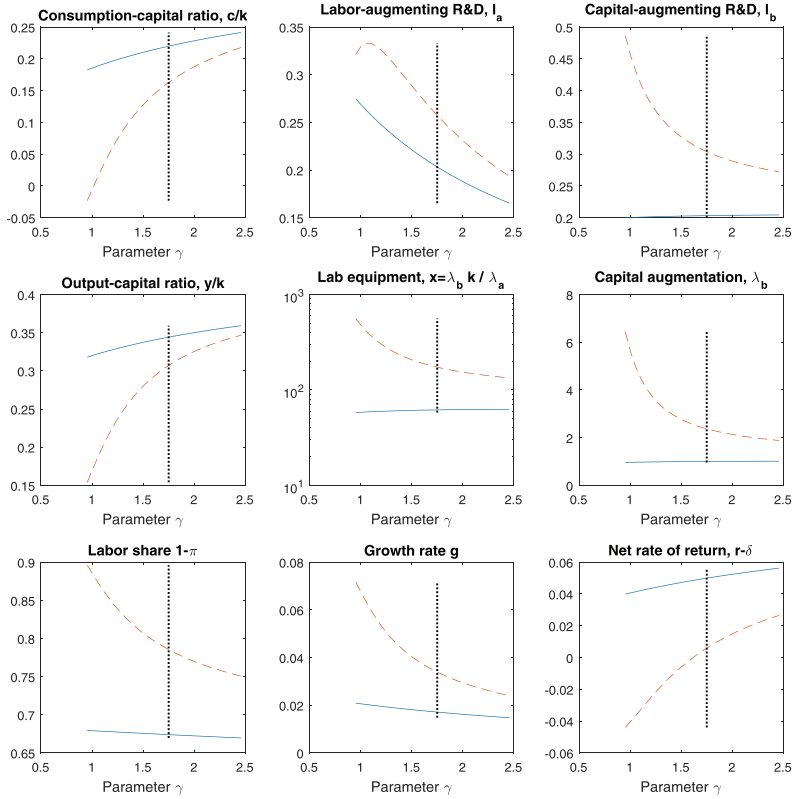


Figure D.3. Comparing Balanced Growth Paths, DA vs. SP: Dependence on the Capital Spillover Exponent in Capital-Augmenting R&D

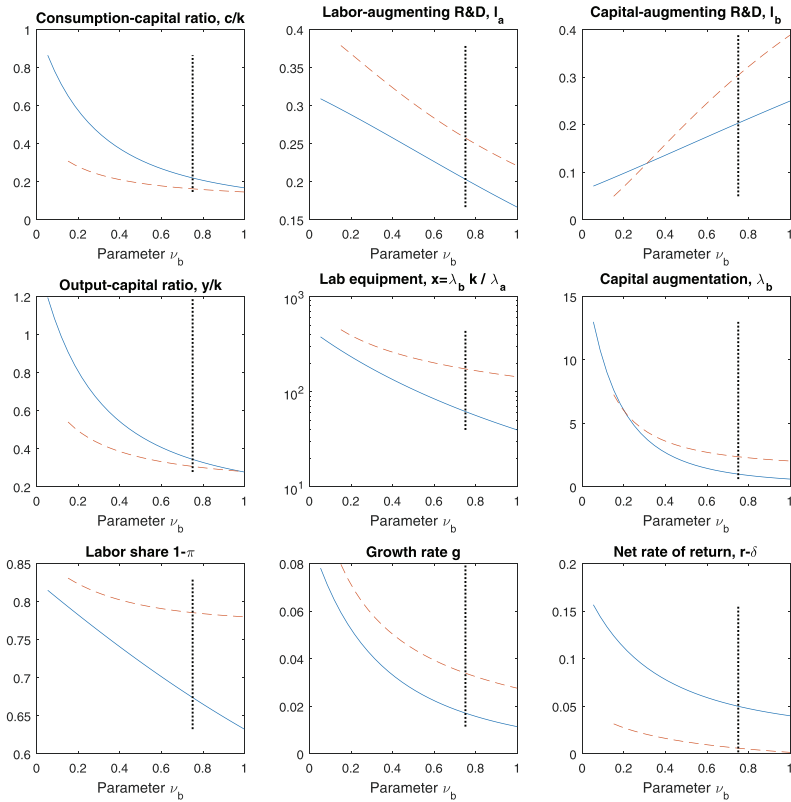
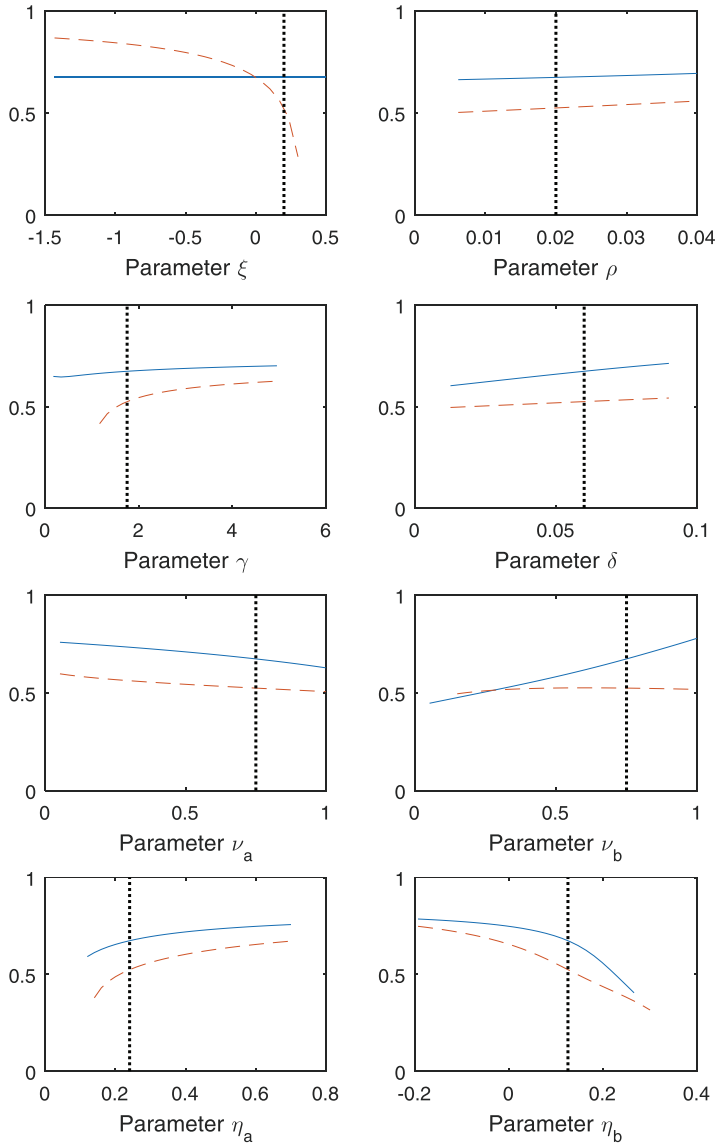
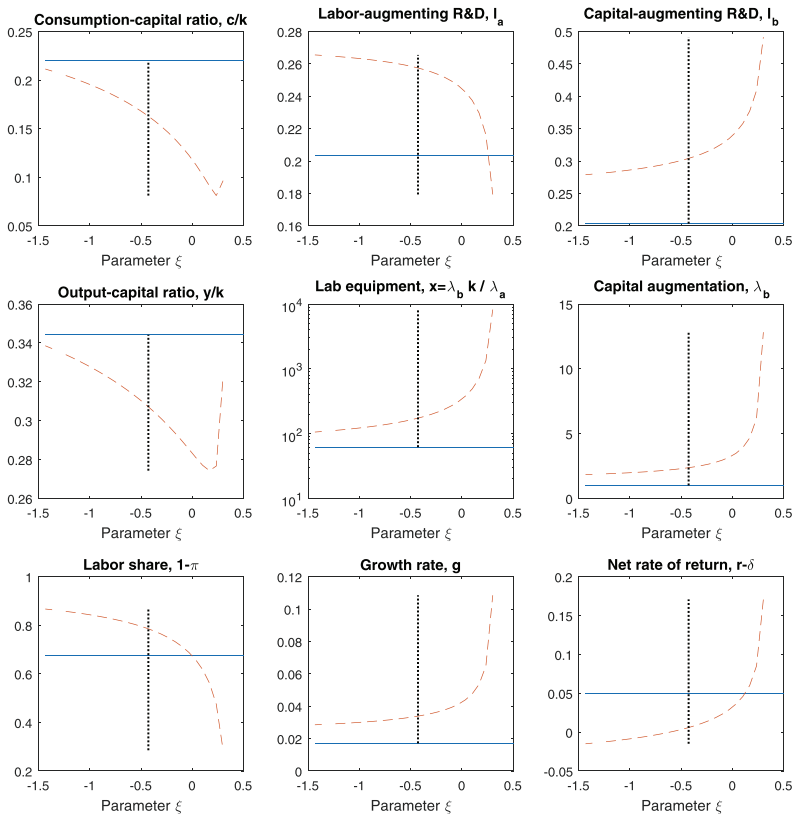


Figure D.4. Dependence of Equilibrium Labor Share on Model Parameters, for the Alternative Calibration of $\sigma = 1.25$ ($\xi = 0.2$)



Notes: $1 - \pi$ on vertical axis; corresponding parameter support on the horizontal axis. Social planner allocation (dashed lines), decentralized equilibrium (solid lines). The vertical dotted line in each graph represents the baseline calibrated parameter value.

Figure D.5. Dependence of BGP on Elasticity of Substitution, DA vs. SP



Notes: Social planner allocation (dashed lines), decentralized equilibrium (solid lines). The vertical dotted line in each graph represents the baseline calibrated parameter value.

References

Acemoglu, D. 2003. "Labor- and Capital-Augmenting Technical Change." *Journal of the European Economic Association* 1 (1): 1–37.

Alvarez-Pelaez, M. J., and C. Groth 2005. "Too Little or Too Much R&D?" *European Economic Review* 49 (2): 437–56.

- Antràs, P. 2004. "Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution." *Contributions to Macroeconomics* 4: 1–34.
- Arrow, K., H. B. Chenery, B. S. Minhas, and R. M. Solow. 1961. "Capital-Labor Substitution and Economic Efficiency." *Review of Economics and Statistics* 43: 225–50.
- Atkinson, A. B. 2015. *Inequality: What Can Be Done?* Harvard University Press.
- Atkinson, A. B., T. Piketty, and E. Saez. 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49 (1): 3–71.
- Azariadis, C. 1996. "The Economics of Poverty Traps — Part One: Complete Markets." *Journal of Economic Growth* 1 (4): 449–86.
- Barro, R. J., and X. X. Sala-i-Martin. 2003. *Economic Growth*. MIT Press.
- Bénassy, J.-P. 1998. "Is There Always Too Little Research in Endogenous Growth with Expanding Product Variety?" *European Economic Review* 42 (1): 61–69.
- Bengtsson, E. 2014. "Labour's Share in Twentieth-Century Sweden: A Reinterpretation." *Scandinavian Economic History Review* 62 (3): 290–314.
- Caselli, F. 2005. "Accounting for Cross-Country Income Differences." In *Handbook of Economic Growth*, Vol. 1A, ed. P. Aghion and S. N. Durlauf, 679–741 (chapter 9). North-Holland.
- Charpe, M., S. Bridji, and P. McAdam. 2020. "Labor Share and Growth in the Long Run." *Macroeconomic Dynamics* 24 (7): 1720–57.
- de La Grandville, O. 1989. "In Quest of the Slutsky Diamond." *American Economic Review* 79 (3): 468–81.
- . 2012. "How Much Should a Nation Save? A New Answer." *Studies in Nonlinear Dynamics and Econometrics* 16 (2): 1–36.
- . 2016. *Economic Growth: A Unified Approach*. Second edition. Cambridge University Press.
- de Loecker, J., J. Eeckhout, and G. Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics* 135 (2): 561–644.
- Dixit, A. K., and J. E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67 (3): 297–308.

- Galí, J., and M. Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44 (2): 195–222.
- Groth, C., and J. B. Madsen. 2016. "Medium-term Fluctuations and the 'Great Ratios' of Economic Growth." *Journal of Macroeconomics* 49 (September): 149–76.
- Growiec, J. 2007. "Beyond the Linearity Critique: The Knife-Edge Assumption of Steady-State Growth." *Economic Theory* 31: 489–99.
- Growiec, J., P. McAdam, and J. Mućk. 2018. "Endogenous Labor Share Cycles: Theory and Evidence." *Journal of Economic Dynamics and Control* 87 (February): 74–93.
- Growiec, J., and J. Mućk. 2020. "Isoelastic Elasticity of Substitution Production Functions." *Macroeconomic Dynamics* 24 (7): 1597–1634.
- Guo, J. T., and K. J. Lansing. 2009. "Capital–Labor Substitution and Equilibrium Indeterminacy." *Journal of Economic Dynamics and Control* 33 (12): 1991–2000.
- Herrendorf, B., C. Herrington, and A. Valentinyi. 2015. "Sectoral Technology and Structural Transformation." *American Economic Journal: Macroeconomics* 7 (4): 104–33.
- Irmen, A., and A. Tabakovic. 2020. "Factor Income Distribution and Endogenous Economic Growth: Piketty Meets Romer." *Economic Inquiry* 58 (3): 1342–61.
- Jaimovich, N., and H. E. Siu. 2020. "Job Polarization and Jobless Recoveries." *Review of Economics and Statistics* 102 (1): 129–47.
- Jalava, J., M. Pohjola, A. Ripatti, and J. Vilmunen. 2006. "Biased Technical Change and Capital–Labor Substitution in Finland, 1902–2003." *B.E. Journal of Macroeconomics* 6 (1): 1–20.
- Jones, C. I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103 (4): 759–84.
- . 1999. "Growth: With or Without Scale Effects?" *American Economic Review* 89 (2): 139–44.
- . 2005. "Growth and Ideas." In *Handbook of Economic Growth*, Vol. 1B, ed. P. Aghion and S. N. Durlauf, 1063–1111 (chapter 16). North-Holland.
- Jones, C. I., and J. C. Williams. 2000. "Too Much of a Good Thing? The Economics of Investment in R&D." *Journal of Economic Growth* 5 (1): 65–85.

- Jones, L. E., and R. E. Manuelli. 1990. "A Convex Model of Equilibrium Growth: Theory and Policy Implications." *Journal of Political Economy* 98 (5, Part 1): 1008–38.
- Kaas, L., and L. von Thadden. 2003. "Unemployment, Factor Substitution and Capital Formation." *German Economic Review* 4 (4): 475–95.
- Karabarbounis, L., and B. Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129 (1): 61–103.
- Kennedy, C. 1964. "Induced Bias in Innovation and the Theory of Distribution." *Economic Journal* 74 (295): 541–47.
- Klump, R. 2002. "Verteilung und Wirtschaftswachstum: Eine neoklassische Verallgemeinerung des Wachstumsmodell von Kaldor." In *Zur Theorie, Empirie und Politik der Einkommensverteilung*, ed. L. Menkhoff and F. Sell, 11–26. Berlin-Heidelberg-New York: Springer.
- Klump, R., and O. de La Grandville. 2000. "Economic Growth and the Elasticity of Substitution: Two Theorems and Some Suggestions." *American Economic Review* 90 (1): 282–91.
- Klump, R., P. McAdam, and A. Willman. 2007. "Factor Substitution and Factor-Augmenting Technical Progress in the US." *Review of Economics and Statistics* 89 (1): 183–92.
- . 2012. "The Normalized CES Production Functions: Theory and Empirics." *Journal of Economic Surveys* 26 (5): 769–99.
- Klump, R., and H. Preissler. 2000. "CES Production Functions and Economic Growth." *Scandinavian Journal of Economics* 102 (1): 41–56.
- Knoblach, M., and F. Stöckl. 2020. "What Determines the Elasticity of Substitution between Capital and Labor? A Literature Review." *Journal of Economic Surveys* 34 (4): 847–75.
- Kortum, S. 1993. "Equilibrium R&D and the Patent-R&D Ratio: U.S. Evidence." *American Economic Review* 83 (2): 450–57.
- Laeven, L., P. McAdam, and A. Popov. 2018. "Credit Shocks, Employment Protection, and Growth: Firm-level Evidence from Spain." ECB Working Paper No. 2166.
- León-Ledesma, M. A., P. McAdam, and A. Willman. 2010. "Identifying the Elasticity of Substitution with Biased Technical Change." *American Economic Review* 100 (4): 1330–57.
- León-Ledesma, M. A., and M. Satchi. 2019. "Appropriate Technology and Balanced Growth." *Review of Economic Studies* 86 (2): 807–35.

- Li, C.-W. 2000. "Endogenous vs. Semi-endogenous Growth in a Two-R&D-Sector Model." *Economic Journal* 110 (462): C109–C122.
- Liesner, T. 1989. *One Hundred Years of Economic Statistics*. Facts on File Publications.
- McAdam, P., and A. Willman. 2004. "Supply, Factor Shares and Inflation Persistence: Re-Examining Euro-Area New-Keynesian Phillips Curves." *Oxford Bulletin of Economics and Statistics* 66 (s1): 637–70.
- . 2013. "Medium Run Redux." *Macroeconomic Dynamics* 17 (4): 695–727.
- Muék, J. 2017. "Elasticity of Substitution between Labor and Capital: Robust Evidence from Developed Economies." Working Paper No. 271, Narodowy Bank Polski, Economic Research Department.
- Muék, J., P. McAdam, and J. Growiec. 2018. "Will the 'True' Labor Share Stand Up? An Applied Survey on Labor Share Measures." *Journal of Economic Surveys* 32 (4): 961–84.
- Oberfield, E., and D. Raval. 2018. "Micro Data and Macro Technology." *Econometrica* 89 (2): 703–32.
- Palivos, T., and G. Karagiannis. 2010. "The Elasticity of Substitution as an Engine of Growth." *Macroeconomic Dynamics* 14 (5): 617–28.
- Pessoa, A. 2005. "'Ideas' Driven Growth: The OECD Evidence." *Portuguese Economic Journal* 4 (1): 46–67.
- Piketty, T. 2014. *Capital in the Twenty-First Century*. Harvard University Press. Translated from French by Arthur Goldhammer.
- Piketty, T., and G. Zucman. 2014. "Capital is Back: Wealth–Income Ratios in Rich Countries, 1700–2010." *Quarterly Journal of Economics* 129 (3): 1255–1310.
- Pitchford, J. 1960. "Growth and the Elasticity of Factor Substitution." *Economic Record* 36 (76): 491–504.
- Ramsey, F. P. 1928. "A Mathematical Theory of Saving." *Economic Journal* 38 (152): 543–59.
- Romer, P. M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5, Part 2): S71–S102.
- Rowthorn, R. 1999. "Unemployment, Wage Bargaining and Capital–Labour Substitution." *Cambridge Journal of Economics* 23 (4): 413–25.

- Saam, M. 2008. "Openness to Trade as a Determinant of the Macroeconomic Elasticity of Substitution." *Journal of Macroeconomics* 30 (2): 691–702.
- Solow, R. M. 1960. "Investment and Technical Progress." In *Mathematical Methods in the Social Sciences*, ed. K. Arrow, S. Karlin, and P. Suppes, 89–104. Stanford University Press.
- Uzawa, H. 1961. "Neutral Inventions and the Stability of Growth Equilibrium." *Review of Economic Studies* 28 (1): 117–24.