Volume 17, Number 1 March 2021

IJCB

INTERNATIONAL JOURNAL OF CENTRAL BANKING

International Journal of Central Banking

Volume 17, Number 1 March 2021

The contents of this journal, together with additional materials provided by article authors, are available without charge at ***www.ijcb.org***.

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

ISSN: 1815-4654

Introduction

This issue of the *International Journal of Central Banking* includes two of the papers presented at the conference entitled "Structural Changes in the Financial System: New Theory and Evidence" hosted by the European Central Bank on August 20–21, 2020. The conference was co-organized with the European Central Bank and Danmarks Nationalbank. The two papers, chosen using the same rigorous standards applied to all *International Journal of Central Banking* content, are "Alternative Models of Interest Rate Pass-Through in Normal and Negative Territory" by Mauricio Ulate; and "The Rise of Fintech Lending to Small Businesses: Businesses' Perspectives on Borrowing" by Brett Barkley and Mark Schweitzer. The program committee for the conference was Tobias Adrian, Elena Carletti, Huberto Ennis, Linda Goldberg, Luc Laeven, and Steven Ongena.

Alternative Models of Interest Rate Pass-Through in Normal and Negative Territory*

Mauricio Ulate
Federal Reserve Bank of San Francisco

In the aftermath of the Great Recession, many countries used low or negative policy rates to stimulate the economy. These policies gave rise to a rapidly growing literature that seeks to understand and quantify their impact. A fundamental step when studying the effectiveness of low and negative policy rates is to understand their transmission to loan and deposit rates. This paper proposes two models of pass-through from policy rates to loan and deposit rates that can match important stylized facts while remaining parsimonious. These models can be used to study the transition between positive and negative policy rates and to quantify the impact of negative rates on banks.

JEL Codes: E32, E44, E52, E58, G21.

1. Introduction

During the Great Recession of 2008–09 many countries cut their policy rates to zero (or its vicinity) to fight the downturn and stimulate the economy. The slow recovery that followed the recession featured nominal rates that remained at zero in many advanced countries and even became negative in others. The effectiveness of these low and negative rates has been debated in the press, central banks, and the academic literature, but the matter remains unsettled. A fundamental issue when studying low or negative policy rates

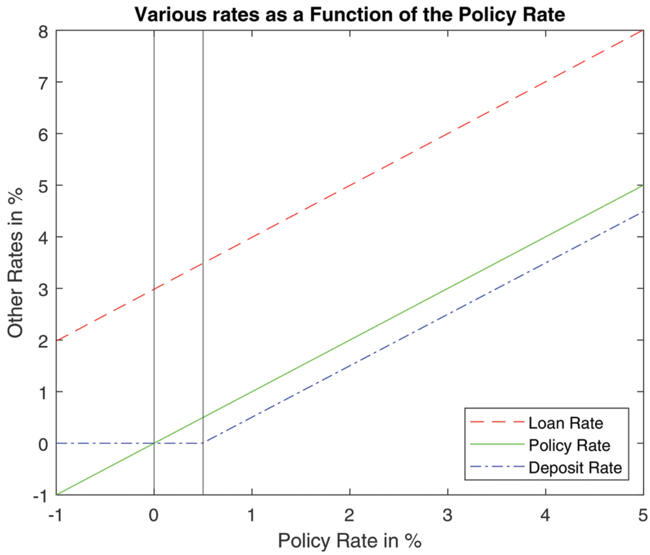
*The views in this paper do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System. For useful comments I thank Jose Vasquez, Ashley Lannquist, Rupal Kamdar, Yuriy Gorodnichenko, Andrés Rodríguez-Clare, Walker Ray, Juan Herreño, and participants in various seminars and conferences. All errors are my own. Author e-mail: mauricio.ulate@sf.frb.org.

is their transmission to other interest rates that play an important role in the broader economy. Two such rates are the interest rate that commercial banks charge on loans (hereafter referred to as the "loan rate") and the interest rate that commercial banks pay their customers for deposits (hereafter referred to as the "deposit rate"). The pass-through of the policy rate to loan and deposit rates is a crucial component in determining the effectiveness of cutting the policy rate in low or negative territory.

Empirically, papers like Drechsler, Savov, and Schnabl (2017) have found that the pass-through of the policy rate to deposit rates is between 0.5 and 0.6 when rates are in their normal range, while papers like Eisenschmidt and Smets (2019) have documented that this pass-through is close to zero when rates are very low or negative. For loan rates, Altavilla et al. (2019) and Ulate (2019) have documented a pass-through of between 0.5 and 1 when rates are in their normal range. The value of the loan rate pass-through when rates are low or negative is a more contested issue, with papers like Amzallag et al. (2019) and Eggertsson et al. (2019) claiming that the pass-through is close to zero (or negative), and papers like Eisenschmidt and Smets (2019) and Ulate (2019) finding that it is still positive. Even though there are disagreements in this literature, and the topic is still evolving, a rough consensus of the facts is that the pass-through of the policy rate to loan and deposit rates is positive but incomplete (say between 0.5 and 0.8) in normal times, the pass-through of the policy rate to the deposit rate is roughly zero in negative territory, and the pass-through of the policy rate to the loan rate is intermediate in negative territory.

In this paper, I propose two models of interest rate pass-through that can capture the facts mentioned in the previous paragraph while remaining tractable. These models extend and modify the static banking model of Ulate (2019), which is unable to capture non-unitary pass-through. The original model of Ulate (2019) contains separate borrowers and savers that solve a two-period problem. Additionally, it assumes that customers (i) choose a single bank from a continuum of possibilities over which they have differentiated preferences, (ii) choose their bank before other quantities of interest, (iii) have a unitary intertemporal elasticity of substitution (IES), and (iv) can only save/borrow through banks. These four assumptions imply that customers have constant elasticity of substitution (CES)

Figure 1. Behavior of Rates in the Original Model of Ulate (2019)

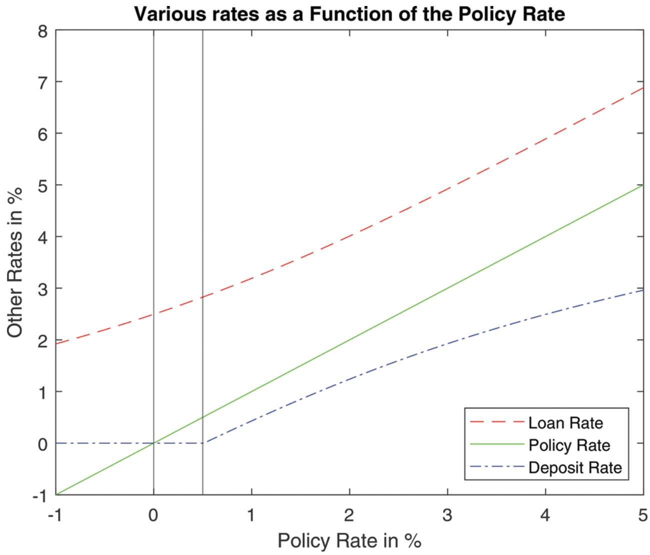


Note: This figure shows the loan rate, policy rate, and deposit rate behavior as a function of the policy rate in the original model of Ulate (2019).

preferences between banks in loan demand and deposit supply. As a consequence, banks set the loan rate as a markup on the policy rate and the deposit rate as a markdown on the policy rate during "normal times" (i.e., when the policy rate is above a threshold of roughly 50 basis points). These markups and markdowns are almost constant, generating a pass-through of the policy rate to the deposit rate and the loan rate which is essentially 1. This complete pass-through (illustrated in figure 1) during normal times is inconsistent with the stylized facts mentioned above.

The first extension developed in this paper maintains assumptions (i), (ii), and (iv) but deviates from Ulate (2019) by relaxing assumption (iii). Specifically, I assume that borrowers and savers have a CES utility function between today and tomorrow with an intertemporal elasticity of substitution greater than 1. Consequently, this extension is denoted the "High Intertemporal Substitution" model. For borrowers, an intertemporal elasticity of substitution

Figure 2. Behavior of Rates in the "High Intertemporal Substitution" Model



Note: This figure shows the loan rate, policy rate, and deposit rate behavior as a function of the policy rate in the "High Intertemporal Substitution" extended model.

greater than 1 means that when rates are high they want to borrow a small share of their income. This gives lenders "less monopoly power" and makes them charge a smaller loan spread. In contrast, savers want to save a higher share of their income when rates are high, which means that deposit-taking banks have "more monopoly power" and charge a higher deposit spread. This leads to a behavior of rates, illustrated in figure 2, which is consistent with the stylized facts about pass-through discussed earlier.

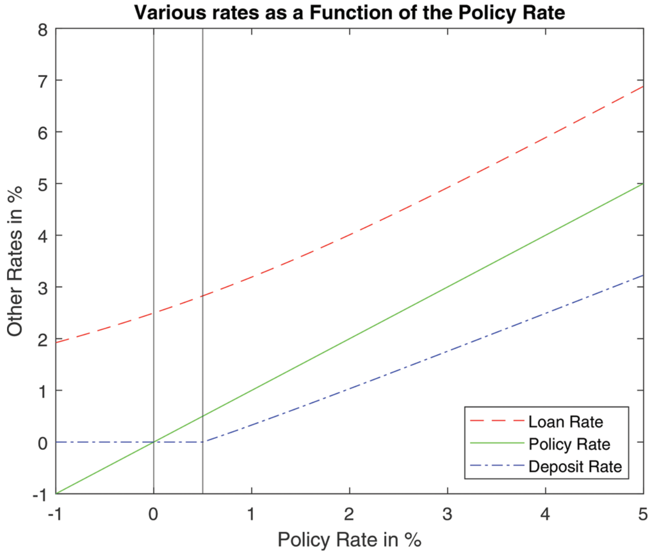
The second extension developed in this paper maintains assumptions (i), (ii), and (iii) but deviates from Ulate (2019) by relaxing assumption (iv). Specifically, savers are allowed to use three type of instruments: cash, deposits, and bonds. Furthermore, cash and deposits (combined through a CES aggregator) provide liquidity, which is valued by customers. Consequently, this extension is denoted the "Liquidity and Bonds" model. This setup implies that the choice of how many deposits to maintain is determined by the comparison of the price of deposits to the price of liquidity. In contrast, the choice of bank is determined by comparing a bank's price of liquidity with the price of liquidity offered by other banks. The total amount of deposits supplied to a bank is a combination of the amount that its customers want to deposit and the probability that a given customer chooses that bank. Hence, banks face two different margins of substitution, one given by the elasticity between deposits and cash in the liquidity aggregator, and the other one given by the dispersion in preferences across banks. This leads to a behavior of rates which I illustrate in figure 3. The behavior of the loan rate is identical to the one in the first model, but the behavior of the deposit rate is different, as this variable grows linearly with the policy rate.[1]

After discussing the models and their implications for loan and deposit rate pass-through, I proceed to discuss their implications for the return on equity (ROE) of banks. The behavior of ROE under the three models is displayed in figure 4. While the pattern of ROE is not exactly the same in the extended models as in the original model of Ulate (2019), the behavior is not too different. In all three models, ROE falls steeply with the policy rate below a certain threshold $\tilde{\imath}$, but has a more moderate behavior above the threshold.[2] In the static model of Ulate (2019) the slope of ROE below the threshold is around 5, while above the threshold it is around 1. In the "High Intertemporal Substitution" model, the slope above the threshold is around 0 at first and eventually becomes greater than 1. In the "Liquidity and Bonds" model, the slope above the threshold is slightly higher than in the original model throughout. Even though there are slight differences, the overall behavior of ROE is similar under models that feature non-unitary pass-through. This serves to

[1]The "Liquidity and Bonds" model relies on the same mechanism as the "High Intertemporal Substitution" model to obtain a non-unitary pass-through for borrowers, but uses a completely different mechanism to obtain a non-unitary pass-through for savers, as explained in the text. That is why the behavior of the loan rate is the same in both models but the behavior of the deposit rate is different.

[2]The value of the threshold and the reason for its existence are discussed extensively in Ulate (2019) and will also be covered in section 2. A second threshold $\underline{i}$ is also present but will not be discussed as much in this paper, since it is likely to be below -1.5 percent.

Figure 3. Behavior of Rates in the "Liquidity and Bonds" Model



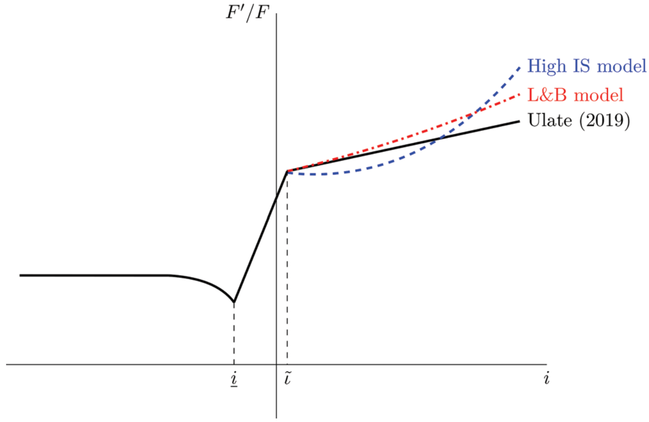
Note: This figure shows the loan rate, policy rate, and deposit rate behavior as a function of the policy rate in the "Liquidity and Bonds" extended model.

reassure the reader that the results in Ulate (2019) are not reliant on the assumption of (approximately) unitary pass-through.

The extended models proposed in this paper can be used to study the impact of negative nominal interest rates on banks. Since they feature a more realistic pass-through in "normal times," they can also be used to build models that more seamlessly capture the transition between positive and negative territory. Even though both extended models produce a similar pass-through in normal times, there is still value in having two alternative models, since researchers might want to include or exclude alternative saving vehicles (i.e., cash or bonds) in their dynamic stochastic general equilibrium (DSGE) models depending on their specific purposes.

To my knowledge, this is the first paper that develops banking models with non-unitary pass-through that contain a continuum of banks and monopoly power. Papers like Drechsler, Savov, and Schnabl (2017), Balloch and Koby (2019), Kurlat (2019), or Wang et al. (2019) have developed models where the pass-through of the

Figure 4. Model-Implied Relationship between ROE and *i*



Notes: This figure describes the model-implied relationship between bank (gross) return on equity (F'/F, denoted ROE), on the y-axis, and the policy rate (i), on the x-axis, for three different models. The levels $\tilde{\imath}$ and $\underline{i}$ represent thresholds where commercial banks start reacting differently to the policy rate; their expressions are given in section 2. The model of Ulate (2019), explained in section 2, is represented by the black line. The "High Intertemporal Substitution" extended model of section 3 is represented by the blue line. Finally, the "Liquidity and Bonds" extended model of section 4 is represented by the red line.

policy rate to the deposit rate in normal territory can differ from 1. Certain parameterizations of those models can produce a deposit pass-through in the 0.5 to 0.6 range. These models are related to the "Liquidity and Bonds" model, as agents can save not only via deposits with banks but also in cash or bonds. Cash and deposits provide liquidity services, while bonds do not. When rates are low, and bonds and money have a similar return, deposits are not very useful and banks have little monopoly power, so they set small spreads.

In the papers mentioned in the previous paragraph, the mechanism for non-unitary pass-through relies on having a limited number of banks. These banks, which have significant size, realize that they affect the aggregate deposit rate, which changes their rate-setting behavior. If these same models are modified to have a medium or large number of banks (or, in the limit, a continuum), then even with the introduction of bonds and cash, the pass-through of the policy rate to the deposit rate approaches 1. The timing assumption in the "Liquidity and Bonds" model, where customers must choose their bank before their saving amount, is what separates the second extended model from previous papers.

To illustrate the importance of the point in the previous paragraph, consider the model of Drechsler, Savov, and Schnabl (2017). For a given set of parameter values, and a single bank ($N = 1$), the pass-through of a cut in the policy rate from 2 percent to 1 percent is exactly half (i.e., 0.5).[3] However, if all parameters are kept fixed but the number of banks is increased to five ($N = 5$), the same measure of pass-through increases from 0.5 to 0.93. This means that even with a medium, yet realistic, number of banks, the model approximately delivers unitary pass-through. For a researcher that is interested in a local labor market with a small number of banks, these types of models can be useful to capture non-unitary pass-through. In contrast, for a researcher trying to calibrate a DSGE model at the national level, introducing a realistic number of banks would lead to a nearly complete pass-through in this family of models. Another downside of having a finite number of banks is that assumptions must be made about the evolution of the number of banks in order to be able to solve the model. In this paper, I develop models that feature non-unitary pass-through during normal times while featuring a continuum of banks, so that the setup remains tractable and can be used to analyze a national economy in a general equilibrium setup.

This paper is related to the theoretical literature that studies the usefulness of negative or low policy rates, while being more limited in scope. Brunnermeier and Koby (2018) study the "reversal rate" (the level of the interest rate where decreasing the policy rate further becomes contractionary for lending) in a model with monopoly power and capital gains in banks. Sims and Wu (2019) propose a framework to study three types of unconventional policies in a unified DSGE model. Eggertsson et al. (2019) propose a monetary DSGE model with banks that does not contain channels through which negative rates can be effective. De Groot and Haas (2020) study the signaling channel, a mechanism through which negative

[3]Specifically, the parameter values used here are $\delta = 1, \eta = 1.1, \varepsilon = 2$, and $\rho = 0.5$. Other parameter choices would deliver different values for the pass-through, but the overall message that increasing N pushes the pass-through toward unity would remain intact.

rates can stimulate the economy even if current deposit rates are stuck at zero. Wang (2019) studies how a low-rate environment can hurt banks and transfer the burden of the net interest margin from depositors to borrowers. Balloch and Koby (2019) also study the effects of a low-rate environment with an emphasis on Japan in a model with heterogeneous banks that have significant size. Rognlie (2016) studies negative rates in a model without banks where money demand does not become unbounded at zero. None of these papers contain models of non-unitary pass-through with monopoly power and a continuum of banks like the ones proposed in this paper.

While the current paper does not present any empirical results, it is motivated by the empirical literature that discusses the effectiveness of low and negative nominal interest rates. This literature includes papers such as Borio, Gambacorta, and Hofmann (2017), Nucera et al. (2017), Altavilla, Boucinha, and Peydro (2018), Basten and Mariathasan (2018), Claessens, Coleman, and Donnelly (2018), Hong and Kandrac (2018), Ampudia and Van den Heuvel (2019), Amzallag et al. (2019), Bottero et al. (2019), Demiralp, Eisenschmidt, and Vlassopoulos (2019), Eisenschmidt and Smets (2019), Heider, Saidi, and Schepens (2019), Lopez, Rose, and Spiegel (2020), etc. This literature has used different exposure measures (or cross-country panel identification) to study the effectiveness of negative rates, with conflicting results that would be impossible to summarize coherently in limited space. None of these papers propose models of banking like the ones developed here, but they support some of the stylized facts about pass-through mentioned above.

In the models developed in this paper, as in the static model of Ulate (2019), all financial contracts (both loans and deposits) have a duration of one period. While this sidestepping of maturity transformation in banking is partially justified by recent work (c.f. Drechsler, Savov, and Schnabl 2017, 2018), this is still a simplification adopted for tractability. This allows the models to deliver realistic pass-through properties without carrying around complicated asset and liability structures. Wang (2019) develops a model that can accommodate flexible maturity structures but doesn't contain monopoly power. More generally, recent papers like Begenau, Piazzesi, and Schneider (2015), Gomez et al. (2016), English, Van den Heuvel, and Zakrajšek (2018), and Hoffmann et al. (2019) discuss in much more detail the issue of banks' risk exposure.

This paper does not deal with the distinction between the short-run and the long-run effects of low or negative nominal interest rates. There are at least two short-run considerations which are not included in this paper. First, loan rates and deposit rates could react to the policy rate with a lag due to adjustment costs, as in Gerali et al. (2010). Second, changes in the nominal interest rate can give rise to short-run capital gains for the banking sector. These gains can stem from the maturity mismatch present in most commercial banks, or from long-lived securities that increase in value after a cut in the policy rate. This channel is present in papers like Brunnermeier and Koby (2018) or Wang (2019). Additionally, the prospect of a long period in low or negative territory might change bank behavior, since the adjustment costs of modifying their balance sheets or revamping their cash storage facilities become less relevant. A model that incorporates all of these issues would be useful, but it could also be too complex to serve as an intuition-building mechanism.

Recent work by Martinez-Miera and Repullo (2020) (see also Repullo 2004; Dell'Ariccia, Laeven, and Marquez 2014) has examined the implications of different levels of market power in banks that monitor risky loans with an unobservable and costly technology. They show that the impact of the safe rate on the risk-taking decisions of banks can vary with the amount of competition. When there is low market power, lower safe rates lead to lower intermediation margins and higher risk-taking. In contrast, when there is high market power, lower safe rates lead to higher intermediation margins and lower risk-taking. Since the models in this paper do not contain heterogeneous borrowers, they cannot speak to risk-taking effects. Nevertheless, the mode of competition is also important in this paper. If banks were in perfect competition, the results from this paper would no longer apply, and the pass-through in normal times would be complete.[4]

The rest of the paper proceeds as follows. Section 2 briefly describes the static model of Ulate (2019) and its implications for interest rate pass-through and bank ROE. Section 3 describes the "High Intertemporal Substitution" model, its assumptions, setup,

[4] Among others, see Berger et al. (2004), Claessens and Laeven (2004), Degryse and Ongena (2008), and Drechsler, Savov, and Schnabl (2017) for papers providing evidence of market power in the banking sector.

and intuition. Similarly, section 4 describes the "Liquidity and Bonds" model. Finally, section 5 concludes.

2. The Static Model of Ulate (2019)

In the model of Ulate (2019), there is a continuum of banks, indexed by j, between 0 and 1. Each bank is granted an amount of equity F_j as an endowment at the beginning of the period, which it combines with an amount of deposits D_j. On the asset side, banks issue loans L_j and hold reserves H_j. Banks seek to maximize their resources at the end of the period, once loans and deposits have been repaid. Each bank faces a downward-sloping loan demand and an upward-sloping deposit supply captured through a CES aggregator.

The maximization problem that individual bank j faces is the following:

$$\max_{i_j^l, L_j, i_j^d, D_j, H_j} \quad (1+i_j^l)L_j + (1+i)H_j - (1+i_j^d)D_j$$

$$\text{s.t.}$$

$$L_j = \left(\frac{1+i_j^l}{1+i^l}\right)^{-\varepsilon^l} \mathbf{L} \tag{1}$$

$$D_j = \begin{cases} \left(\frac{1+i_j^d}{1+i^d}\right)^{-\varepsilon^d} \mathbf{D} & \text{if } i_j^d \geq 0 \\ 0 & \text{if } i_j^d < 0 \end{cases} \tag{2}$$

$$L_j + H_j = F_j + D_j \tag{3}$$

$$H_j \geq 0. \tag{4}$$

Equation (1) represents loan demand. Equation (2) represents deposit supply, and it indicates that a bank obtains no deposits if it sets negative nominal deposit rates. Equations (1) and (2) can be derived directly from the behavior of borrowers and savers using the four assumptions mentioned in the introduction, as illustrated in appendixes A.1–A.3 of Ulate (2019). The aggregate amounts of loans demanded by firms and deposits supplied by households are $\mathbf{L}$ and $\mathbf{D}$, respectively. These aggregate quantities are assumed to be unaffected by any rates in this partial equilibrium model, but they can be made endogenous in more elaborate general equilibrium models. Equation (3) is the bank balance sheet constraint, indicating that total assets (loans plus reserves) have to equal liabilities (which are just deposits) plus equity. Equation (4) states that reserves at the central bank must be nonnegative.

This model assumes that $\varepsilon^l > 1$ and $\varepsilon^d < -1$, that all banks are given the same amount of initial equity $F_j = \mathbf{F}$, and that $\mathbf{D} > \mathbf{L} > \mathbf{F}$. The formal solution to the bank problem is described in Ulate (2019); here I provide a brief summary. The solution consists of regimes that apply depending on the level of the policy rate. Regime 1 applies when $i \geq \tilde{\imath}$, regime 2 when $\underline{i} \leq i < \tilde{\imath}$, and regime 3 when $i < \underline{i}$. The thresholds are given by $\tilde{\imath} \equiv -\frac{1}{\varepsilon^d} > 0$ and

$$\underline{i} = \frac{\left(\frac{\mathbf{L}}{\mathbf{F}}\right)^{\frac{1}{\varepsilon^l}} \frac{\varepsilon^l}{\varepsilon^l - 1} - \frac{1}{\varepsilon^l - 1}\frac{\mathbf{L}}{\mathbf{F}} - 1}{1 + \frac{1}{\varepsilon^l - 1}\frac{\mathbf{L}}{\mathbf{F}} + \frac{\mathbf{D}}{\mathbf{F}} - \left(\frac{\mathbf{L}}{\mathbf{F}}\right)^{\frac{1}{\varepsilon^l}} \frac{\varepsilon^l}{\varepsilon^l - 1}} < 0.$$

In regime 1, when the policy rate is in "normal" territory, all banks set the same gross loan and deposit rates, which are given as a markup and markdown on the gross policy rate:

$$1 + i_j^l = \frac{\varepsilon^l}{\varepsilon^l - 1}(1 + i), \qquad 1 + i_j^d = \frac{\varepsilon^d}{\varepsilon^d - 1}(1 + i).$$

This is reminiscent of the solution of the pricing problem of a monopolistically competitive good producer. In this model, the absolute values of ε^l and ε^d will be high in order to match the steady-state spreads between the loan rate and the policy rate and between the policy rate and the deposit rate. Consequently, the values of $\frac{\varepsilon^l}{\varepsilon^l - 1}$ and $\frac{\varepsilon^d}{\varepsilon^d - 1}$ will be close to 1, and pass-through will be nearly complete. As mentioned in Ulate (2019), in this regime all banks obtain an amount of deposits equal to the aggregate deposit supply ($\mathbf{D}$), give an amount of loans equal to the aggregate demand of loans ($\mathbf{L}$), and hold a positive amount of reserves.

In regime 2, when $\underline{i} \leq i < \tilde{\imath}$, all banks set $i_j^d = 0$, receive an amount of deposits $\mathbf{D}$, give an amount of loans $\mathbf{L}$, and still hold a positive amount of reserves at the central bank. In this regime the loan-rate-setting behavior of banks is the same as in regime 1, since the marginal use of commercial banks' resources is still as reserves at the central bank. Intuitively, regime 2 exists because there is a range of low and negative policy rates where banks prefer to receive deposits even if they earn a low or negative spread on them, because it allows them to maintain their leverage and earn more on their loan franchise. Regime 2 stops applying when the policy rate crosses the threshold $\underline{i} < 0$, where offering deposits at a zero rate is so costly that at least one commercial bank has incentives to stop doing so. Regime 3, which applies when $i < \underline{i}$, is described in detail in Ulate (2019), but will not be discussed here.

The behavior of interest rates with respect to the policy rate in this model is described in figure 1. Since the policy rate is in both axes, the green line is simply the diagonal.[5] Additionally, it is clear that the loan rate is a markup over the policy rate and the deposit rate is a markdown over the policy rate. Moreover, the spreads are essentially constant when the policy rate is above $\tilde{\imath}$ (which is around 50 basis points). The x axis in figure 1 only contains realizations of the policy rate that are above $\underline{i}$, since $\underline{i} \approx -2\%$ in this model. The behavior of return on equity (ROE) is depicted in figure 4 with a solid black line. The interest rate $\tilde{\imath}$ represents the threshold where further cuts in the policy rate would turn deposit rates negative in the absence of the deposit zero lower bound (ZLB). However, since deposit rates are constrained by zero, $\tilde{\imath}$ instead represents the point where lowering the policy rate further starts affecting banks disproportionately, because they cannot charge their usual spread on deposits.

The nearly complete pass-through displayed by this model in "normal territory" makes it unable to match the stylized facts described in the introduction. In the following sections I modify this model in order to capture a non-unitary pass-through in normal times.

3. "High Intertemporal Substitution" Model

As mentioned in the introduction, the first extended model relaxes the assumption of a unitary intertemporal elasticity of substitution

[5]For color versions of the figures, see the online version of the paper at http://www.ijcb.org.

and instead assumes that agents (both borrowers and savers) have an elasticity of substitution between today and tomorrow which is greater than 1. This seemingly small change has profound implications for the loan and deposit pass-through. For borrowers, it means that when rates are high they do not want to borrow much. For savers, in contrast, it means that when rates are high they want to save a lot. Consequently, high rates amplify the monopoly power of banks on the deposit side but decrease it on the loan side. Therefore, high rates lead to small loan spreads and high deposit spreads, allowing this model to capture the stylized pass-through facts mentioned in the introduction.

Agents in this model choose their bank before their allocations (i.e., consumption and saving/borrowing), and they have a preference shock (with an extreme value distribution) across different banks. Having agents choose their bank before their allocations captures frictions like switching costs or limited attention spans. These frictions correspond to the realistic feature that customers usually choose their bank once and stick with it for long periods of time.[6] Additionally, assuming a preference shock across banks captures the fact that due to idiosyncratic or geographical characteristics, certain customers might prefer a given bank for reasons orthogonal to its interest rates.[7]

The assumptions in this model imply that deposit supply and loan demand for a given bank contain two different elasticities, one related to the elasticity of substitution between today and tomorrow, and another one related to the elasticity of substitution between different banks (stemming from the preference shock). If the intertemporal elasticity is greater than the elasticity across banks, increases in the policy rate will increase the loan elasticity but decrease the deposit elasticity, leading to smaller loan spreads but higher deposit spreads. In the following subsections, I describe the problem of the saver, the problem of the borrower, and the problem of the bank, respectively.

[6] Brunetti, Ciciretti, and Djordjevic (2016) find evidence in an Italian data set that less than one-quarter of households switch their bank in a horizon of two years.

[7] Papers like Repullo (2004) and Andres and Arce (2012) have used geographical variation to model heterogeneous preferences in the banking sector.

3.1 The Problem of the Savers

A representative saver has CES preferences between today and tomorrow characterized by the following utility function:

$$U(C_0, C_1) = \left[(\alpha^d)^{\frac{1}{\theta^d}} C_0^{\frac{\theta^d - 1}{\theta^d}} + (1-\alpha^d)^{\frac{1}{\theta^d}} C_1^{\frac{\theta^d - 1}{\theta^d}} \right]^{\frac{\theta^d}{\theta^d - 1}},$$

where C_0 is consumption today, C_1 is consumption tomorrow, α^d is the importance of consumption today, and θ^d is the elasticity of substitution between today and tomorrow. The saver has income $\overline{Y}^d$ today and no income tomorrow. Therefore, he must save in order to consume tomorrow. Saving can only be done in a continuum of banks between 0 and 1.

An individual bank is indexed by j. Bank j offers a deposit rate i_j^d. Savers must first choose the bank that they will put their savings into, and then the amount that they will save. The budget constraint of the saver, conditional on the choice of bank j, is given by

$$C_0 + \frac{C_1}{1 + i_j^d} = \overline{Y}^d.$$

The solution to this problem is

$$C_0 = \alpha^d \left(\frac{1}{p_j^d} \right)^{-\theta^d} \frac{\overline{Y}^d}{p_j^d}, \qquad C_1 = (1-\alpha^d) \left(\frac{1/(1+i_j^d)}{p_j^d} \right)^{-\theta^d} \frac{\overline{Y}^d}{p_j^d},$$

where

$$p_j^d \equiv \left(\alpha^d + (1-\alpha^d) \left(\frac{1}{1+i_j^d} \right)^{1-\theta^d} \right)^{\frac{1}{1-\theta^d}}$$

is the price index of aggregate consumption for a saver that chooses bank j. The indirect utility function of this consumer is $v_j^d = \ln(\overline{Y}^d) - \ln(p_j^d)$.[8]

[8] After adding a logarithm to the utility function, which does not alter the maximization problem.

Up to now, the quantities being discussed are conditional on choosing bank j. The next step is to characterize the choice of bank, which is the first stage of the decision process. I assume that the bank choice stage can be described by a stochastic utility approach, where the total utility of choosing a given bank is the sum of the indirect utility obtained in the second stage, and a stochastic component that varies across banks.[9] Mathematically,

$$V_j^d = v_j^d + \mu^d \epsilon_j^d,$$

where μ^d is a positive constant and ϵ_j^d is a random variable with zero mean and unit variance.

Assuming that the ϵ_j^d random variable is independently and identically distributed with type 1 extreme value distribution, the probability of choosing bank j is given by

$$Pr_j^d = Pr(V_j^d = \max_r V_r^d) = \frac{e^{v_j^d/\mu^d}}{\int_0^1 e^{v_r^d/\mu^d} dr} = \frac{(p_j^d)^{-\frac{1}{\mu^d}}}{\int_0^1 (p_r^d)^{-\frac{1}{\mu^d}} dr},$$

as in McFadden (1973). Substituting $1/\mu^d$ for $\varepsilon^d - 1$, the previous expression becomes

$$Pr_j^d = \frac{(p_j^d)^{1-\varepsilon^d}}{\int_0^1 (p_r^d)^{1-\varepsilon^d} dr} = \left(\frac{p_j^d}{p^d}\right)^{1-\varepsilon^d},$$

where p^d is the usual price index: $p^d = \left(\int_0^1 (p_r^d)^{1-\varepsilon^d} dr\right)^{\frac{1}{1-\varepsilon^d}}$. This indicates that the probability of choosing a given bank is determined by the ratio of the price of aggregate consumption offered by that bank over the price of aggregate consumption offered by the "average" bank, with an "elasticity" ε^d which captures how sensitive the probability is to deviations from the average price. However, what matters for banks is not only the probability that they are chosen but also the amount of deposits that they receive. This is the multiplication of the probability that they are chosen by the amount

[9]As mentioned earlier, this stochastic component can be due to several things: geographic variation, switching costs, recommendations of family or friends, etc.

of deposits that they receive conditional on being chosen. Multiplying the probability that bank j is chosen (Pr_j^d) with the amount of deposits held at bank j if it is chosen (d_j), one obtains

$$d_j Pr_j^d = (1-\alpha^d)(1+i_j^d)^{\theta^d-1}(p_j^d)^{\theta^d-\varepsilon^d}(p^d)^{\varepsilon^d-1}\overline{Y}^d.$$

I interpret $(1-\alpha^d)(1+i^d)^{\theta^d-1}(p^d)^{\theta^d-1}\overline{Y}^d$ as aggregate deposits and denote it with $\mathbf{D}$. Even though this quantity varies with the policy rate, here I will keep $\mathbf{D}$ fixed and ignore its dependence on the policy rate. I do this in order to preserve the partial equilibrium nature of the model in Ulate (2019), which assumes that the banks optimize in response to changes in the policy rate, but the aggregate amount of loans and deposits remains fixed.[10] Additionally, I interpret $d_j Pr_j^d$ as the amount deposited to bank j once the whole population of savers is taken into account, and denote this by D_j. Then

$$D_j = \left(\frac{1+i_j^d}{1+i^d}\right)^{\theta^d-1}\left(\frac{p_j^d}{p^d}\right)^{\theta^d-\varepsilon^d}\mathbf{D}. \tag{5}$$

This means that deposit supply for bank j has two distinct elasticity margins.

Using equation (5) and the definition of p_j^d, the elasticity of deposit supply with respect to the gross deposit rate can be written as

$$\begin{aligned}\gamma_j^d \equiv \frac{\partial D_j}{\partial(1+i_j^d)}\frac{1+i_j^d}{D_j} &= s_j^d(\varepsilon^d-1)+(1-s_j^d)(\theta^d-1)\\ &= (\theta^d-1)-s_j^d(\theta^d-\varepsilon^d),\end{aligned} \tag{6}$$

where $s_j^d \equiv ((1-\alpha^d)(1+i_j^d)^{\theta^d-1})/(\alpha^d+(1-\alpha^d)(1+i_j^d)^{\theta^d-1})$, $0 \leq s_j^d \leq 1$, and $\frac{\partial s_j^d}{\partial i_j^d} > 0$. When the deposit rate charged by bank

[10]There are several margins besides the ones considered here which can affect deposit supply. Those margins might even dominate the influence of the policy rate. That is why in this paper I choose to abstract from analyzing changes in aggregate deposit supply and focus instead on the allocation of such aggregate supply.

j is high, the weight s_j^d is high, and the elasticity is driven toward $\varepsilon^d - 1$. Conversely, when i_j^d is low, the weight s_j^d is low, and the elasticity is driven toward $\theta^d - 1$. To the extent that $\theta^d > \varepsilon^d$, increasing the policy rate (which will increase the deposit rate of all banks) decreases the elasticity and leads to higher markups.[11] This implies that the pass-through from the policy rate to the deposit rate is smaller than 1.

When i_j^d is low, the price of consumption of bank j (p_j^d) tends to 1. Hence, the second parenthesis in (5) plays a smaller role and the main elasticity left in the expression for deposit supply is $\theta^d - 1$. Intuitively, when the deposit rate is low, it is not an important factor in how customers substitute between banks, and hence the main elasticity that determines deposit supply is the intertemporal elasticity of substitution. In contrast, a high deposit rate i_j^d tilts the price of consumption for bank j toward $(1 + i_j^d)^{-1}$, allowing the second parenthesis in (5) to be combined with the first, with an elasticity of $\varepsilon^d - 1$. Intuitively, with a high deposit rate (and $\theta^d > 1$), most consumption happens tomorrow, making the IES irrelevant and turning $\varepsilon^d - 1$ into the crucial elasticity governing deposit supply.

3.2 The Problem of the Borrowers

The problem of the borrower is somewhat related to the one of the saver, but it is slightly different. A borrower has CES preferences between today and tomorrow,

$$U(C_0, C_1) = \left[(\alpha^l)^{\frac{1}{\theta^l}} C_0^{\frac{\theta^l - 1}{\theta^l}} + (1 - \alpha^l)^{\frac{1}{\theta^l}} C_1^{\frac{\theta^l - 1}{\theta^l}} \right]^{\frac{\theta^l}{\theta^l - 1}},$$

where C_0 is consumption today, C_1 is consumption tomorrow, α^l is the importance of consumption today, and θ^l is the elasticity of substitution between consumption today and consumption tomorrow. In contrast to the saver, the borrower only has income $\overline{Y}^l$ tomorrow. He needs to borrow in order to consume today. He can borrow

[11] In the limit of a continuous-time model, the elasticity of substitution between "today" and "tomorrow" should be high, approaching infinity. In contrast, the elasticity of substitution between banks will remain bounded because the switching costs operate across several periods. It is then natural to expect $\theta^d > \varepsilon^d$.

from a continuum of banks between 0 and 1. The budget constraint conditional on the choice of bank j can be expressed as

$$(1+i_j^l)C_0 + C_1 = \overline{Y}^l.$$

The solution to this problem is

$$C_0 = \alpha^l \left(\frac{1+i_j^l}{p_j^l}\right)^{-\theta^l} \frac{\overline{Y}^l}{p_j^l}, \qquad C_1 = (1-\alpha^l)\left(\frac{1}{p_j^l}\right)^{-\theta^l} \frac{\overline{Y}^l}{p_j^l},$$

where

$$p_j^l \equiv \left(\alpha^l (1+i_j^l)^{1-\theta^l} + 1 - \alpha^l\right)^{\frac{1}{1-\theta^l}}$$

is the price index of aggregate consumption for a borrower that chooses bank j.

Taking the same stochastic utility approach as in the case of the saver, the probability for a consumer of choosing bank j is given by $Pr_j^l = \left(p_j^l/p^l\right)^{1-\varepsilon^l}$, where p^l is the usual CES price index: $p^l = \left(\int_0^1 (p_r^l)^{1-\varepsilon^l} dr\right)^{\frac{1}{1-\varepsilon^l}}$. Multiplying the amount borrowed from bank j if it is chosen (B_j) by this probability, one obtains

$$B_j Pr_j^l = \alpha^l (1+i_j^l)^{-\theta^l} (p_j^l)^{\theta^l - \varepsilon^l} (p^l)^{\varepsilon^l - 1} \overline{Y}^l.$$

Interpret $\alpha^l (1+i^l)^{-\theta^l} (p^l)^{\theta^l - 1} \overline{Y}^l$ as aggregate borrowing and denote it as $\mathbf{L}$. Additionally, interpret $B_j Pr_j^l$ as the amount borrowed from each bank j once the whole population of borrowers is taken into account, and denote this by L_j. Then

$$L_j = \left(\frac{1+i_j^l}{1+i^l}\right)^{-\theta^l} \left(\frac{p_j^l}{p^l}\right)^{\theta^l - \varepsilon^l} \mathbf{L}. \tag{7}$$

The interpretation of this equation is similar to the one of equation (5). The elasticity of loan demand with respect to the gross loan rate is

$$\gamma_j^l \equiv \frac{\partial L_j}{\partial(1+i_j^l)} \frac{1+i_j^l}{L_j} = -s_j^l \varepsilon^l - (1-s_j^l)\theta^l = -\theta^l + s_j^l(\theta^l - \varepsilon^l), \tag{8}$$

where $s_j^l \equiv (\alpha^l(1+i_j^l)^{1-\theta^l})/(\alpha^l\left(1+i_j^l\right)^{1-\theta^l}+1-\alpha^l)$, $0 \leq s_j^l \leq 1$, and $\frac{\partial s_j^l}{\partial i_j^l} < 0$.

3.3 The Problem of the Banks

The setup of the banking problem is similar to the one in section 2, but here I also introduce exogenous costs of issuing loans and deposits (μ^l and μ^d). Banks choose the interest rate they charge on loans i_j^l, the amount they lend, the interest rate they pay on deposits i_j^d, the amount of deposits they take, and the amount of reserves they hold at the central bank, which earns the policy rate i, subject to several constraints. The maximization problem that the individual bank j faces is therefore the following:

$$\max_{i_j^l, L_j, i_j^d, D_j, H_j} \quad (1+i_j^l-\mu^l)L_j + (1+i)H_j - (1+i_j^d+\mu^d)D_j$$

$$\text{s.t.}$$

$$L_j = \left(\frac{1+i_j^l}{1+i^l}\right)^{-\theta^l}\left(\frac{p_j^l}{p^l}\right)^{\theta^l-\varepsilon^l}\mathbf{L}$$

$$D_j = \begin{cases} \left(\frac{1+i_j^d}{1+i^d}\right)^{\theta^d-1}\left(\frac{p_j^d}{p^d}\right)^{\theta^d-\varepsilon^d}\mathbf{D} & \text{if } i_j^d \geq 0 \\ 0 & \text{if } i_j^d < 0 \end{cases}$$

$$L_j + H_j = F_j + D_j$$

$$H_j \geq 0,$$

where μ^l is the cost of issuing loans and μ^d is the cost of issuing deposits.

In regime 1, where the banks can solve their problem unconstrained by the ZLB on deposits and optimally hold positive reserves, the first-order conditions (FOCs) with regard to the gross loan rate and the gross deposit rate are

$$0 = L_j + [(1+i_j^l) - (1+i) - \mu^l]\frac{\partial L_j}{\partial 1+i_j^l}$$

$$0 = -D_j + [(1+i) - (1+i_j^d) - \mu^d]\frac{\partial D_j}{\partial 1 + i_j^d}.$$

Using the elasticities provided in equations (6) and (8), these equations can be simplified to

$$1 + i_j^l = \frac{\gamma_j^l}{\gamma_j^l + 1}(1 + i + \mu^l), \qquad 1 + i_j^d = \frac{\gamma_j^d}{\gamma_j^d + 1}(1 + i - \mu^d).$$

Since the elasticities γ_j^l and γ_j^d contain i_j^l and i_j^d, respectively, these equations don't provide a closed-form solution for the loan rate and the deposit rate, but they can be solved numerically. Nevertheless, the previous equations are still very useful, since they clarify that the gross loan rate is set as a markup (since $\gamma_j^l < -1$) on the gross policy rate and the gross deposit rate is set as a markdown (since $\gamma_j^d > 0$) on the gross policy rate. Since all banks are identical, they all charge the same loan rate and pay the same deposit rate (denoted by i^l and i^d). Return on equity for banks is then given by

$$\frac{F'}{F} - 1 = i + (i^l - i - \mu^l)\frac{\mathbf{L}}{F} + (i - i^d - \mu^d)\frac{\mathbf{D}}{F}.$$

In regime 2 banks pay a zero rate on deposits and obtain a fixed amount of deposits $\mathbf{D}$, and choose the interest rate they charge on loans i_j^l, the amount they lend, and the amount of reserves they hold in the central bank. The solution for the loan rate is exactly the same as in regime 1. Return on equity has the same expression as in regime 1 after setting $i^d = 0$. The solution for regime 3 is a bit complicated but works very similarly to that of regime 3 in the original static model of Ulate (2019).

If I assume parameter values $\alpha^d = \alpha^l = 0.9, \varepsilon^d = \varepsilon^l = 10, \theta^l = \theta^d = 100, \mu^l = 0.8\%, \mu^d = -0.6\%, \mathbf{D}/F = 9$, and $\mathbf{L}/F = 10$, then the behavior of rates is the one illustrated in figure 2 and the behavior of ROE is the one illustrated by the blue (dashed) line in figure 4. Importantly, the model exhibits non-unitary pass-through similar to the one in the data. While the parameter values that I assume (in order to obtain a pass-through that can match the stylized facts described in the introduction) are not carefully calibrated, this setup illustrates the fact that models with a non-unitary pass-through can still feature a behavior of bank ROE that is similar to the one in Ulate (2019).

4. "Liquidity and Bonds" Model

The second extended model relies on different mechanisms to generate a non-unitary pass-through in the loan rate and the deposit rate. On the loan side the mechanism is exactly the same as in the previous model. Consequently, the problem of the borrowers is not described here. The problem of the savers is different, since the intertemporal elasticity of substitution is once again assumed to be unitary, but agents can now save in cash, bonds, or deposits with a continuum of banks. The next subsection describes the problem of the saver, and the following one describes the bank's problem.

4.1 The Problem of the Savers

I assume that there is an individual consumer that lives for two periods, denoted 0 and 1. This consumer has a total income of $\overline{Y}^d$ in the first period and he can consume in both periods. To consume in period 0 is easy for this consumer; it can be done directly. However, to consume in period 1, the consumer must save some of his current income $\overline{Y}^d$. He can save in three ways: through one of a continuum of banks between 0 and 1 (indexed with j), in cash (which offers a nominal return of 0 percent), or in bonds that pay a gross return of $(1 + i)$.

The decision process of this consumer happens in two stages. In the first stage, the consumer decides which bank he wants to save with, and in the second stage he chooses the amounts he wants to allocate to cash, deposits, and bonds. First, I will describe the problem of a consumer that has already chosen bank j, and then I will describe the way that the bank choice is made. I assume that the direct utility function of the consumer conditional on his choice of bank j is given by

$$U(C_0, C_1, \mathcal{L}_j) = \ln(C_0) + \beta \ln(C_1) + \gamma \ln(\mathcal{L}_j),$$

where β is the discount factor between periods, γ is the importance of liquidity in utility, and C_t is consumption in period t. Additionally, $\mathcal{L}_j$ represents liquidity services, which are the following combination of deposits in bank j and cash:

$$\mathcal{L}_j = \left((\alpha^d)^{\frac{1}{\theta^d}} d_j^{\frac{\theta^d - 1}{\theta^d}} + (1-\alpha^d)^{\frac{1}{\theta^d}} M_j^{\frac{\theta^d - 1}{\theta^d}} \right)^{\frac{\theta^d}{\theta^d - 1}},$$

where α^d is the importance of deposits in liquidity provision, θ^d is the elasticity of substitution between cash and deposits in liquidity provision, d_j are deposits at bank j, and M_j is the amount of cash held conditional on the choice of bank j.

The first- and second-period budget constraints of the saver (again, conditional on the choice of bank j) are

$$\begin{aligned} P_0 C_0 &= P_0 \overline{Y}^d - d_j - M_j - B_j \\ P_1 C_1 &= (1 + i_j^d) d_j + (1+i) B_j + M_j, \end{aligned}$$

where $1 + i_j^d$ is the gross deposit rate paid between periods 0 and 1 by bank j (which is known by the consumer with certainty), B_j is the amount of bonds held conditional on the choice of bank j, i is the policy rate (which is assumed to be the return on bonds), and P_t is the price index in period t. The aggregate budget constraint can then be expressed as

$$C_0 = \overline{Y}^d - \frac{1}{1+i}\frac{P_1}{P_0} C_1 - \frac{i - i_j^d}{1+i}\frac{d_j}{P_0} - \frac{i}{1+i}\frac{M_j}{P_0}.$$

The solution to the saver's problem conditional on the choice of bank is

$$C_0 = \frac{\overline{Y}^d}{1+\beta+\gamma}, \qquad C_1 = \frac{\beta(1+r)}{1+\beta+\gamma}\overline{Y}^d,$$

$$\mathcal{L}_j = \frac{\gamma(1+i)}{1+\beta+\gamma}\frac{\overline{Y}^d}{p_j^d}, \qquad d_j = \alpha^d \left(\frac{i - i_j^d}{p_j^d} \right)^{-\theta^d} \mathcal{L}_j,$$

where P_0 has been normalized to one, $1 + r \equiv (1+i)\frac{P_0}{P_1}$, and the price of liquidity p_j^d is given by

$$p_j^d \equiv \left[\alpha^d (i - i_j^d)^{1-\theta^d} + (1-\alpha^d) i^{1-\theta^d} \right]^{\frac{1}{1-\theta^d}}.$$

With these quantities, the indirect utility function conditional on borrowing from bank j can be expressed as

$$v_j^d = (1+\beta+\gamma)(\ln(\overline{Y}^d) - \ln(1+\beta+\gamma)) + \beta\ln(\beta) \\ + \gamma\ln(\gamma) + \beta\ln(1+r) + \gamma\ln(1+i) - \gamma\ln(p_j^d).$$

Then, as in Anderson, de Palma, and Thisse (1988), assume that the first stage (the bank choice stage), is described by a stochastic utility approach:

$$V_j^d = v_j^d + \mu^d \epsilon_j^d,$$

where μ^d is a positive constant and ϵ_j^d is a random variable with zero mean and unit variance. Assuming that the ϵ_j^d random variables are independently and identically distributed with type 1 extreme value distribution, the probability for a consumer of choosing bank j is

$$Pr_j^d = Pr(V_j^d = \max_r V_r^d) = \frac{e^{v_j^d/\mu^d}}{\int_0^1 e^{v_r^d/\mu^d} dr} = \frac{(p_j^d)^{-\frac{\gamma}{\mu^d}}}{\int_0^1 (p_r^d)^{-\frac{\gamma}{\mu^d}} dr}.$$

Substituting $-\gamma/\mu^d$ for $1-\varepsilon^d$, the previous expression can be rewritten as

$$Pr_j^d = \frac{(p_j^d)^{1-\varepsilon^d}}{\int_0^1 (p_r^d)^{1-\varepsilon^d} dr} = \left(\frac{p_j^d}{p^d}\right)^{1-\varepsilon^d},$$

where p^d is the aggregate price of liquidity defined in the usual way. Multiplying d_j by this probability and simplifying, one obtains

$$d_j Pr_j^d = \alpha^d \frac{\gamma(1+i)}{1+\beta+\gamma}\frac{\overline{Y}^d}{p^d}\left(\frac{i-i^d}{p^d}\right)^{-\theta^d}\left(\frac{i-i_j^d}{i-i^d}\right)^{-\theta^d}\left(\frac{p_j^d}{p^d}\right)^{\theta^d-\varepsilon^d}.$$

Interpret $\alpha^d \frac{\gamma(1+i)}{1+\beta+\gamma}\frac{\overline{Y}^d}{p^d}\left(\frac{i-i^d}{p^d}\right)^{-\theta^d}$ as aggregate deposits and denote them with **D**. Additionally, interpret $d_j Pr(j)$ as the amount deposited in bank j once the whole population of consumers is taken into account, and denote this by D_j. Then

$$D_j = \left(\frac{i - i_j^d}{i - i^d}\right)^{-\theta^d} \left(\frac{p_j^d}{p^d}\right)^{\theta^d - \varepsilon^d} \mathbf{D}. \tag{9}$$

This is related to equation (5), but it is different in several aspects. First, the exponent of the first term is $-\theta^d$ instead of $\theta^d - 1$. Second, the quantity inside the first parenthesis is a ratio of spreads $(i - i_j^d)$ instead of a ratio of gross interest rates (because now the customers have a bigger selection of saving instruments). Third, the definition of p_j^d is different in this context.

4.2 The Problem of the Banks

The setup of the banking problem is exactly the same as in section 3.3, with a single change to make deposit supply follow (9) instead of (5). The maximization problem that individual bank j faces is therefore the following:

$$\max_{i_j^l, L_j, i_j^d, D_j, H_j} \quad (1 + i_j^l - \mu^l)L_j + (1 + i)H_j - (1 + i_j^d + \mu^d)D_j$$

s.t.

$$L_j = \left(\frac{1 + i_j^l}{1 + i^l}\right)^{-\theta^l} \left(\frac{p_j^l}{p^l}\right)^{\theta^l - \varepsilon^l} \mathbf{L}$$

$$D_j = \begin{cases} \left(\frac{i - i_j^d}{i - i^d}\right)^{-\theta^d} \left(\frac{p_j^d}{p^d}\right)^{\theta^d - \varepsilon^d} \mathbf{D} & \text{if } i_j^d \geq 0 \\ 0 & \text{if } i_j^d < 0 \end{cases}$$

$$L_j + H_j = F_j + D_j$$

$$H_j \geq 0.$$

The FOC for the loan rate is exactly the same as in section 3.3. Meanwhile, the derivative of deposit supply with regard to i_j^d is

$$\frac{\partial D_j}{\partial i_j^d} = \theta^d \frac{D_j}{i - i_j^d} - (\theta^d - \varepsilon^d)\frac{D_j}{p_j^d}\alpha^d \left(\frac{i - i_j^d}{p_j^d}\right)^{-\theta^d}.$$

The FOC with regard to i_j^d is the following:

$$0 = -D_j + (i - i_j^d - \mu^d)\frac{\partial D_j}{\partial i_j^d}.$$

Combining the previous two equations, one obtains

$$\begin{aligned}0 = \alpha^d(i - i_j^d)^{2-\theta^d}(1-\varepsilon^d) + (i - i_j^d)(1-\alpha^d)i^{1-\theta^d}(1-\theta^d)\\ + \mu^d\theta^d(1-\alpha^d)i^{1-\theta^d} + \mu^d\varepsilon^d\alpha^d(i - i_j^d)^{1-\theta^d}.\end{aligned}$$

As in the previous extended model, this equation cannot be solved explicitly for i_j^d, but it can be solved numerically. Return on equity for banks is described by the same expression as in the first extended model, and regimes 2 and 3 work in a similar way as well.

If I assume parameter values $\alpha^l = 0.9$, $\varepsilon^l = 10$, $\theta^l = 100$, and $\mu^l = 0.8\%$ on the loan side, and $\alpha^d = 0.5$, $\varepsilon^d = 2$, $\theta^d = 0.5$, and $\mu^d = 0.1\%$ on the deposit side, as well as $\mathbf{D}/F = 9$ and $\mathbf{L}/F = 10$, then the behavior of rates is the one illustrated in figure 3 and the behavior of ROE is the one illustrated by the red (dash-dotted) line in figure 4. As in the case of the "High Intertemporal Substitution" model, this second model also exhibits non-unitary pass-through similar to that in the data. The parameter values that I assume on the loan side are the same as the ones assumed in the "High Intertemporal Substitution" model. On the deposit side, $\alpha^d = 0.5$ indicates that deposits and cash have the same importance in the liquidity aggregator, $\theta^d = 0.5$ indicates that deposits and cash are not very substitutable, and $\varepsilon^d = 2$ indicates that banks have substantial monopoly power. Kurlat (2019) uses data to estimate a version of θ^d in a related model, and finds a value of 0.52, consistent with the value of 0.5 used in this paper.

While this model ends up delivering a behavior of rates similar to that in the "High Intertemporal Substitution" model, it relies on a completely different mechanism to deliver non-unitary deposit pass-through. This can be beneficial to researchers that want to include cash and bonds in their general equilibrium models for alternative reasons. Additionally, the saving parameter values are more realistic in this extension.

5. Conclusion

This paper proposes static and partial equilibrium models of the banking sector in order to study the pass-through of the policy rate to the loan rate and the deposit rate. First, the paper discusses the partial equilibrium model of Ulate (2019). This model is useful to convey intuition and to study negative nominal interest rates, but it features a complete pass-through of the policy rate to loans and deposit rates in "normal territory." This complete pass-through is not consistent with stylized facts indicating that the pass-through of the policy rate to loan and deposit rates is between 0.5 and 0.8 during normal times.

Next, the paper modifies the static framework of Ulate (2019) and proposes two models which can match the aforementioned stylized facts while remaining parsimonious. Importantly, the proposed models do not rely on having large banks to obtain a realistic pass-through, as they can deliver a non-unitary pass-through even with a continuum of banks.

The first model relies on a CES utility function between today and tomorrow, a sequential choice of bank and saving (or borrowing) amounts, and differentiation between banks. For borrowers, an intertemporal elasticity of substitution greater than 1 implies that they want to borrow a small amount when rates are high. This gives lenders "less monopoly power" and makes them charge a smaller loan spread when rates are high. In contrast, banks charge a higher deposit spread when rates are high.

In the second model, agents can save using cash, deposits in a continuum of banks, or bonds. Additionally, cash and bonds provide liquidity services through a CES aggregator. Savers must first choose a bank and only then choose their allocations (amount of deposits, cash, or bonds). When rates are high, the return differential between bonds and cash is high, making deposits valuable and allowing banks to charge a high deposit spread.

Overall, the extended models provide a parsimonious way of capturing non-unitary pass-through in normal territory, while also providing realistic pass-through in negative territory. Additionally, they do not require a small number of banks and hence sidestep the associated complication of determining the evolution of the number of banks. Moreover, the extended models have similar implications for return on equity as the static model of Ulate (2019), and they suggest that having a non-unitary pass-through in normal territory does not modify substantially the analysis under negative rates.

In this paper, I don't discuss how changing the amount of total reserves in the system affects the economy or the effectiveness of negative rates. Nevertheless, in the models discussed in this paper, the effect of increasing the central bank balance sheet (through programs like quantitative easing or targeted longer-term refinancing operations) would be to increase the amount of reserves in the system. This would increase the exposure of commercial banks to negative rates, thereby diminishing their effectiveness. Balance sheet expansion could also have an effect in normal times, although the models in this paper cannot speak to that. This topic is studied more explicitly by papers such as Ray (2019) or Sims and Wu (2019), among many others.

The three models discussed in this paper indicate that a cut in the policy rate in negative territory affects banks more than usual. However, this does not indicate that negative nominal interest rates (or, to be more precise, rates below $\tilde{\imath}$) are ineffective or harmful. As discussed in Ulate (2019), even if commercial bank profitability is being adversely affected by negative rates, a cut in the policy rate in negative territory can still be expansionary in a general equilibrium model. This occurs for several reasons. First, lower loan rates can stimulate investment and output. Second, higher loan demand allows banks to substitute reserves for loans, shielding them from negative rates. Third, negative rates can signal lower rates in the future (via the signaling channel emphasized in de Groot and Haas 2020).

In the DSGE model of Ulate (2019), the effectiveness of a cut in negative territory is between 60 percent and 90 percent of its effectiveness in positive territory (in terms of welfare). The models proposed in this paper have similar implications for bank profitability and pass-through in negative territory as the model in Ulate (2019). Hence, they would also indicate that negative rates are expansionary until the policy rate reaches the disintermediation threshold $\underline{i}$ (which is between –1.5 percent and –2 percent). This indicates that the effective lower bound (ELB) can be lower than zero (the ZLB). Additionally, this can occur despite the fact that commercial banks start being disproportionately affected by policy cuts even above the ZLB (since the first threshold $\tilde{\imath}$ is above zero).

References

Altavilla, C., M. Boucinha, and J.-L. Peydro. 2018. "Monetary Policy and Bank Profitability in a Low Interest Rate Environment." *Economic Policy* 33 (96): 531–86.

Altavilla, C., L. Burlon, M. Giannetti, and S. Holton. 2019. "Is There a Zero Lower Bound? The Effects of Negative Policy Rates on Banks and Firms." ECB Working Paper No. 2289.

Ampudia, M., and S. J. van den Heuvel. 2019. "Monetary Policy and Bank Equity Values in a Time of Low and Negative Interest Rates." Finance and Economics Discussion Series No. 2019-064, Board of Governors of the Federal Reserve System.

Amzallag, A., A. Calza, D. Georgarakos, and J. Sousa. 2019. "Monetary Policy Transmission to Mortgages in a Negative Interest Rate Environment." ECB Working Paper No. 2243.

Anderson, S. P., A. de Palma, and J.-F. Thisse. 1988. "The CES and the Logit: Two Related Models of Heterogeneity." *Regional Science and Urban Economics* 18 (1): 155–64.

Andres, J., and O. Arce. 2012. "Banking Competition, Housing Prices and Macroeconomic Stability." *Economic Journal* 122 (565): 1346–72.

Balloch, C., and Y. Koby. 2019. "Low Rates and Bank Loan Supply: Theory and Evidence from Japan." Technical Report.

Basten, C., and M. Mariathasan. 2018. "How Banks Respond to Negative Interest Rates: Evidence from the Swiss Exemption Threshold." Working Paper No. 6901, CESifo Group Munich.

Begenau, J., M. Piazzesi, and M. Schneider. 2015. "Banks' Risk Exposures." Technical Report, Mimeo.

Berger, A. N., A. Demirgüç-Kunt, R. Levine, and J. G. Haubrich. 2004. "Bank Concentration and Competition: An Evolution in the Making." *Journal of Money, Credit and Banking* 36 (3, Part 2): 433–51.

Borio, C., L. Gambacorta, and B. Hofmann. 2017. "The Influence of Monetary Policy on Bank Profitability." *International Finance* 20 (1): 48–63.

Bottero, M., C. Minoiu, J.-L. Peydro, A. Polo, A. F. Presbitero, and E. Sette. 2019. "Expansionary Yet Different: Credit Supply and Real Effects of Negative Interest Rate Policy." Working Paper No. 1090, Barcelona Graduate School of Economics.

Brunetti, M., R. Ciciretti, and L. Djordjevic. 2016. "The Determinants of Household's Bank Switching." *Journal of Financial Stability* 26 (October): 175–89.

Brunnermeier, M. K., and Y. Koby. 2018. "The Reversal Interest Rate." NBER Working Paper No. 25406.

Claessens, S., N. Coleman, and M. Donnelly. 2018. "Low-For-Long Interest Rates and Banks' Interest Margins and Profitability: Cross-country Evidence." *Journal of Financial Intermediation* 35 (Part A): 1–16.

Claessens, S., and L. Laeven. 2004. "What Drives Bank Competition? Some International Evidence." *Journal of Money, Credit and Banking* 36 (3, Part 2): 563–83.

de Groot, O., and A. Haas. 2020. "The Signalling Channel of Negative Interest Rates." CEPR Discussion Paper No. 14268.

Degryse, H., and S. Ongena. 2008. "Competition and Regulation in the Banking Sector: A Review of the Empirical Evidence on the Sources of Bank Rents." In *Handbook of Financial Intermediation and Banking*, ed. A. V. Thakor and A. W. A. Boot, 483–554 (chapter 15). Elsevier.

Dell'Ariccia, G., L. Laeven, and R. Marquez. 2014. "Real Interest Rates, Leverage, and Bank Risk-taking." *Journal of Economic Theory* 149 (January): 65–99.

Demiralp, S., J. Eisenschmidt, and T. Vlassopoulos. 2019. "Negative Interest Rates, Excess Liquidity and Retail Deposits: Banks' Reaction to Unconventional Monetary Policy in the Euro Area." Working Paper No. 2283, European Central Bank.

Drechsler, I., A. Savov, and P. Schnabl. 2017. "The Deposits Channel of Monetary Policy." *Quarterly Journal of Economics* 132 (4): 1819–76.

———. 2018. "Banking on Deposits: Maturity Transformation without Interest Rate Risk." NBER Working Paper No. 24582.

Eggertsson, G. B., R. E. Juelsrud, L. H. Summers, and E. G. Wold. 2019. "Negative Nominal Interest Rates and the Bank Lending Channel." NBER Working Paper No. 25416.

Eisenschmidt, J., and F. Smets. 2019. "Negative Interest Rates: Lessons from the Euro Area." In *Monetary Policy and Financial Stability: Transmission Mechanisms and Policy Implications*, ed. A. Aguirre, M. Brunnermeier, and D. Saravia, 13–42 (chapter 2). Vol. 26 of Central Banking, Analysis, and Economic Policies Book Series. Central Bank of Chile.

English, W. B., S. J. Van den Heuvel, and E. Zakrajšek. 2018. "Interest Rate Risk and Bank Equity Valuations." *Journal of Monetary Economics* 98 (October): 80–97.

Gerali, A., S. Neri, L. Sessa, and F. M. Signoretti. 2010. "Credit and Banking in a DSGE Model of the Euro Area." *Journal of Money, Credit and Banking* 42 (s1): 107–41.

Gomez, M., A. Landier, D. Sraer, and D. Thesmar. 2016. "Banks' Exposure to Interest Rate Risk and the Transmission of Monetary Policy." ESRB Working Paper No. 13.

Heider, F., F. Saidi, and G. Schepens. 2019. "Life below Zero: Bank Lending under Negative Policy Rates." *Review of Financial Studies* 32 (10): 3728–61.

Hoffmann, P., S. Langfield, F. Pierobon, and G. Vuillemey. 2019. "Who Bears Interest Rate Risk?" *Review of Financial Studies* 32 (8): 2921–54.

Hong, G. H., and J. Kandrac. 2018. "Pushed Past the Limit? How Japanese Banks Reacted to Negative Interest Rates." IMF Working Paper No. 18/131.

Kurlat, P. 2019. "Deposit Spreads and the Welfare Cost of Inflation." *Journal of Monetary Economics* 106 (October): 78–93.

Lopez, J. A., A. K. Rose, and M. M. Spiegel. 2020. "Why Have Negative Nominal Interest Rates Had Such a Small Effect on Bank Performance? Cross Country Evidence." *European Economic Review* 124 (May): Article 103402.

Martinez-Miera, D., and R. Repullo. 2020. "Interest Rates, Market Power, and Financial Stability." CEPR Discussion Paper No. 15063.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics,* ed. P. Zarembka, 105–42. New York: Academic Press.

Nucera, F., A. Lucas, J. Schaumburg, and B. Schwaab. 2017. "Do Negative Interest Rates Make Banks Less Safe?" *Economics Letters* 159 (October): 112–15.

Ray, W. 2019. "Monetary Policy and the Limits to Arbitrage: Insights from a New Keynesian Preferred Habitat Model." Technical Report.

Repullo, R. 2004. "Capital Requirements, Market Power, and Risk-taking in Banking." *Journal of Financial Intermediation* 13 (2): 156–82.

Rognlie, M. 2016. "What Lower Bound? Monetary Policy with Negative Interest Rates." Mimeo.

Sims, E. R., and J. C. Wu. 2019. "Evaluating Central Banks' Tool Kit: Past, Present, and Future." NBER Working Paper No. 26040.

Ulate, M. 2019. "Going Negative at the Zero Lower Bound: The Effects of Negative Nominal Interest Rates." Working Paper No. 2019-21, Federal Reserve Bank of San Francisco.

Wang, O. 2019. "Banks, Low Interest Rates, and Monetary Policy Transmission." Technical Report.

Wang, Y., T. M. Whited, Y. Wu, and K. Xiao. 2019. "Bank Market Power and Monetary Policy Transmission: Evidence from a Structural Estimation." Technical Report.

The Rise of Fintech Lending to Small Businesses: Businesses' Perspectives on Borrowing*

Brett Barkley and Mark Schweitzer
Federal Reserve Bank of Cleveland

Online lending through fintech firms is a rapidly expanding segment of the financial market that is receiving much attention from investors and increasing scrutiny from regulators. To assess how fintech firms' entry is altering the choices and outcomes of small businesses that borrow from them, we analyze data from the Federal Reserve's Small Business Credit Survey, a unique data source on the experiences of business owners with new and traditional sources of credit. We find that fintech lenders have substantially expanded the small business finance market by reaching borrowers less likely to be served by traditional lenders and that businesses using online lenders are younger, smaller, and less profitable than the average small or medium-sized enterprise in the United States. After controlling for compositional differences between online and bank borrowers, we find that businesses using fintech lenders generally apply for smaller loan amounts but value the option of fintech loans. Businesses that receive fintech loans expect more revenue and employment growth than those receiving a bank loan; however, they are less satisfied than businesses that borrow from banks but more satisfied than businesses that were denied credit.

JEL Codes: G21, G23, G28, C31.

*Brett Barkley is a data scientist in the Supervision and Regulation Department of the Federal Reserve Bank of Cleveland (brett.barkley@clev.frb.org). Mark E. Schweitzer is a senior vice president in the Research Department of the Federal Reserve Bank of Cleveland (mark.schweitzer@clev.frb.org). The views stated herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System.

1. Introduction

Fintech firms are a rapidly growing set of technology companies providing alternatives to traditional banking services, most often exclusively in an online environment. Fintech firms compete in financial services markets including consumer payments, asset management, and consumer and business lending. Overall, fintech lenders averaged nearly $12 billion in quarterly originations through the first half of 2018 (Darden, Dixit, and Mason 2018), and their lending to small businesses increased from approximately $121 million in quarterly originations during 2013 to $2 billion in quarterly originations during 2018. The 2020 pandemic and recession affected fintech lenders' existing business models, but several of them had substantial roles in providing Paycheck Protection Program (PPP) loans, with 19 fintech lenders originating more than 250,000 PPP loans amounting to approximately $6 billion (U.S. Small Business Administration 2020); other PPP loans were made by financial institutions like Cross River Bank, WebBank, and Celtic Bank on behalf of fintech lenders, accounting for an additional $12.5 billion (Federal Reserve 2020). The entrance of new types of lenders raises potential coordination challenges (Goldstein, Jagtiani, and Klein 2019) and important regulatory issues as new lenders increasingly compete with more heavily regulated banking institutions (Philippon 2018). Despite substantial investments and growing activity levels, fintech lenders have been lightly regulated to date (U.S. Department of the Treasury 2016 and Basel Committee on Banking Supervision 2018).

Only a few studies have explored fintech as a financing alternative for small businesses (Slattery 2014; Jagtiani and Lemieux 2019; and Balyuk, Berger, and Hackney 2020). Of these, our work is closest to Balyuk, Berger, and Hackney (2020). They use state-level changes in bank structures to show that two online-only, small business lenders have increased in the markets where the presence of local banks declined. Similar to our findings, they find that these two fintech lenders offer somewhat riskier loans. But all of these studies, including Balyuk, Berger, and Hackney (2020), have been constrained in their examination of fintech lending by having access only to data that have been released by particular fintech lenders, and those data do not include the set of all possible borrowers.[1] Our analysis complements these studies by using borrower-side data obtained from a survey of small businesses, which allows us to examine a broader set of borrowers and a fuller range of credit outcomes. This is important because, for example, if small businesses denied by banks are similar to businesses approved by fintech lenders, comparing the two provides a more complete picture as to whether fintech is merely substituting for bank credit in places where the latter has declined or truly expanding the credit market.

An older literature has focused on the roles different types of banking entities play in the financing and growth of small businesses. Community banks have long been recognized as an important source of small business credit (Berger and Udell 2002; Wiersch and Shane 2013; Robb and Robinson 2014). Despite a growing market share for large banks in small business lending dating back to the 1990s, several studies have shown that community banks still have an advantage in providing appropriate credit products to this market (Berger et al. 2005; Deyoung, Glennon, and Nigro 2008; Deyoung et al. 2011). As evidence of community banks' staying power in the small business lending market, note that 45 percent of the $525 billion in PPP loans were made by banks with less than $10 billion in assets (U.S. Small Business Administration 2020). We examine how different types of traditional lenders (large banks, community banks, and credit unions) differ from online lenders in providing financing to small businesses and how these new lending alternatives have been working for the small businesses that use them.

To collect data on the financing needs and experiences of small businesses, Federal Reserve Banks have conducted an annual survey of firms (the Small Business Credit Survey, or SBCS), which reached national coverage starting in 2016. Since that time, the SBCS has included questions about online lenders as well as traditional lenders. The survey focuses on measuring the financial needs and outcomes of businesses with fewer than 500 full- or part-time employees.[2] While the survey participants include thousands of small businesses,

[1] Mach, Carter, and Slattery (2014) and Jagtiani and Lemieux (2019) both examine LendingClub's publicly available data. Balyuk, Berger, and Hackney (2020) examine LendingClub and Funding Circle data.

[2] The survey includes nonemployer firms, but for this analysis we focus on businesses with at least one employee.

they are not a stratified random sample. Instead, participants are contacted through partner organizations and then the sample is weighted to reflect national small business characteristics according to census data. At this point, we are aware of no alternative data sources on the experiences of small businesses with both fintech firms and banks.

While banks have historically played an important role in meeting small businesses' financing needs, the SBCS reveals that fintech firms are now a substantial source of credit: in 2018, about 32 percent of small businesses that sought financing applied with a fintech or online lender[3] versus 44 percent with small banks and 49 percent with large banks. We use SBCS data from 2016 to 2018 to analyze the extent to which borrowers using online sources (the term used in the survey) would have been likely to have had their needs met by traditional lenders (a category that includes large and small banks and credit unions). To investigate the value of these loans, we then apply treatment effect estimators which flexibly control for compositional differences of the credit applicants and measure the impact of and ex post borrower satisfaction with online lenders. Overall, we find that fintech lenders have expanded lending to small businesses largely to the benefit of those businesses.

2. Small Business Credit Survey Design and Coverage

The Federal Reserve's Small Business Credit Survey is an annual survey of business establishments with fewer than 500 employees. It collects information about business performance, financing needs and choices, and borrowing experiences. The survey is designed to inform policymakers about how the small business credit environment affects firm operation and growth.[4]

The Federal Reserve partners with more than 400 organizations—including chambers of commerce, industry associations, development authorities, and other civic and nonprofit partners—to field the SBCS via an online questionnaire. The sampling frame consists of

[3]Throughout the paper, we use the terms "fintech lenders" and "online lenders" interchangeably.

[4]See https://www.fedsmallbusiness.org for more information.

businesses on the membership list or registry of partner organizations and is, therefore, a convenience sample. Across each participating Federal Reserve district, businesses receive an e-mail from partner organizations on behalf of the respective Federal Reserve Bank requesting their participation and providing an online link to the survey. Response rates for each partner organization are tracked in real time, and partners with initially low response rates may be encouraged to send out additional e-mails to businesses on their distribution lists until the survey officially closes. In total, responses were collected from 6,614 firms in 2018; 8,169 firms in 2017; and 10,303 firms in 2016 across all 50 states and the District of Columbia.

Unweighted, the SBCS sample is likely to reflect the firms favored by the Federal Reserve's collection process. For example, given that the sampling frame primarily consists of distribution lists of chambers of commerce and industry associations—organizations less likely to be connected to younger, less established firms—it is reasonable to expect that such firms would be underrepresented in the SBCS sample. In order to correct for gross sampling deviations from population data, the Federal Reserve uses a ratio-adjustment weighting method and demographic data on firm age, employee size, and industry to make the sample more representative of the population distribution of firms.[5] Age-of-firm data come from the Census Bureau's Business Dynamics Statistics. Industry and employee size data are from County Business Patterns.

3. Adoption of the Fintech Alternative to Banks

There is no question that fintech lenders are increasingly active in small business finance, but financial regulators need to know whether that activity is expanding access to credit for small businesses. Treasury officials noted in a recent report on nonbank financials, fintech, and innovation (U.S. Department of the Treasury 2018) that the use of alternative models and data sources could expand credit availability particularly for consumers and businesses that might be constrained by traditional credit-scoring models, an observation echoed in a 2019 interagency statement from the five federal financial

[5] Most econometric studies instead weight by an observation's inverse probability of selection. The SBCS poses certain limitations in this regard.

regulators.[6] However, identifying when fintech loans are expanding credit and when they are just substituting for banks and other credit providers has not been previously quantified in this market. In the context of consumer loans, Jagtiani and Lemieux (2018) show that while there are substantive differences between LendingClub's borrowers and those of traditional lenders (suggesting that LendingClub is penetrating potentially undeserved areas), the average FICO score of LendingClub's borrowers "is only very slightly below the average of overall Equifax customers." Jagtiani and Lemieux (2018) interpret this as evidence that much of the expansion might be substantially drawn from firms that previously borrowed or could borrow from traditional banks.

We use information available in the SBCS on the businesses that received financing from an online lender to compare the characteristics of these businesses with those of businesses that received bank loans and those of businesses that were denied financing. In simple comparisons, online borrowers are on average younger firms with fewer employees and less revenue (table 1). A larger proportion of firms operating at a loss also tend to turn to online lenders compared with firms receiving loans from traditional lenders, as do a larger proportion of minority-, women-, and veteran-owned businesses. In terms of industry (though not reported in table 1), firms in health care, administrative services, and retail are the most likely customers for fintech loans. The differences support the argument that online lenders reach groups that are less likely to be served by banks, but these firm characteristics are correlated with each other, so a model is needed to evaluate the relative importance of these factors on the type of financing received, if any.

3.1 Which Businesses Receive Which Financing?

We do not observe the specific factors which banks or online lenders use in their lending decisions, but any of the business characteristics identified in table 1 could be a factor in those decisions. At the same

[6]See "CA Letter 19-11 Interagency Statement on the Use of Alternative Data in Credit Underwriting" at https://www.federalreserve.gov/supervisionreg/caletters/caltr1911.htm.

Table 1. Basic Weighted Sample Characteristics, Survey Years 2016–18

	Denied Financing	Online Lender	Bank/CU Financing
Age			
0–2 Years	24.4	15.6	15.5
3–5 Years	18.8	22.1	12.8
6–10 Years	23.9	27.0	21.3
11–15 Years	13.1	15.9	14.3
16–20 Years	6.0	7.7	10.2
21+ Years	13.8	11.7	25.9
Employer Size			
1–4 Employees	59.1	54.4	37.0
5–9 Employees	20.7	22.6	19.7
10–19 Employees	10.4	13.0	18.2
20–49 Employees	6.9	7.8	14.6
50–499 Employees	2.9	2.2	10.5
Revenue			
< \$100K	25.1	12.2	9.9
\$100K–\$1M	53.6	64.7	42.1
\$1M–\$10M	19.9	21.9	39.2
\$10M+	1.4	1.2	8.7
Profitability			
At a Loss	38.7	35.6	22.4
Break Even	25.2	21.2	16.0
At a Profit	36.1	43.2	61.6
Minority-Owned Business			
Non-minority	74.2	79.2	83.9
Minority	25.8	20.8	16.1
Female-Owned Business			
Male	74.6	79.2	80.9
Female	16.1	17.7	14.6
Did Not Respond	9.3	3.0	4.5
Veteran-Owned Business			
Non-veteran	67.5	72.9	76.1
Veteran	11.5	15.0	10.2
Did Not Respond	21.0	12.1	13.7
Unemployment Rate (Change), 2015–16			
Mean	−0.447	−0.443	−0.403
Unemployment Rate (Change), 2016–17			
Mean	−0.514	−0.510	−0.516
Unemployment Rate (Change), 2017–18			
Mean	−0.471	−0.464	−0.435
N	1,376	1,004	4,904

Notes: Sample characteristics represent the percentage of survey respondents in each treatment group, except for the unemployment rate variables which represent the average change in the state unemployment rate for the state in which a firm is located during the noted time period. Of the firms in the Bank/CU financing treatment group, 164 were also approved for financing by a nonbank online lender after their approval by a bank lender. Of the firms in the Online financing group, 225 were also approved by a bank or credit union after their approval by an online lender.

time, correlations between firm characteristics may result in indirect associations of outcomes with observed characteristics that are not actually the factors used to make lending decisions. We apply a multinomial logit model to identify the factors with the greatest impacts on the funding outcomes of the small businesses that applied for financing. We specify a firm's financing status as a function of its size (in terms of employees), age, industry, revenue, profitability, credit risk status, and the demographic variables minority owned, woman owned, and/or veteran owned with all covariates specified as categorical variables around conventional cutoffs. In addition, we include controls for changes in state unemployment rates to account for local economic conditions.

The multinomial logit model implies that the probability of an outcome, also known as the propensity score, is

$$P\left(w=1|x_i\right)=\frac{e^{X_i\beta_1}}{1-\sum_{o=1}^{O-1}e^{X_i\beta_o}}.$$

The sum of the probabilities of all outcomes w is equal to 1 by construction. In our estimation, financing outcomes are online, bank or credit union, and denied: $w_i = O, B$, or D.

Table 2 shows the average marginal effects of the key variables.[7] Average marginal effects are measured as the difference in propensity scores for a predicted outcome ($w = O$) for a particular variable ($z = 1$) versus ($z = 0$), averaging across all observations of other variables x regardless of the realized outcome of the observations:

$$\begin{aligned}AME\left(w=O,z=1\right)=\sum_{n=0}^{N}&\left(P(w=O|z=1,x_n)\right.\\&\left.-P\left(w=O|z=0,x_n\right)\right)/N.\end{aligned}$$

Because the sample is composed of all businesses applying for credit regardless of outcome, it represents the average effect of a categorical variable for an otherwise typical business applying for

[7]The multinomial logit model's full results are shown in appendix table A.1. The samples vary some based on the outcome questions. We include the largest possible sample for each outcome, so there are four similar but not identical logit models shown in table A.1.

Table 2. Average Marginal Effects of Key Variables on Receiving Financing, Survey Years 2016–18

	Denied Financing	Online Lender	Bank/CU Financing
Age			
0–2 Years	0.026	−0.054***	0.029
	(0.018)	(0.015)	(0.020)
3–5 Years	0.017	0.051***	−0.067***
	(0.016)	(0.017)	(0.019)
6–10 Years	0.002	0.028*	−0.030*
	(0.014)	(0.014)	(0.016)
11–15 Years	0.001	0.038**	−0.038**
	(0.018)	(0.019)	(0.019)
16–20 Years	−0.041*	−0.001	0.042
	(0.021)	(0.024)	(0.026)
21+ Years	−0.019	−0.049***	0.068***
	(0.015)	(0.013)	(0.016)
Employees	−0.001**	−0.001*	0.002***
	(0.001)	(0.001)	(0.001)
Profitable	−0.044***	−0.019***	0.063***
	(0.007)	(0.007)	(0.008)
Revenue > $1M	−0.052***	−0.036***	0.088***
	(0.011)	(0.011)	(0.013)
Minority-Owned Firm	0.035**	0.001	−0.037*
	(0.017)	(0.015)	(0.019)
Woman-Owned Firm	−0.024*	0.012	0.012
	(0.014)	(0.014)	(0.017)
Veteran-Owned Firm	−0.015	0.056**	−0.041*
	(0.020)	(0.024)	(0.024)
Medium/High Credit Risk	0.057***	0.052***	−0.109***
	(0.008)	(0.008)	(0.009)
Unemployment Rate (Change),	−0.053***	−0.036*	0.089***
2015–16	(0.020)	(0.019)	(0.022)
Unemployment Rate (Change),	0.011	0.027	−0.038
2016–17	(0.028)	(0.024)	(0.030)
Unemployment Rate (Change),	−0.064**	−0.030	0.093***
2017–18	(0.027)	(0.027)	(0.030)
Year			
2016	0.007	−0.058***	0.051***
	(0.010)	(0.009)	(0.011)
2017	0.004	−0.002	−0.002
	(0.011)	(0.011)	(0.013)
2018	−0.011	0.062***	−0.051***
	(0.010)	(0.011)	(0.012)

Notes: Standard errors are in parentheses. ***, **, and * denote significance at $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively. Employee and unemployment rate variables are continuous; all other variables are discrete. Credit risk is determined by the self-reported business credit score or personal credit score, depending on which is used to obtain financing for their business. If the firm uses both, the higher risk rating is used. Low credit risk is an 80–100 business credit score or a 720+ personal credit score. Medium credit risk is a 50–79 business credit score or a 620–719 personal credit score. High credit risk is a 1–49 business credit score or a <620 personal credit score. For full results of multinomial logit estimates, see table A.1.

credit. The average marginal effects also net to zero across rows because the columns represent the full set of options.

The borrowing outcomes of small businesses do depend on a range of characteristics, but not necessarily monotonically. The effect of a business being in one of the younger age categories (firm age between 3 and 15 years) is to boost the likelihood of receiving credit from an online lender and lower the likelihood of bank financing. In contrast, most age groups of firms are not statistically distinguishable for being denied financing, with statistically significant results only for firms between 16 and 20 years old (−4 percentage points). Those in the oldest age category of small businesses, 21+ years, are most likely to receive bank financing (7 percentage points).

Increased employee counts (included as a continuous variable and its square) make bank financing statistically more likely, with similar reductions in being denied financing or the use of online financing. The negative coefficient on the squared term of employment size (table A.1) implies that these effects diminish as firms grow. That said, for most of the firm sizes in our sample, these effects are not that large: Going from 1 employee to 10 employees increases the likelihood of bank financing by about 2 percentage points and lowers the likelihood of online financing by 1 percentage point.

The profitability of businesses is a critical factor for banks, boosting the likelihood of bank financing by about 6 percentage points. That higher probability of bank lending is mirrored by lower likelihoods of both denials (−4 percentage points) and online-lender financing (−2 percentage points) for profitable firms. The coefficients imply that online-lender financing is more likely for unprofitable firms, all else held constant. Even accounting for profitability, higher-revenue firms are 9 percentage points more likely to receive bank financing, with most of the offsetting probability coming from denials. Finally, being evaluated by a credit bureau as medium or high risk substantially lowers the likelihood of bank financing (by 11 percentage points) and evenly raises the likelihood of both denial and online-lender financing. These key financial variables clearly help to determine which firms receive which financing outcomes.

The demographic characteristics of the heads of businesses are relatively less influential on the outcomes, but there are still some statistically significant differences after accounting for the other variables. Minority status lowers the likelihood of bank financing by roughly 4 percentage points, with the associated higher frequency being in denials. Women-owned businesses have a lower likelihood of being denied financing, while veteran-owned businesses are more likely to receive online financing with an associated lower probability of bank financing.

We included the change in state unemployment rates to account for (generally) improving market conditions on lending outcomes. Banks seem less likely to lend in areas where the unemployment rate is declining (with associated higher levels of denials), but the changes are relatively small in most of this period, a finding that suggests a relatively small role for local economic conditions in the determination of individual lending outcomes.

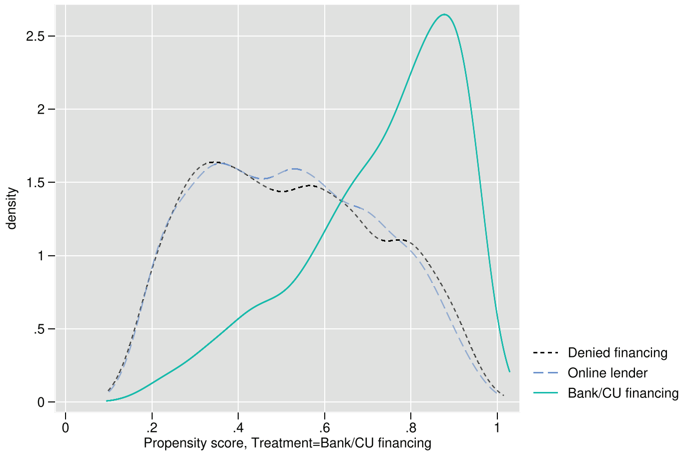
Finally, we included year dummy variables to account for other changes over time. This variable seems to primarily pick up the relative rise in online lending relative to bank lending. All else equal, the outcome of getting online financing is 12 percentage points more likely in 2018 than it was in 2016, with most of that effect being accounted for by offsetting reductions in the likelihood of being a bank borrower.

3.2 Are Online Lenders Expanding the Financing Options of Small Businesses?

The substantial differences seen in the probabilities reported in table 2 motivate the importance of the controls and the value of a model to assess lending decisions by banks and online lenders. We can use the associated propensity scores to evaluate the proportion of online-lender financing that could be substituting for bank financing rather than representing a new source of business financing. The relevant comparison uses the propensity of borrowers to receive bank financing given the full set of characteristics of each small business[8]:

[8]We group the financing received from large and small banks with credit union financing into the category of traditional financing. Credit unions remain a smaller actor in small business financing but are important enough to include: 8 percent of our businesses seeking financing received their first financing from a credit union.

Figure 1. Kernel Density ("overlap") Plots, Survey Years 2016–18



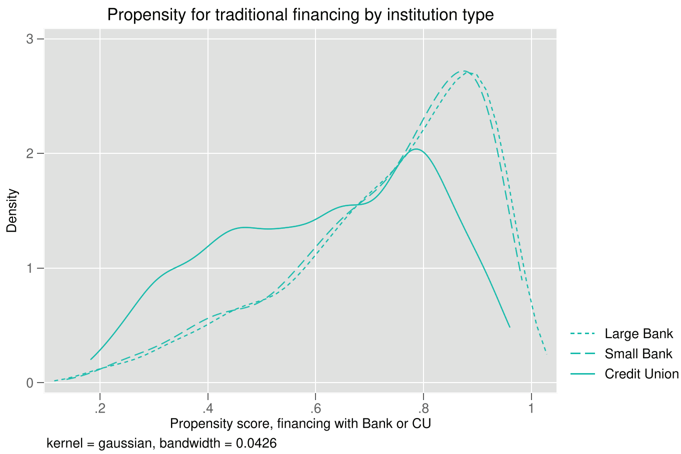
Notes: Predicted probabilities of being approved for bank/credit union financing shown for each treatment group. For full results of multinomial logit estimates, see table A.1.

$P(w = B|x_n)$. These propensities can then be compared for businesses that received online financing, those that received financing from banks, and those rejected for financing (figure 1).[9]

Not surprisingly, the majority of businesses that actually received financing from either large or small banks have propensity scores for traditional financing of above 0.70; the median propensity score for a business that received traditional financing is 0.77. In contrast, online lenders appear substantially more likely to provide credit to firms that the model expects to be denied credit. The median propensity score for businesses that use online-lender financing is 0.51, which is identical to the median propensity score of businesses that were denied credit. This means that half of those either using online financing or being denied financing were evaluated by the model as being in a region of characteristics where bank financing is uncommon.

[9]The estimates are smoothed by a Gaussian kernel density estimator to deemphasize small differences in estimated propensities that particularly appear when the model includes discrete variables.

Figure 2. Kernel Density Plots, Survey Years 2016–18



Notes: Predicted probabilities of being approved for bank/credit union financing shown for firms actually approved by a small bank, large bank, or credit union. For full results of multinomial logit estimates, see table A.1.

To formalize this point, we construct a measure of added lending activity (A) associated with the existence of online lenders. It sums the excess mass of the online lender outcome, whenever the density for online lenders is higher than traditional lenders:

$$A = \sum \left(f_{w=O}\left(z_d\right) - f_{w=B}\left(z_d\right)\right) \cdot I\left(f_{w=O}\left(z_d\right) > f_{w=B}\left(z_d\right)\right),$$

where $z_d(x) = P\left(w = B|x_d\right)$ and the densities, f, are estimated using a kernel density procedure. The summation can then be applied across the full data set. For the period of 2016 to 2018, we would estimate that 44 percent of businesses served by online lenders look unlikely to have been served by banks. This is a conservative estimate of the extra firms financed, because the entry and expansion of online lenders has likely also drawn in more businesses to apply for financing than would have been the case without the new option.

For figure 1 we grouped all of the existing traditional financing options together, but given the long-standing research on the roles of small banks and the relatively recent entry of credit unions into small business finance, it is worthwhile to compare these lenders. Figure 2 shows the densities of propensity scores for traditional financing by the type of institution that provided each business's first financing. This comparison is offered as a way to assess whether the banking options are similar. It is the case that small and large banks are essentially equally likely to provide financing at any given level of the propensity score. Figure 2 does reveal that credit unions more frequently lend to businesses with a lower propensity score for traditional financing. That said, the difference between these categories of lenders is much smaller than the difference between traditional financing and online lending.

4. Using Treatment Effects to Evaluate Financial Alternatives

The expansion of credit to small businesses is an important question, but policymakers and regulators are also interested in whether a credit source is beneficial and appropriate for the borrower. This is a hard assessment to make in the best of circumstances because we observe only one set of outcomes per firm, so the outcomes associated with a counterfactual funding alternative are never observed. Complicating matters is the fact that many small businesses have reasonably high rates of failure, regardless of whether they have borrowed or not. The SBCS does not follow firms, so we cannot measure failures or defaults, but it does include the businesses' assessments for revenue growth, employment growth, and satisfaction with financing after the lending outcome. Table 3 shows business expectations with no controls applied other than weighting to match population statistics. Without compositional controls, firms that received online financing have the most positive expectations about future firm growth for revenue, while firms that were denied financing had the strongest outlook for employment growth. This could be evidence of the value of online financing, but it could also reflect the role of sorting based on the age of the firm: younger (and riskier) firms expect more growth and are more willing to use online financing.

Differences in satisfaction levels across treatment groups are much more pronounced, with only 5.3 percent of firms that were denied financing being satisfied with their lender(s) compared with 37.7 percent among firms approved by fintech lenders, and 69.6 percent among firms approved by traditional bank lenders. These

Table 3. Treatment Group Comparison, Survey Years 2016–18

	Denied Financing	**Online Lender**	**Bank/CU Financing**
Outcomes of Interest			
Expects Future Revenue Growth (%)	75.8	76.9	73.2
N	1,376	1,004	4,904
Expects Future Employment Growth (%)	52.9	52.1	50.7
N	1,343	990	4,829
Satisfied with Lender (%)	5.3	37.7	69.6
N	1,243	1,001	4,873

Notes: Respondents are asked in separate questions how they expect revenue and the number of employees to change over the next 12 months with the option to select "Decrease," "No Change," or "Increase." Comparisons of each outcome of interest represent the percentage of respondents who selected "Increase." Of the firms in the Bank/CU financing treatment group, 164 were also approved for financing by a non-bank online lender after their approval by a bank lender. Of the firms in the Online financing group, 225 were also approved by a bank or credit union after their approval by an online lender.

differences are large, but again we should be concerned about the compositional differences.

4.1 Treatment Effects Estimators

Ideally, we would like to observe the counterfactual scenarios of each firm, that is to say, what the expectations of a firm denied financing would have been if it had been approved by an online lender and likewise if it had been approved by a traditional lender. However, by construction, we will never see all three financing treatments for the same owner because they are mutually exclusive. Furthermore, our data are not the product of a large-scale randomized experiment, which could make other important characteristics of the owner or firm asymptotically irrelevant. These weaknesses imply that confounding variation (like the age and profitability of the business) could affect the likelihood of observing a given financing treatment and, potentially, the outcomes of interest given a financing treatment.

To address these issues we apply semiparametrically estimated treatment effects given the likelihood that firms with specific characteristics are provided financing $w_i = O$, B, or D. Specifically, we will estimate potential-outcome means for all firms regardless of outcome, for receiving online financing $(E[Y_i | w_i = O])$, for receiving bank financing $(E[Y_i | w_i = B])$, and for seeking financing but being denied $(E[Y_i | w_i = D])$. Using these terms, we can evaluate an average treatment effect for online financing as $ATE(O) = E[Y_i | w_i = O] - E[Y_i | w_i = D]$ along with a parallel estimate for traditional bank financing, $ATE(B) = E[Y_i | w_i = B] - E[Y_i | w_i = D]$. Finally, we can also construct a relative treatment effect of online financing relative to bank financing: $RTE(O, B) = E[Y_i | w_i = O] - E[Y_i | w_i = B]$.

In our analysis we estimate these values using inverse probability weighting (IPW) and inverse-probability-weighted regression adjustment (IPWRA) as described in Imbens (2004) and Wooldridge (2015). IPW is simply the sample average of the outcome weighting by $\hat{p}(w, x_i)$ the estimated probability that observation i experiences treatment W:

$$\hat{\mu}(W) = N^{-1} \sum_{i=1}^{N} \frac{I(w_i = W)Y_i)}{\hat{p}(w, x_i)},$$

where I() is an indicator function.

Weighting by the inverse of the propensity for an outcome, w, given x_i, balances the observations across the full range of characteristics regardless of outcome. In our case, $\hat{p}(w, x_i)$ is implemented by the simple multinomial logit model discussed previously. An advantage of IPW is that assumptions about the nature of the outcomes with respect to covariates are limited, given an effective model of the probability of treatment.

IPWRA combines this weighting with regression-based adjustment for differences in outcomes based on the set of characteristics x_i solving the following minimization:

$$\hat{\mu}(W) = \min_{\alpha_1, \beta_1} \sum_{i=1}^{N} \frac{I(w_i = W)(Y_i - \alpha_1 - \beta_1 x_{i1}))^2}{\hat{p}(w, x_i)}.$$

While there is no particular justification for different control variables in the two steps, x_i and x_{i1} need not be identical. The IPWRA is a “doubly robust technique” in that it is asymptotically unbiased if either the model of treatment probabilities or the model of conditional means is correct (Wooldridge 2015).

Importantly, regardless of the estimation technique, reliable estimates of these values rely on two assumptions: (i) *unconfoundedness*, or conditional independence, which requires that treatment assignment be independent of the treatment effect when conditioned on appropriate control variables, and (ii) *overlap of the treatments*, which requires that the probability of observing a treatment value must be greater than 0 for all relevant x.

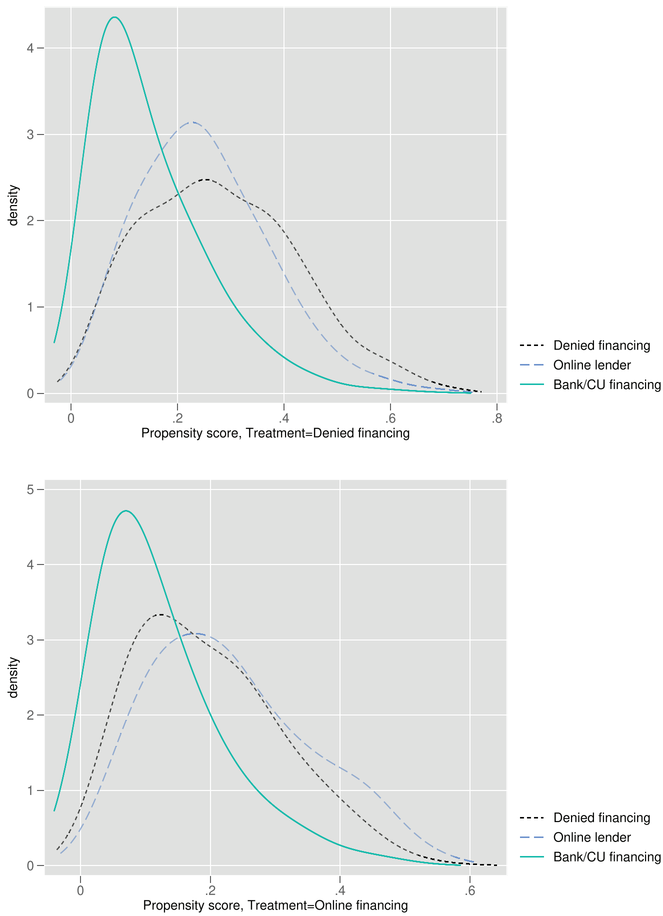
In the case of small business lending, firm-specific variables that are likely to alter the approval of loans are key controls that are likely to satisfy assumption (i). We intentionally included all reasonable variables available in the SBCS including revenue, profitability, age of firm, and the demographic characteristics of the business owner. These variables should inform predictions of financing approval and were shown in table 2 to be important factors.

4.2 Overlap of Treatments

For the measurement of the businesses’ response to the two lending treatments, it is important to confirm that there are relevant observations to compare according to the treatment model. The fundamental issue is that if online borrowers were always riskier than any observed bank borrower, then it would require strong assumptions to estimate what their outcomes would have been had they received a bank loan. A lack of overlap makes it particularly difficult to reliably predict the counterfactual scenarios that are needed to obtain accurate treatment effects.

The plot in figure 1, while informative about the expansion of credit, is called an overlap plot in the treatment effects literature. It shows the distribution of predicted probabilities of receiving each financing treatment and of denial for firms according to their propensity to receive bank and credit union financing. From an overlap perspective, we want to see that there are observations experiencing each outcome for any given propensity of bank and credit union

Figure 3. Kernel Density ("overlap") Plots, Survey Years 2016–18



Notes: Predicted probabilities of being denied financing and receiving online financing, respectively, shown for each treatment group. For overlap plot of receiving bank/credit union financing, see figure 1. For full results of multinomial logit estimates, see table A.1.

financing. This is generally the case, with the only possible exceptions coming at the far tails of the densities, when none of the outcomes are likely. This is excellent for being able to estimate treatment effects across the full range of firms in the data. Figure 3 completes the set of overlap plots, by showing the plots based on propensities to receive online financing and to be denied financing. The plot on the bottom displays the estimated density of the predicted probabilities for receiving online financing. The plot on the top shows the propensity of denial for the different treatment outcomes. There is again substantial overlap through much of the distribution, although bank borrowers crowd to the left (low online or denial probability) in figure 3, making conclusions about riskier borrowers less robust. Importantly, while profitability, revenues, and so on have a very strong effect on financing treatment, the observed firms do not have most of their mass at opposite ends of the distribution—but rather each example appears to have substantial overlapping cases for each treatment.

5. Effects of Banking Alternatives on Firm Outcomes

5.1 Loan Size Differences

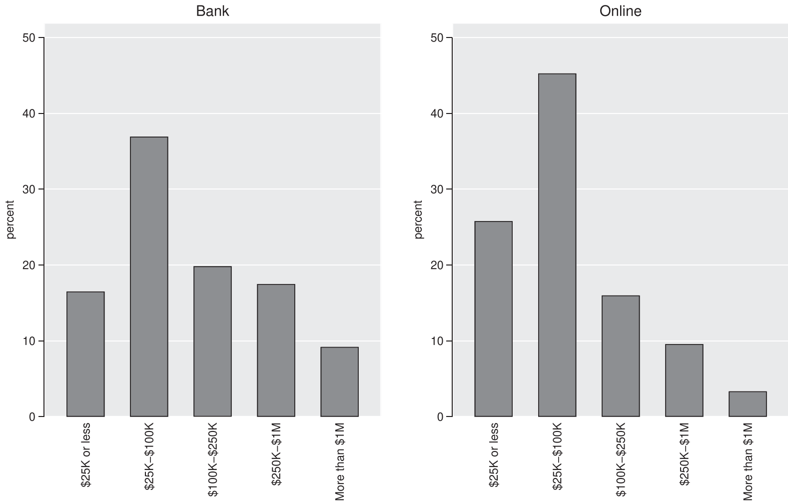
An important difference in alternative lending channels is the size of the loan offered. In order to support a higher survey response rate, the SBCS asks for loan amounts in terms of five bins. The loan application amounts are clearly lower for online loans than for bank loans, but again this could reflect firm differences rather than any difference in the treatment channel.

To counter the tendency for firm characteristics to distort the lender differences, we applied inverse probability weighting to the histograms to produce an estimate of the loan size distribution once the composition is accounted for. Figure 4 shows that after compositional adjustments, applicants at online lenders still make smaller requests, with more than 70 percent of loan applications requesting less than $100,000 versus roughly 56 percent of adjusted loan applications with traditional lenders.

5.2 Revenue and Employment Growth

Businesses typically can use loan proceeds to make capital purchases to support operations, so we should expect approved businesses to anticipate revenue growth and potentially employment growth, although the unobserved terms of the financing may also hinder the

Figure 4. Distribution of Loan Size after Inverse Probability Weighting



growth of firms. Future revenue growth and capital expenditures are measured by the owner's short-term expectations (next 12 months); while not ex post, these measures may show differences in likely outcomes as a result of the financing channel chosen.

In table 4, we report the composition-adjusted potential-outcome mean for being denied financing and then the treatment effects for receiving online or bank financing, followed by the relative treatment effect between online and bank financing. First it is worth noting that regardless of the estimator, the majority of the composition-balanced businesses (75.2 percent) expect revenue and employment growth even if they were denied financing. The results indicate that there is no statistically significant difference in expected revenue growth for either bank or online financing options relative to being denied financing. However, the difference between online and bank financing on revenue and employment growth are statistically significant in all cases.

We might have anticipated online loans being less effective than bank loans either because they are smaller or because their terms might differ unfavorably, but this conclusion is rejected in our analysis. Still, the estimated impact of fintech financing on a firm's

Table 4. Likelihood of Reporting Future Firm Growth or Satisfaction with Lender, by Model Specification and Treatment Group, Survey Years 2016–18

	Potential-Outcome Mean	Average Treatment Effect		
	Denied Financing	**Online vs. Denied**	**Bank/CU vs. Denied**	**Bank/CU vs. Online**
Expects Future Revenue Growth				
IPW	0.752	0.029	−0.013	−0.043**
		(0.028)	(0.023)	(0.021)
IPWRA	0.750	0.032	−0.010	−0.041**
		(0.024)	(0.019)	(0.021)
Expects Future Employment Growth				
IPW	0.527	0.036	−0.018	−0.054**
		(0.032)	(0.026)	(0.024)
IPWRA	0.518	0.043	−0.009	−0.052**
		(0.030)	(0.023)	(0.025)
Satisfied with Lender				
IPW	0.053	0.360***	0.619***	0.259***
		(0.026)	(0.015)	(0.027)
IPWRA	0.055	0.355***	0.616***	0.261***
		(0.025)	(0.015)	(0.026)

Notes: Standard errors are in parentheses. ***, **, and * denote significance at $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively. Of the firms in the Bank/CU financing treatment group, 164 were also approved for financing by a nonbank online lender after their approval by a bank lender. Of the firms in the Online financing group, 225 were also approved by a bank or credit union after their approval by an online lender.

self-reported business outlook in table 4 is somewhat ambiguous, in that firms in the bank and online treatment groups do not perform statistically differently from firms that were denied financing.

5.3 Satisfaction with the Lending Experience

The SBCS asks firms whether they are satisfied, dissatisfied, or neutral with regard to the lender applied to. Respondents are specifically prompted as they answer the question to consider the application process as well as terms of repayment for lenders that approved their application. The descriptive statistics shown in table 3 reveal that there are significant differences in satisfaction levels with the type of lender businesses used, but this result could also be substantially affected by the characteristics of the treated samples.

After IPW adjustments for composition, just 5.3 percent of applicants for credit are satisfied after a financing denial. Adjusted satisfaction levels are higher for online lenders, with a treatment effect of 36 percentage points, which is statistically different from the denial outcome. Bank financing results in a treatment effect on satisfaction of 61.9 percentage points, which is again statistically significant. Thus the difference after compositional adjustments between satisfaction with online lenders and banks is 25.9 percentage points, with firms more likely to be satisfied with bank lender(s) than with online financing. The same qualitative results are maintained when the IPWRA procedure is applied.

These results suggest room for improvement for online lenders in their customer satisfaction levels. To further investigate where this difference comes from, the SBCS includes an identification of the type of online lender in 2017 and 2018. Table 5 shows the breakdown of satisfaction rates by type of online lender. We neither adjust for composition nor calculate standard errors given the smaller numbers of survey respondents, but merchant cash advance lenders stand out for their relatively low satisfaction figures. That said, average satisfaction rates for all types of online lenders are still below the bank average of 69.6 percent (unadjusted, from table 3).

The 2017 and 2018 surveys also follow up with a question on challenges experienced during the application process. Table 6 shows that the top three challenges reported by businesses applying for

Table 5. Types of Online Lenders Applied to by Applicants in Online Treatment Group, Survey Years 2017–18

	# of Applicants	% of Applicants	% of Applicants Satisfied
Direct Lender	360	57.9	41.9
Retail/Payments Processor	90	14.5	45.6
Peer-to-Peer Lender	58	9.3	39.7
Merchant Cash Advance Lender	87	14.0	26.7
Other	28	4.5	53.6

Notes: Frequency counts and percentages are unweighted. For a survey respondent's two most recent credit applications—if one or both applications were with an online lender—the respondent is asked: *Which type of online lender did you apply to?* The question was not included in the 2016 survey. Percentages in column 2 do not add to 100 because firms were only asked the given question if their application was among their two most recent applications. "Direct Lender" includes OnDeck, Kabbage, Blue Vine, etc.; "Retail/Payments Processor" includes Paypal Working Capital, Square Capital, Amazon Capital Services, etc.; "Peer-to-Peer Lender" includes LendingClub, Funding Circle, etc.; "Merchant Cash Advance Lender" includes RapidAdvance, CAN Capital, BizFi, etc.

Table 6. Challenges Experienced during Application Process, Survey Years 2017–18

	Online Treatment Group		Bank/CU Treatment Group	
	# of Applicants	% of Applicants	# of Applicants	% of Applicants
High Interest Rate	204	32.8	128	4.8
Unfavorable Repayment Terms	118	19.0	53	2.0
Long Wait for Decision	28	4.5	161	6.1
Difficult Application Process	29	4.7	124	4.7
Lack of Transparency	32	5.1	35	1.3
Other Challenges	15	2.4	81	3.1
Experienced No Challenges	114	18.3	745	28.2

Notes: Frequency counts and percentages are unweighted. For a survey respondent's two most recent credit applications, the respondent is asked: *Did your business experience any challenges in applying for the [given product]? Select all that apply.* The question was not included in the 2016 survey. Percentages in columns 2 and 4 do not add to 100 because firms were only asked the given question if their application was among their two most recent applications.

online loans are high interest rates (32.8 percent), unfavorable repayment terms (19 percent), and lack of transparency (5.1 percent). Challenges for bank borrowers are all lower, but their top three challenges are the long wait for decision (6.1 percent), high interest rates (4.9 percent), and the difficult application process (4.7 percent).

6. Conclusion

While there are still many open questions about the value and effects of online business lending, particularly in the long run, our results based on Small Business Credit Survey data provide some useful insights into this expanding sector of the financial market. One important finding is that the businesses that pursue bank or online options or are denied credit are not equivalent entities. Thus, to accurately compare the lending outcomes of these businesses, adjustments have to be made to account for compositional differences. We use a treatment effects approach, which, although it cannot solve underlying sampling defects, can help to evaluate the role of different lending outcomes when the characteristics of firms vary substantially between those outcomes.

The 2018 Treasury report notes the potential for fintech to expand credit "to borrower segments that may not otherwise have access to credit through traditional underwriting approaches." But the Treasury report is able to provide little evidence to support this conjecture. We show that the entry of online lenders has meaningfully altered the range of firms that receive financing, with 44 percent of online borrowers not likely to receive credit from traditional sources. Overall, our evidence suggests that the characteristics of online borrowers are closer to those of businesses rejected for credit than those served by banks, which increases the financing available in the small business financing marketplace.

On the effectiveness of online credit, we find that growth expectations from online lenders are better than those for bank borrowers. This is despite controlling for compositional differences that are strongly predictive of which firms receive credit from banks and from fintech firms, including profitability, revenue growth, and self-reported credit scores of the business or owner. This result is supportive of the position that financial innovation, at least in this case, has been beneficial to borrowers, particularly when combined with the greater financial inclusion shown by fintech lenders.

While the effects on expectations for growth are relatively small, the ordering of customer satisfaction across lender types is clear: bank borrowers are more satisfied than online borrowers, who are more satisfied than businesses that were denied credit. This may point to issues that both the lenders and regulators may want to address as online lending continues to expand.

Table A.1. Multinomial Logit Regressions for Probability of Receiving Financing (i.e., the treatment models used as inputs into outcome models)[a]

	Revenue Model	Employment Model	Satisfaction Model	Application Amount Model
Online Lender				
Employees (Continuous)	−0.000	0.002	−0.001	−0.001
	(0.005)	(0.006)	(0.006)	(0.006)
Employees Squared (Continuous)	−0.000	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Age				
3–5 Years	0.649**	0.611**	0.650**	0.646**
	(0.209)	(0.211)	(0.213)	(0.212)
6–10 Years	0.612**	0.611**	0.660**	0.641**
	(0.212)	(0.214)	(0.216)	(0.216)
11–15 Years	0.664**	0.606*	0.733**	0.726**
	(0.236)	(0.240)	(0.241)	(0.241)
16–20 Years	0.690*	0.699*	0.691*	0.676*
	(0.282)	(0.288)	(0.285)	(0.285)
21+ Years	0.255	0.251	0.278	0.291
	(0.232)	(0.235)	(0.235)	(0.235)
Revenue Size				
$1M+	0.085	0.090	0.056	0.057
	(0.166)	(0.168)	(0.170)	(0.171)
Profitability				
Profitable	0.211	0.186	0.252	0.267*
	(0.131)	(0.133)	(0.134)	(0.135)
Minority-Owned Business				
Minority	−0.190	−0.193	−0.233	−0.231
	(0.161)	(0.163)	(0.164)	(0.164)

(*continued*)

Table A.1. (Continued)

	Revenue Model	**Employment Model**	**Satisfaction Model**	**Application Amount Model**
Female-Owned Business				
Female	0.170	0.174	0.208	0.191
	(0.146)	(0.146)	(0.148)	(0.148)
Did Not Respond	-0.976^{***}	-0.983^{***}	−0.420	−0.435
	(0.254)	(0.262)	(0.277)	(0.277)
Veteran-Owned Business				
Veteran	0.339	0.353	0.317	0.316
	(0.204)	(0.206)	(0.205)	(0.205)
Did Not Respond	−0.340	−0.324	−0.285	−0.281
	(0.182)	(0.186)	(0.186)	(0.187)
Medium/High Credit Risk	0.035	0.041	0.072	0.077
	(0.129)	(0.131)	(0.131)	(0.131)
Revenue Growth in Past 12 Months				
Increased	0.017	0.048	0.037	0.033
	(0.132)	(0.133)	(0.134)	(0.134)
Change in Unemployment Rate (Continuous)				
2015–16	0.050	0.006	0.073	0.072
	(0.169)	(0.170)	(0.172)	(0.172)
2016–17	0.097	0.096	0.035	0.040
	(0.223)	(0.225)	(0.228)	(0.229)
2017–18	0.138	0.041	0.098	0.126
	(0.236)	(0.238)	(0.240)	(0.239)
Survey Year Dummy				
2017	0.390^{*}	0.395^{*}	0.362^{*}	0.367^{*}
	(0.158)	(0.160)	(0.161)	(0.161)
2018	0.770^{***}	0.780^{***}	0.772^{***}	0.772^{***}
	(0.152)	(0.153)	(0.155)	(0.155)
Constant	-0.954^{***}	-1.024^{***}	-1.023^{***}	-1.007^{***}
	(0.279)	(0.283)	(0.284)	(0.284)

(*continued*)

Table A.1. (Continued)

	Revenue Model	**Employment Model**	**Satisfaction Model**	**Application Amount Model**
Bank/CU				
Employees (Continuous)	0.011**	0.012**	0.011**	0.011**
	(0.004)	(0.004)	(0.004)	(0.004)
Employees Squared (Continuous)	−0.000**	−0.000**	−0.000**	−0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Age				
3–5 Years	−0.196	−0.217	−0.211	−0.232
	(0.169)	(0.171)	(0.173)	(0.173)
6–10 Years	−0.032	−0.037	0.005	−0.012
	(0.170)	(0.173)	(0.175)	(0.175)
11–15 Years	−0.049	−0.109	0.001	−0.001
	(0.184)	(0.186)	(0.190)	(0.191)
16–20 Years	0.368	0.387	0.354	0.350
	(0.221)	(0.227)	(0.225)	(0.225)
21+ Years	0.306	0.277	0.301	0.320
	(0.180)	(0.182)	(0.183)	(0.185)
Revenue Size				
$1M+	0.740***	0.735***	0.719***	0.718***
	(0.125)	(0.126)	(0.129)	(0.130)
Profitability				
Profitable	0.755***	0.759***	0.792***	0.801***
	(0.105)	(0.106)	(0.109)	(0.109)
Minority-Owned Business				
Minority	−0.313*	−0.318*	−0.358*	−0.357*
	(0.138)	(0.140)	(0.141)	(0.141)
Female-Owned Business				
Female	0.153	0.133	0.189	0.177
	(0.125)	(0.126)	(0.127)	(0.127)
Did Not Respond	−0.387*	−0.382	0.099	0.066
	(0.192)	(0.198)	(0.231)	(0.232)

(*continued*)

Table A.1. (Continued)

	Revenue Model	Employment Model	Satisfaction Model	Application Amount Model
Veteran-Owned Business				
Veteran	−0.081	−0.091	−0.138	−0.135
	(0.165)	(0.168)	(0.168)	(0.168)
Did Not Respond	−0.255	−0.261	−0.199	−0.195
	(0.145)	(0.147)	(0.151)	(0.152)
Medium/High Credit Risk	−0.951***	−0.935***	−0.890***	−0.890***
	(0.103)	(0.104)	(0.105)	(0.105)
Revenue Growth in Past 12 Months				
Increased	0.151	0.153	0.163	0.171
	(0.104)	(0.105)	(0.107)	(0.108)
Change in Unemployment Rate (Continuous)				
2015–16	0.475***	0.452**	0.505***	0.498***
	(0.137)	(0.139)	(0.142)	(0.142)
2016–17	−0.147	−0.171	−0.221	−0.215
	(0.194)	(0.195)	(0.200)	(0.201)
2017–18	0.538**	0.516**	0.468*	0.501**
	(0.185)	(0.186)	(0.189)	(0.188)
Survey Year Dummy				
2017	−0.112	−0.118	−0.120	−0.126
	(0.120)	(0.121)	(0.123)	(0.124)
2018	−0.165	−0.153	−0.149	−0.164
	(0.124)	(0.125)	(0.127)	(0.128)
Constant	1.195***	1.173***	1.110***	1.131***
	(0.219)	(0.221)	(0.224)	(0.224)

[a]Variable specification is identical for all treatment models, but coefficient estimates vary given that the sample size varies depending on the outcome question asked in the survey.

Notes: Coefficient estimates are relative to the base outcome of not receiving any financing. Standard errors are in parentheses. ***, **, and * denote significance at $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively. Credit risk is determined by the self-reported business credit score or personal credit score, depending on which is used to obtain financing for their business. If the firm uses both, the higher risk rating is used. Low credit risk is an 80–100 business credit score or a 720+ personal credit score. Medium credit risk is a 50–79 business credit score or a 620–719 personal credit score. High credit risk is a 1–49 business credit score or a <620 personal credit score.

References

Balyuk, T., A. N. Berger, and J. Hackney. 2020. "What is Fueling FinTech Lending? The Role of Banking Market Structure." Working Paper. http://dx.doi.org/10.2139/ssrn.3633907.

Basel Committee on Banking Supervision. 2018. *Sound Practices: Implications of Fintech Developments for Banks and Bank Supervisors.* Bank for International Settlements. https://www.bis.org/bcbs/publ/d431.htm.

Berger, A., N. Miller, M. Petersen, R. Rajan, and J. Stein. 2005. "Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks." *Journal of Financial Economics* 76 (2): 237–69.

Berger, A. N., and G. F. Udell. 2002. "Small Business Credit Availability and Relationship Lending: The Importance of Bank Organisational Structure." *Economic Journal* 112 (477): F32–F53.

Darden, K., N. Dixit, and T. Mason. 2018. "2018 US Fintech Market Report." S&P Global Market Intelligence. https://www.spglobal.com/marketintelligence/en/documents/2018-us-fintech-market-report.pdf.

DeYoung, R., S. Frame, D. Glennon, and P. Nigro. 2011. "The Information Revolution and Small Business Lending: The Missing Evidence." *Journal of Financial Services Research* 39 (1–2): 19–33.

DeYoung, R., D. Glennon, and P. Nigro. 2008. "Borrower–Lender Distance, Credit Scoring, and Loan Performance: Evidence from Informational-Opaque Small Business Borrowers." *Journal of Financial Intermediation* 17 (1): 113–43.

Federal Reserve. 2019. "2018 Small Business Credity Survey: Report on Employer Firms." Special Report.

———. 2020. "PPPLF Transaction-Specific Disclosures (September 8, 2020)." https://www.federalreserve.gov/monetarypolicy/ppplf.htm.

Goldstein, I., J. Jagtiani, and A. Klein. 2019. "FinTech and the New Financial Landscape." *Banking Perspectives* (Bank Policy Institute). https://www.bankingperspectives.com/fintech-and-the-new-financial-landscape/.

Imbens, G. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics* 86 (1): 4–29.

Jagtiani, J., and C. Lemieux. 2018. "Do Fintech Lenders Penetrate Areas that Are Underserved by Traditional Banks?" *Journal of Economics and Business* 100 (November-December): 43–54. https://doi.org/10.1016/j.jeconbus.2018.03.001.

———. 2019. "The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform." *Financial Management* 48 (4): 1009–29. https://doi.org/10.1111/fima.12295.

Mach, T. L., C. M. Carter, and C. R. Slattery. 2014. "Peer-to-Peer Lending to Small Businesses." Finance and Economics Discussion Series No. 2014-10, Board of Governors of the Federal Reserve System.

Philippon, T. 2018. "The FinTech Opportunity." Working Paper, Wharton Pension Research Council. https://repository.upenn.edu/prc_papers/1.

Robb, A. M., and D. T. Robinson. 2014. "The Capital Structure Decisions of New Firms." *Review of Financial Studies* 27 (1): 153–79. https://doi.org/10/1093/rfs/hhs072.

U.S. Department of the Treasury. 2016. "Opportunities and Challenges in Online Marketplace Lending." White Paper.

———. 2018. "A Financial System That Creates Economic Opportunities: Nonbank Financials, Fintech, and Innovation."

U.S. Small Business Administration. 2020. "Paycheck Protection Program (PPP) Report: Approvals through 08/08/2020." https://home.treasury.gov/system/files/136/SBA-Paycheck-Protection-Program-Loan-Report-Round2.pdf.

Wiersch, A. M., and S. Shane. 2013. "Why Small Business Lending Isn't What It Used to Be." Economic Commentary No. 2013-10, Federal Reserve Bank of Cleveland.

Wooldridge, J. M. 2015. *Introductory Econometrics: A Modern Approach.* 6th Edition. Boston: Cengage Learning.

Monetary Policy with Negative Interest Rates: De-linking Cash from Digital Money*

Katrin Assenmacher[a] and Signe Krogstrup[b]
[a]European Central Bank
[b]Danmarks Nationalbank

Monetary policy space remains constrained by the lower bound on nominal interest rates in many countries, limiting the policy options available to address future deflationary shocks. The existence of cash prevents central banks from cutting interest rates much below zero. In this paper, we consider the practical feasibility of recent proposals for de-linking cash from digital money to achieve a negative yield on cash which would remove the lower bound constraint on monetary policy. We discuss how central banks could design and operate such a system, and highlight some issues that require further research.

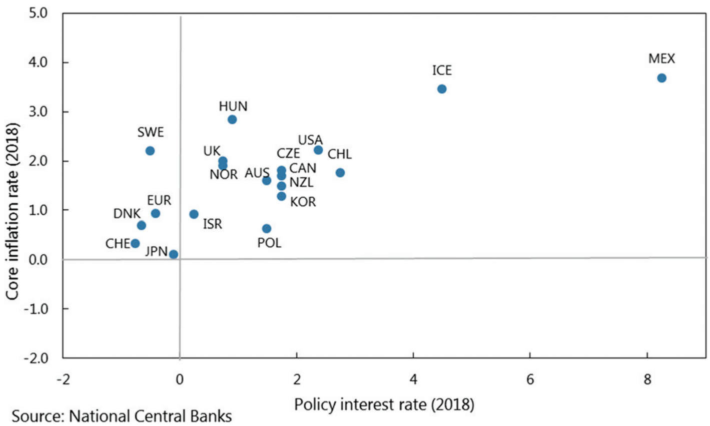
JEL Codes: E42, E52, E58.

1. Introduction

The global financial crisis brought policy rates to the so-called zero lower bound (ZLB) in many countries. Most of these remain in the vicinity of this lower bound 10 years after, as illustrated in figure 1. Central banks may not have sufficient policy space to counter the next recession, as the normal playbook would suggest countries

*The views expressed in this paper are those of the authors and do not necessarily represent those of the European Central Bank or the Danmarks Nationalbank. This paper was written while Signe Krogstrup was at the Research Department of the International Monetary Fund. We thank Larry Christiano, Giovanni Dell'Ariccia, Daniel Heller, Umang Rawat, Will Roberds, and Ted Truman as well as seminar participants at the Federal Reserve Bank of Atlanta, the University of Lausanne, the ROME Workshop Frankfurt, and the Swiss National Bank for comments. Wenjie Li and Jens Vedelsdal Aurup provided excellent research assistance. Author contact: Assenmacher: European Central Bank, Sonnemannstr. 20, 60314 Frankfurt am Main, Germany; e-mail: Katrin.Assenmacher@ecb.europa.eu. Krogstrup: Danmarks Nationalbank, Havnegade 5, DK-1093 Copenhagen K, Denmark; e-mail: skro@nationalbanken.dk.

Figure 1. Monetary Policy Space in OECD Countries (Turkey, omitted, had a policy rate of 24% in 2018)



Source: National Central Banks

should be able to cut rates by around 500 basis points to effectively respond to important negative shocks.[1] Figure 1 shows that the majority of OECD countries have monetary policy space of less than 250 basis points. Only two OECD countries, Mexico and Turkey (not shown), have policy space that exceeds 500 basis points.

The lower bound on interest rates is due to the existence of cash, which by design yields a nominal interest rate of zero.[2] If a central bank attempts to move its policy rate significantly below zero, commercial banks will see their interest margin compressed as long as they do not pass on the negative interest rate fully to all deposits.[3] Sufficiently negative interest rates on bank deposits may cause depositors to switch from (negative) interest-bearing deposits to cash, which could lead to a substantial outflow of deposits from the banking sector.[4]

[1]See Summers (2018).

[2]In this article, the term cash refers to physical currency, i.e., coins and banknotes.

[3]In most countries with negative interest rates, banks have, to date, refrained from passing on the negative interest to retail deposits. This compression is more pronounced if a bank relies more on deposit funding relative to funding from money and capital markets.

[4]Banks themselves face a similar tradeoff. When the negative interest rate on reserves exceeds the storage cost for vault cash, they may decide to convert excess

This risk of substitution from bank deposits into cash is at the core of the existence of a lower bound on interest rates. The experience of recent years suggests that the lower bound is somewhat below zero, as storing and handling cash is associated with cost and inconvenience compared with using money in a deposit account. No large-scale substitution toward cash has been observed in connection with negative interest rates as of yet. But there is no doubt that substitution would eventually set in and erode banks' funding base if interest rates were to become sufficiently negative. The intended stimulating effects of interest rate cuts substantially below zero would be undermined as the zero rate on cash, and not the negative rate on central bank reserves, would become the economically relevant interest rate.

The lower bound on interest rates hence poses a hard constraint on the ability of monetary policy to counter cyclical downturns, deflation, and unemployment in an environment where interest rates are already low. Numerous proposals have been made to increase the ability of monetary policy to provide stimulus when faced with the lower bound on short-term interest rates, such as adjusting exchange rate policy, raising the inflation target, conducting large-scale asset purchases, or phasing out cash to allow for substantially negative interest rates (Svensson 2003, Blanchard, Dell'Ariccia, and Mauro 2010, Rogoff 2014, and Ball et al. 2016). Each proposal has advantages and drawbacks, and only the latter fully removes the lower bound constraint.

In this paper, we discuss the practical feasibility of de-linking the value of cash from digitally issued central bank reserves as a way of fully removing the lower bound constraint on monetary policy while preserving a role for cash. In current monetary systems, banknotes and central bank reserves are issued and exchanged at par at the central bank, i.e., central bank reserves can be exchanged one-for-one into banknotes and vice versa on demand at the central bank's cash window. In a dual domestic currency system that we analyze in

reserves into cash. Storing large amounts of cash centrally, however, will create security and insurance issues. Moreover, in many countries vault cash is part of the minimum-reserve regulation, giving the central bank a lever to influence cash holdings at banks (see, e.g., Switzerland). Affecting the behavior of private agents is much more difficult.

this paper, cash can still be exchanged on demand for central bank reserves but at a time-varying cash reserve exchange rate rather than at par.

In the more recent literature, Buiter (2007) is the first to discuss such a proposal. Agarwal and Kimball (2015, 2019), Goodfriend (2016), and Pfister and Valla (2018) consider the depreciation of cash relative to reserves as a means to overcome the lower bound on interest rates. Our paper is close to Agarwal and Kimball (2015, 2019) in that we discuss how a central bank could design and operate such a system. Our contributions include considering the transmission and financial stability implications in more detail. We also discuss how a dual domestic currency system would work within different exchange rate regimes, as well as its interrelations with central bank digital currency (CBDC). While Agarwal and Kimball (2019) frame their proposal as a time-varying transaction fee for net deposits of cash at the central bank, we prefer to analyze the scheme in terms of an exchange rate between cash and digital currency, allowing us to apply economic concepts such as uncovered interest parity to investigate credibility and robustness. We stress that the central bank would need to apply this exchange rate symmetrically for deposits and withdrawals of cash at its cash window in order to not interfere with the cash cycle. A simpler, one-sided fee for cash withdrawals would not ensure that cash in circulation flows back to the central bank for quality checking and redistribution.[5]

De-linking the value of cash from central bank reserves establishes a system in which two different types of domestic currency circulate. The dual domestic currency system would allow the central bank to stabilize the economy in a severe downturn by implementing substantially negative interest rates. The negative interest rate would not trigger a large-scale substitution into cash, because the system would feature a similarly negative yield on cash in terms of central bank reserves. Banks, subject to both the negative interest

[5] Bordo and Levin (2017) analyze CBDC and propose a schedule of fees for transferring funds between CBDC and cash, which increase with the frequency and the amount of a transfer, in order to remove the ZLB. In their setup, however, they do not discuss how digital money on bank deposit accounts would be treated. A fee on transfers of cash into CBDC would not remove the ZLB if—like today—agents can still switch at no cost between bank deposits and cash, eroding banks' funding base.

rate on reserves and the depreciating rate of cash at the central bank's cash window, would transmit both to their deposit holders, implying that not only the value of central bank reserves would de-link from the value of cash but any form of digital money, such as bank deposits, would de-link as well.

We discuss some remaining unanswered questions concerning the system's legal and institutional implications, the transmission of monetary policy, financial stability and seigniorage revenues, and point to areas where more research is needed. We conclude that de-linking the value of cash from central bank reserves would fully restore monetary policy space by removing the lower bound. The system would be technically feasible and not require fundamental changes to central banks' operating frameworks. Moreover, it would be entirely reversible once the need for negative interest rates disappears. It would work in both fixed and flexible exchange rate systems and could be implemented unilaterally, even by small and highly open economies. Its benefits and drawbacks should be considered alongside the pros and cons of other proposals for increasing monetary policy space in a low interest rate environment.

The paper is structured as follows. The next section sets the stage by first considering why abolishing cash may neither be practically feasible nor desirable in many countries. In section 3, we describe how the dual domestic currency system that preserves the role of cash but allows for negative interest rates would work in practice. We consider the central bank's operating framework, the relationship with foreign currency and exchange rate policy in fixed and flexible systems, and the transmission of negative interest rates to cash and deposits, as well as payments in the broader economy. Section 4 addresses some open questions, and the final section concludes.

2. Why Not Simply Phase Out Cash?

Rather than de-linking cash from digital money, it would arguably be much simpler to phase out cash altogether, which would achieve many of the same advantages in terms of allowing monetary policy to operate below the lower bound. However, while some countries, notably Sweden, are quickly heading in the direction of a

Figure 2. Cash in Circulation in Percent of GDP in 2018

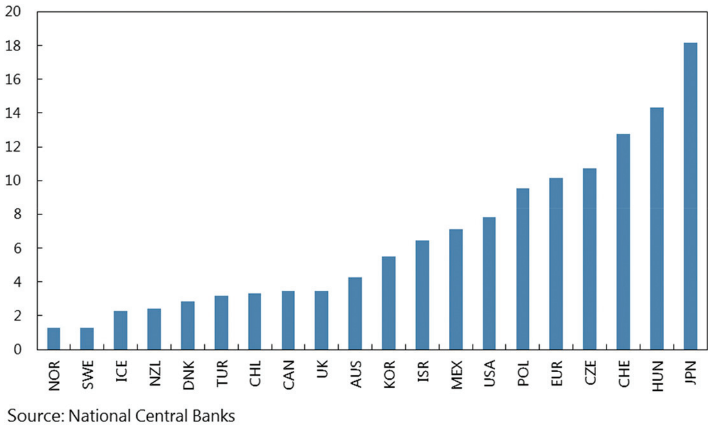
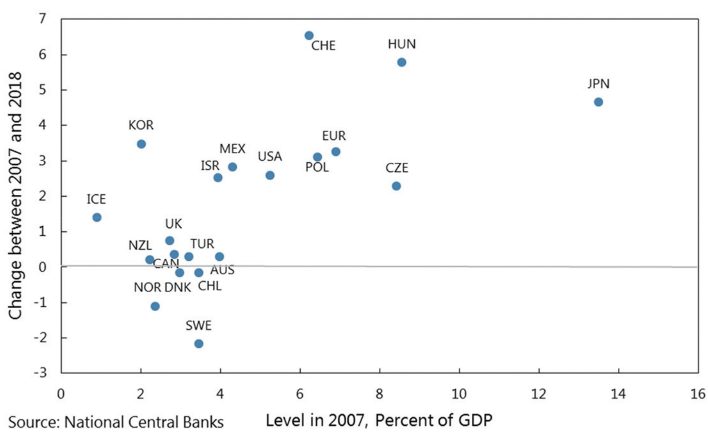


Figure 3. Cash in Circulation in Percent of GDP Level in 2007 and Change since 2007



cashless society, other countries remain strongly reliant on cash, as shown in figure 2. Figure 3 shows that only two countries—Sweden and Norway—saw an outright reduction in currency in circulation in percent of gross domestic product (GDP) in the past decade. For Canada, Engert, Fung, and Hendry (2018) conclude that the emergence of a cashless society would not generally cause material financial-system-wide problems. Khiaonarong and Humphrey (2019) estimate that, with the exception of India, cash usage will further fall in a sample of 11 countries, supported—among other factors—by demographic developments. Countries with relatively high outstanding amounts of currency in circulation also had high growth rates of cash circulation in the past decade, illustrating how the development in outstanding cash differs across countries.

Moreover, there are reasons why phasing out cash completely may be premature or undesirable. Cash currently serves three main uses in our societies. It plays a key role in retail payments, it is used for storage (i.e., hoarding of banknotes as a means of saving), and it is used for tax evasion and illegal activities. Rogoff (2014) argues that the first two functions of cash can nowadays be conveniently performed by digital forms of money, whereas the prevalent use of cash in tax evasion and illegal activities is an important reason for central banks to consider phasing it out. An added benefit to phasing out cash, he argues, is that a society without physical currency has no lower bound on nominal interest rates, allowing monetary policy to address cyclical downturns without constraints.

Cash remains important in retail payments in many countries, however. For this reason alone, central banks with mandates to promote the stability of payments systems cannot actively phase out cash. In the euro area, for example, cash is still the dominant payment instrument at the point of sale (Esselink and Hernández 2017), though the use of cash varies strongly between European countries.[6] Removing cash as a payments option before digital means of payment have become near-universal could disrupt the retail economy. Moreover, the use of cash and access to digital means of payment is not evenly distributed across demographic groups. Low-income and older population groups, for example, tend to use digital means of payments less. Phasing out cash could be particularly disruptive for the financial livelihoods of such population segments.[7]

[6] Similar results are found for Switzerland; see Swiss National Bank (2017).

[7] One option is to tailor public policy to achieving universal access to digital means of payment, and some countries—in particular, Sweden—are currently heading down this route (see, e.g., Sveriges Riksbank 2018). But such changes to the structure of payments systems take time and cannot be achieved overnight.

Beyond the stability of payments systems, there are institutional and cultural reasons why some countries may wish to hold on to physical currency. That cash payments are anonymous is seen in some countries as important for ensuring the right to privacy. Another key property exclusive to cash is exactly that it is not digital. If digital systems break down, cash is still usable and hence provides a hedge of the retail financial system against digital disruptions. This has turned out to be of high value in areas plagued by natural disasters that interfere with digital networks. Preparations for natural disasters in fact often include securing the provision of sufficient cash stocks.[8] Finally, if cash were abolished, the decision would be difficult to reverse, whereas in a dual domestic currency system all payment instruments that are in use today could remain in use and the corresponding infrastructures could remain in place and running.

3. De-linking Cash from Central Bank Reserves

Making cash as costly to hold as digital money in bank accounts or short-term money market instruments when interest rates are negative is an alternative to phasing out cash while creating space for monetary policy to stabilize severe downturns. Various proposals have been put forth for how to impose a cost on cash holdings. Gesell (1916) suggested discouraging cash hoarding by introducing a demurrage fee. His idea was that money would need to be stamped at regular intervals to remain valid and that these stamps would have to be purchased. Such a scheme was implemented in some Austrian and German communities during the Great Depression, but the practice was soon stopped by the respective central banks. A similar but untried proposal is to let banknotes expire at certain dates, forcing their holders to pay a conversion fee for changing them into new, valid banknotes (Seltmann 2010). Goodfriend (2000) suggests integrating magnetic strips into banknotes that record when the note

[8]See, e.g., http://ready.gov. Cheney (2006, p. 7) documents that in advance of Hurricane Katrina the Federal Reserve had to dispense exceptionally high amounts of cash to the affected regions. In light of the need to distribute financial relief to affected families, however, schemes to improve the digital payment infrastructure were seen as highly important as well.

was last withdrawn from the banking system and how much carry tax on that note is due. This option gets very close to simply replacing cash with a digital currency. Mankiw (2009) proposes a lottery scheme that declares a certain number of banknotes invalid at regular intervals. While all these ideas achieve a carry cost on cash, they seem impractical or unworkable.

A different proposal for implementing a negative carry cost on cash is to change the one-to-one conversion of reserves held at the central bank into cash. Buiter (2007) shows that, in theory, a negative yield on cash can be achieved by de-linking the value of cash from the value of digital money, such as bank deposits, allowing cash to depreciate over time in terms of digital money. Effectively, the idea entails a split of the domestic base money supply into two different domestic currencies: cash and digital central bank reserves. The central bank could then impose a negative yield on cash in terms of central bank reserves and thereby continue conventional monetary policy below the lower bound. Agarwal and Kimball (2015) show that by introducing a time-varying deposit fee, the central bank effectively would establish an exchange rate between cash and central bank reserves. The implicit negative yield would transmit to the economy through conventional channels while private digital money creation (e.g., bank deposits) can be left to adjust freely. They also consider many practical and operational aspects of how such a system would work.

We argue below that such a system could be implemented with relatively small changes to central bank operating frameworks. It would work in fixed as well as flexible exchange rate systems, even when foreign banknotes are close substitutes to domestic cash. In short, we conclude that a dual domestic currency system should be workable from a technical and operational perspective. Moreover, we argue that interest rate changes in negative territory would transmit in a similar way as conventional interest rate cuts to the real economy once such a system is in place and that financial stability implications are broadly equivalent to those in current low interest rate environments. Nevertheless, in a monetary system in which cash and central bank reserves circulate with different value, new challenges arise with regard to the legal environment as well as communication with and behavioral responses of the public. We discuss some of these issues in section 4.

3.1 How Would It Work? Setup and Operating Framework

The central bank would divide the monetary base into two separate domestic currencies, referred to as cash and reserves in the following. Cash would be issued in physical banknotes and coins. Reserves would be issued only digitally. Reserves would pay nominal interest, possibly negative.[9] Denote the overnight rate on reserves i_t^R, also referred to as the policy rate. Moreover, the central bank would set the spot cash reserve conversion rate (henceforth referred to as the CRC rate) for cash withdrawn from or deposited in the central bank's reserve accounts and supply cash fully elastically on demand against reserves at this price. This cash conversion rate would apply to the central bank's operations with financial institutions that hold reserve accounts with the central bank, i.e., mainly bank counterparties.

Under such a monetary framework, banks depositing cash into their reserve accounts with the central bank would see their reserves credited not at par, but at the prevailing CRC rate. The CRC rate would apply symmetrically, just like any other exchange rate. Banks taking out cash from their reserve account would see their reserve account debited at the CRC rate. The change of the CRC rate, not its level, would determine the yield on cash in terms of reserves. To remove any incentive for banks to move into cash when a negative interest rate is applied to reserves, the central bank would use the CRC rate to steer the demand for cash, by changing the spot conversion rate between reserves and cash over time in order to impose a sufficiently negative yield on cash in terms of reserves.

Assuming that the negative yield on cash would be set equal to the negative interest rate on reserves, the central bank would set the conversion rate such that $i_t^C = \{360 \times (CRC_{t+d} - CRC_t)\}/\{d \times CRC_t\}$, where the subscript d refers to time units (days) between adjustments of the CRC rate, and i_t^C is the annualized yield on cash in terms of reserves during that time. Note that there are no expectations signs in this interest parity because the conversion rate tomorrow and the rate on reserves held from today

[9] Interest on reserves is currently used by several central banks to establish a floor for money market rates and has been discussed as a monetary policy tool in times of high liquidity, especially in the United States; see Gagnon and Sack (2014).

until tomorrow are both known and set by the central bank.[10] In essence, the time-varying CRC rate would give the central bank a tool to control the demand for cash relative to the demand for reserves. For ease of exposition, we assume in the following that the conversion rate is set such that the negative yield on cash equals the interest on reserves.

To ensure a smooth functioning of the scheme, the conversion rate applied to the central bank's cash operations with its counterparties would have to be adjusted continuously, preferably daily. Discrete jumps in the CRC rate would redistribute wealth between cash holders and reserve holders at the moment of the jump, which would be destabilizing. The chosen adjustment frequency would have to be announced up front to ensure transparency and avoid speculation in the timing of adjustments. For example, at the end of each business day, the central bank could announce that tomorrow's rate of conversion of cash into reserves is $CRC_{t+1} = CRC_t \left(1 + i^c_{t+1,t}/360\right)$ Alternatively, the central bank could announce a path for the CRC rate that would be followed until the next policy meeting. If the central bank were to change its policy rate, the corresponding new level for tomorrow (or the new path) of the CRC rate would be announced simultaneously and accordingly. The central bank could use the same setup and frequency for changes in the CRC rate as for changes in its usual policy rate.

Once the system has been used to engineer a negative rate on cash, the central bank cannot exit by simply setting the conversion rate back to par when interest rates move back to zero. Such a discrete "appreciation" of cash in terms of reserves would redistribute wealth from deposit to cash holders and be destabilizing. If the central bank were expected to eventually exit in this way, there would be strong speculation in taking out cash before the exit, which would be counterproductive to the purpose and functioning of the system.

[10]The yield would not need to be equal to the interest rate on reserves. In principle, a slightly less negative yield on cash than the interest on reserves could be enough to deter a shift into cash, given the storage cost of cash. In contrast, if the central bank were aiming to reduce the demand for cash even further—for example, during a run out of bank deposits and into cash—the yield could be made even more negative. In current systems, a return differential between cash and reserves is the norm when reserves pay interest, as in the current Federal Reserve system.

It could also pose financial stability risks. Instead, an exit would not be disruptive when interest rates have been positive for long enough to bring the CRC rate back to par. Cash would appreciate by the same amount that it has depreciated during the preceding negative-interest episode in order to not create arbitrage opportunities for cash holders by switching out of depreciated cash into digital reserves at par. Cash holdings would thus receive an implicit positive interest equal to the positive interest on reserves for as long as necessary to bring the CRC rate back to par.

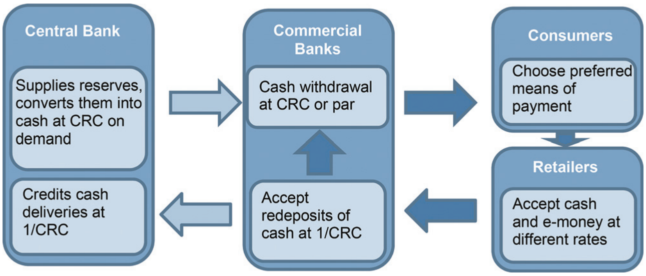
To increase transparency and avoid disruptive speculation in the exit, the central bank could announce up front if and how it plans to exit the system, whether the conversion rate would remain in place or whether it would be abandoned at the moment it is back to par. To illustrate this point and to fix ideas, a numerical example of how a dual domestic currency system could be operated is presented in the next section.

In principle, however, such a dual domestic currency system could remain in place even when the CRC rate has reached par again. The obvious benefit from leaving it in place, once the fixed costs of introducing the system were borne, would be to always be ready for an accommodative monetary policy at negative interest rates. There could also be disadvantages, such as possibly reducing seigniorage revenues, as we discuss in section 4.1 below, or having cash prices deviating from digital money prices permanently. It could also be more politically palatable and easier to introduce if the measure is announced as temporary.

3.2 A Numerical Example

Suppose a central bank moves from a zero to a negative rate of –3 percent per annum (p.a.) on reserve accounts at the central bank. If cash remained convertible one-for-one with central bank reserves, there would be a strong incentive for banks and, in turn, for the non-bank public, to hoard cash. This could trigger a run toward cash, which could endanger economic and financial stability—the lower bound on interest rates would be reached. To prevent a run to cash, the central bank simultaneously announces a shift to a dual domestic currency system with an initial $CRC_0 = 100$ (defined in units of cash per 100 units of reserves). At the same time, the central bank

Figure 4. Transmission of the Cash Reserve Conversion Rate

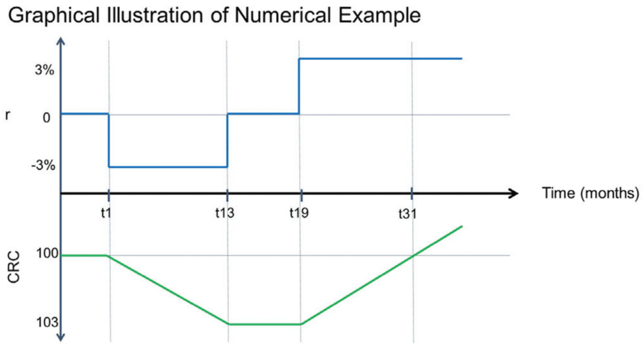


announces that the yield on cash in terms of reserves is set equal to the interest on reserves (for simplicity of the example) and hence that tomorrow's CRC rate will be $CRC_1 = \frac{100}{(\sqrt[360]{1-0.03})} = 100.0085$.[11] After a year with −3 percent p.a. interest on reserves, the CRC rate would be 103, i.e., 100 units of cash would be converted into roughly 97 units of reserves. Figure 4 depicts this case, based on monthly CRC rate adjustments for the sake of illustration. As long as the interest rate is negative, banks would receive increasingly more cash over time when they withdraw funds from their account at the central bank. The effective depreciation of cash over the holding period would correspond exactly to the accumulated negative interest on reserves over the same period.

In the example in figure 5, the negative interest rate is abandoned at month 13. If an exit from the dual local currency system is desired, the conversion rate would have to remain in place for at least the amount of time that it would take to bring the conversion rate back to par. After a period with a negative interest rate, this would imply that the cash yield would have to be positive for a while. In the example in figure 5, the CRC rate is kept at 103 for as long as the interest on reserves remains zero. In month 18, we assume that the central bank wishes to tighten policy by increasing the interest rate to 3 percent p.a. Now the CRC rate starts reversing, in order to ensure that the yield on cash in terms of reserves is also a positive 3

[11] Using compound interest, the square-root expression refers to the daily equivalent gross compound interest of a yearly interest rate of −3 percent, with a year defined as 360 days.

Figure 5. Negative Interest Rates and the Cash-Reserve Conversion Rate



percent p.a. If the interest rate remains at 3 percent, the CRC rate returns to par in month 30. At this point, the system can be safely exited.

3.3 Transmission to Bank Deposits and Beyond Banks

How would the introduction of the conversion rate transmit to the cost of using cash and digital bank deposits for payments in the rest of the economy? The answer to this question depends on behavioral, legal, and other types of responses of the broader economy. There are no obvious empirically relevant historical episodes that can inform these questions.[12] We briefly discuss how banks and wholesale clients

[12] After a failure to stabilize successive cycles of hyperinflation, the Reserve Bank of Zimbabwe (RBZ) demonetized the Zimbabwean dollar from June to September 2015 and converted all remaining currency into U.S. dollars (RBZ 2015). In 2016, the RBZ began issuing so-called USD-denominated bond notes that were pegged 1:1 to the U.S. dollar and could be deposited into existing U.S. dollar accounts (RBZ 2016). As trust in the U.S. dollar accounts (the so-called RTGS dollar) evaporated, U.S. dollar bills were exchanged at a premium against digital U.S. dollars, but also bond notes traded at a (smaller) premium to the digital RTGS dollar; see https://zwnews.com/latest-us-dollar-zimbabwe-bond-note-trgs-exchange-rates-today-9-october-2018/, retrieved on October 27, 2019). Zimbabwe is thus a rare example of a country that experienced a spread between physical and digital currency. We were unable to find any evidence on how this spread affected pricing in the economy. Though confusion on pricing was large (https://www.victoriafalls-guide.net/zimbabwe-currency.html, retrieved on

are likely to respond given current behavior and use these considerations as a bridge to the central questions about broader transmission to financial markets, instruments, and the real economy addressed in section 4.

At first, the central bank's counterparties, i.e., banks, are faced with the CRC rate at the cash window in conjunction with a negative interest rate on their reserve holdings. Whether and how a commercial bank would pass on the conversion rate between reserves and cash to its customers would not need to be dictated by the central bank but could remain a business decision by banks, just as is the case with negative interest rates on central bank reserves, as also pointed out in Agarwal and Kimball (2015). All else equal, providing cash to a bank deposit holder would require the bank to demand cash in exchange for reserves at the central bank's cash window, where the CRC rate applies.

For relatively short and mild episodes of policy rates below zero (e.g., below a margin of a few percentage points and within a reasonably short period), the CRC rate might not be passed on to banks' customers, as the necessary changes to cash systems would imply a one-off cost for banks. Automatic teller machines could continue to work with a unit conversion factor, at least for retail-sized withdrawals and deposits, and merchants would likely continue to accept payments in whatever form the customer prefers.[13] During episodes of mildly negative interest rates on central bank reserves, the negative interest rate would probably not be passed on to retail-sized bank deposits and its transmission would remain limited to wholesale deposits, as we have seen in some countries. It could take time to overcome the psychological or institutional barriers to negative

---

October 27, 2019), other issues related to Zimbabwe's history of monetary instability seemed to be more important than the spread between bond notes and RTGS dollars.

[13]As Agarwal and Kimball (2015) also pointed out, this would correspond to the current system, in which different means of payment also bear different costs, but businesses prefer to take these costs onto their margins and generally do not differentiate prices according to different means of payment. Based on current limits for credit card fees that are absorbed by banks and businesses, one can speculate that this threshold resides somewhere around 3 percent accumulated depreciation of cash relative to reserves.

interest on retail bank deposits that have been observed in recent episodes.

Eventually and with sufficiently negative interest rates, however, banks would have to pass on negative interest to bank deposits in order to remain profitable. The pressure on banks' profits would not result from customers withdrawing cash at par since cash would depreciate in terms of digital money, meaning that customers would actually get more cash for their withdrawal of an equivalent amount of digital money. Instead, the CRC rate would transmit via cash deliveries from businesses that banks would have to credit at less than par to remain profitable. Via this channel, businesses would have a strong incentive to differentiate between cash and digital-money prices once the CRC rate between cash deliveries and digital payments becomes too steep. While a dual domestic currency system has not been tried yet and we cannot analyze consumer behavior in response to different prices for cash and digital money, results of De Grauwe, Rinaldi, and Van Caysele (2006) suggest that consumers react strongly to changes in the cost of using cash or digital means of payment.

How would this affect the demand for cash overall? While cash as a means of payments could become less attractive, a large share of cash seems to be demanded for hoarding purposes (see Bech et al. 2018). Agents would need to understand that—because of the depreciation of cash in terms of digital money—the value of their cash holdings at any point in time during a negative interest episode is exactly equivalent to the balance on a bank account with accrued (negative) interest at the official policy rate. Once this has been understood, banks should be able to pass on the negative interest rate without triggering a run into cash. This requires transparent and convincing communication on the part of the central bank and rational behavior on the part of the public, a point to which we return in section 4.3.

When banks have paid the one-off cost of adapting systems to the CRC rate and the stickiness of retail deposit rates at zero has been overcome, passing on the CRC rate and interest rate cuts into negative territory to retail customers could be immediate. Based on current experiences with banks passing on mildly negative interest rates to institutional clients but not to retail clients, it is probable that the conversion rate would first be passed on to institutional, nonbank financial firms and possibly nonfinancial firms before households and small enterprises would be confronted with it.

Figure 4 provides an illustration of how the CRC rate would be passed on from the central bank via the commercial banking system to retailers and consumers. The figure shows that the behavior of cash-handling companies, which typically carry out the transfer of cash between the central bank and commercial banks, and their price setting, would play a central role in the transmission of the CRC rate to the economy.

To make sure that the transmission is working properly, it is crucial that the CRC rate is applied symmetrically, i.e., that withdrawals are credited at above par, giving the cash-handling companies an incentive to obtain cash at the central bank instead of trying to re-issue banknotes without channeling them through the central bank. If the CRC rate would only apply to deposits of cash and not to withdrawals, there would be an incentive to short-circuit the cash cycle, and the transmission of the CRC rate to the broader economy could be at risk.

Passing on the CRC rate to bank customers should not create any problems for cash circulation or payments. Cash is usually withdrawn by the banks' retail customers from their deposit accounts and spent for purchases of goods and services. The bank would decide whether customers can withdraw cash at par or at the CRC rate from their accounts. Shops and other firms receive cash payments and bring them back to the bank, which would likely credit these cash deposits at the conversion rate that it faces at the central bank, i.e., firms' accounts with their bank would be credited at less than par.[14]

It is uncertain how the dual domestic currency system would transmit to pricing and price quotes—i.e., which currency, cash or reserves, would become the main unit of account. Confronted with a higher cost for depositing cash, firms would perhaps first—before quoting cash prices that differ from prices in digital money to their

[14] The setting of the conversion rate would be a business decision of the bank. There would not be any requirement that this conversion rate has to be identical to the one the central bank is applying. Banks would be free to set a more or less favorable conversion rate for their customers. They could also choose to adjust it less frequently than the central bank, as has generally been the case with deposit rates relative to the policy rates.

customers—have an incentive to influence their customers toward paying with digital money instead of using depreciating cash.[15] If the value of cash were to deviate sufficiently from bank deposits and reserves, however, firms would have to choose which currency to use for price quotes and let prices in the two currencies deviate. Having legislation in place that supports price quotes in digital money would help economic agents to coordinate on a common unit of account and ensure that the dual domestic currency system is able to remove the lower bound on interest rates. We address this question, central for the functioning of a dual domestic currency system, in more detail in section 4.3 below.

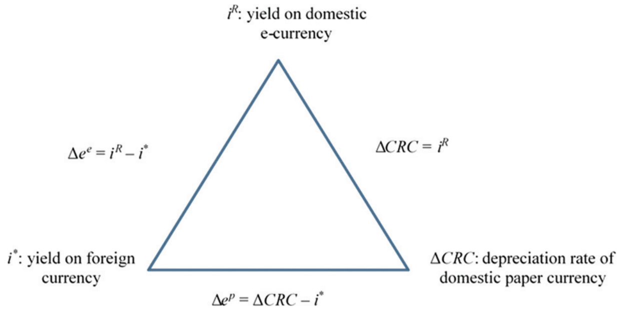
3.4 Small Open Economies and the Exchange Rate Regime

With cash and central bank reserves de-linked into two distinct domestic currencies, there would also be two foreign currency exchange rates, namely one for domestic cash per foreign currency and one for domestic digital central bank reserves and bank deposits per foreign currency, assuming that foreign currency is not unbundled into cash and reserves (in which case, four exchange rates would apply). The relationship between the two foreign exchange rates would in market equilibrium be determined by the central bank's choice of the CRC rate between domestic cash and reserves, as the central bank would be the monopoly supplier of both cash and reserves. The relationship between the three exchange rates would be identical to the relationship between the bilateral exchange rates of three pairs of freely traded separate foreign currencies, except that, in this case, at least one of the exchange rates (the CRC rate) would be deterministically fixed by the central bank. The relationship between the other two exchange rates would be determined by the CRC rate. There would only be one exchange rate left for either the market or the central bank to fix.

Figure 6 illustrates the uncovered interest parity (UIP) relations between foreign currency and the two domestic currencies. If cash were depreciating at a rate consistent with the negative interest rate on reserves, UIP between the three bilateral exchange rates would

[15] For small accumulated rates of depreciation of cash, this could initially take the form of bonuses or coupons for noncash payments.

Figure 6. Illustration of Uncovered Interest Parity with a Dual Domestic Currency System



apply, abstracting from risk premiums. Financial market equilibrium would ensure that agents would be indifferent between currencies (cash, reserves or foreign) to invest in. If the central bank were to let cash depreciate more or less quickly than the negative rate on reserves to accommodate demand fluctuations, the two foreign exchange rates for cash and digital currency would reflect this difference.[16] Overall, this would not create any new arbitrage opportunities as compared with the single domestic currency case. In short, adding a foreign currency to the dual domestic currency system with a corresponding exchange rate does not change the relationship between the two domestic currencies.

In a floating exchange rate system, the exchange rate of domestic central bank reserves against foreign currency would be market determined, and the exchange rate for domestic cash against foreign currency would be given by the exchange rate of digital reserves against foreign currency times the CRC rate. There would be no arbitrage incentive to shift to foreign currency—whether foreign reserves or foreign cash—in light of a negative yield on domestic cash and bank deposits. If the domestic central bank wanted to ease monetary policy further below zero, it could do so by implementing a faster depreciation of domestic cash in terms of digital reserves. This

[16]We would expect the relations on the foreign exchange market to be dominated by the exchange rate for domestic digital currency, as cross-border transactions are primarily digital. In general, foreign exchange rates for cash transactions exhibit almost prohibitively wide spreads.

would transmit to the economy through the normal monetary policy transmission channels, including through the exchange rate channel. Thus, initially, a lower domestic interest rate would increase the demand for foreign currency. This would trigger a domestic currency depreciation which would help boost domestic demand and inflation. The domestic interest rate could in turn be increased, sooner rather than later.

In a fixed exchange rate system, in contrast, the goal of the central bank would be to keep the foreign currency exchange rate of domestic reserves stable, and the interest rate would be chosen with this objective in mind. In this situation too, a dual domestic currency system would provide the tools necessary to meet the mandate in a low interest rate environment. Suppose, for example, that domestic and foreign interest rates were at their lower bound as a negative risk shock in the foreign currency area causes safe-haven-like appreciation pressures on the domestic currency. These appreciation pressures could be countered by cutting domestic interest rates into negative territory by switching unilaterally to a dual domestic currency system. The negative yield on cash and negative interest rate on bank deposits would trigger substitution away from domestic currency and into foreign banknotes, reserves, or bank deposits, thereby alleviating the risk-induced appreciation pressures on the domestic currency.

Suppose instead that the central bank issuing the foreign reference currency cuts interest rates into negative territory in response to a negative shock by unilaterally implementing a dual domestic currency system. The lower foreign interest rate would induce appreciation pressures on the domestic currency. In this case, the domestic central bank could maintain its peg by following suit and implementing a dual domestic currency system too, which would allow the domestic interest rate to fall in tandem with the foreign rate.

It is important to note that the dual domestic currency system discussed here is not akin to so-called dual exchange rate regimes operated in some countries in history. In typical dual exchange rate regimes, central banks impose two different exchange rates for converting a unique domestic currency into foreign currency, where the applied rate depends on the motive for the transaction (e.g., imports or capital account transaction) or the types of counterparties in the transaction. Dual exchange rate regimes suffer from a number of problems, making them unsustainable. If the market for foreign currency is not controlled and fully segmented, there would be unlimited arbitrage opportunities from buying and selling foreign exchange at the two rates. To avoid such arbitrage, the central bank has to restrict or ration access to domestic or foreign currency at some of the rates. In history, black markets and rent seeking in response to the arbitrage opportunities have inevitably developed and led to the downfall of dual exchange rate systems. Similar problems have characterized the historical experiences of early banks in issuing parallel currencies during eras of free banking.[17] In contrast, the dual domestic currency regime that we discuss here is not associated with any unlimited arbitrage opportunities as long as the CRC rate is set according to the principles discussed above. The central bank would issue two domestic currencies, and set the interest rate on reserves and the conversion rate for reserves into cash consistent with this interest rate, allowing the market to set the interest on bank deposits, bonds, and the respective exchange rates vis-à-vis foreign currencies. Arbitrage would be stabilizing for the system. It would ensure that the foreign exchange rates would be consistent with UIP. Currency controls or market rationing would not be required to make the system operational.

We conclude that the practical operation of a dual domestic currency system should not require major changes to current monetary policy frameworks. The interactions of the central bank with its counterparties under a dual domestic currency system would be straightforward and would follow the same lines as monetary policy operations under contemporary systems. Both fixed and floating exchange rate regimes could implement the system, unilaterally as well as in coordination with foreign monetary authorities. We next turn to issues surrounding the introduction of a dual domestic currency system and its ability to provide monetary policy space at the lower bound.

[17] The Bank of Amsterdam operated an early version of a dual local currency system in the 17th century by unbundling its deposits in unredeemable account balances (reserves) and coins that could be withdrawn (Quinn and Roberds 2014). The main difference to our scheme is that in a modern central bank both components, i.e., cash and reserves, are fiat money and thus unredeemable.

4. Issues and Unanswered Questions

As the above discussion suggests, breaking the unit conversion between cash and reserves would be operationally feasible with small changes to central banks' operational frameworks. To make a dual domestic currency regime ready for implementation in case of need, however, would also require changes to current legal frameworks. Concrete proposals are beyond the scope of this paper. Instead, we discuss below what we consider to be the main questions needing attention. Following the structure in figure 5, we first consider issues that might arise for the central bank, such as implications for monetary policy implementation and seigniorage revenues. Next, we discuss issues related to the banking system such as transmission of negative rate to bank lending and implications for financial stability. Finally, we consider the transmission of deeply negative interest rates to the broader economy, which will depend critically on the behavior of firms and consumers.

4.1 Implications for the Central Bank's Balance Sheet

Introducing a dual domestic currency system will have implications for the size of the central bank's balance sheet with potential consequences for monetary policy implementation and seigniorage revenues. Introducing a dual domestic currency system with the digital currency as the relevant unit of account would presumably lead to an increased use of digital payments at the expense of cash. Depending on how much digital money replaces cash in circulation and whether commercial banks or the central bank itself provides nonbanks with digital currency, the central bank's balance sheet might either expand or shrink.

Prospects of a significant reduction of the central bank's balance sheet have led to discussions about the size of the monetary base that is necessary to effectively implement monetary policy. Friedman (1999) expresses concerns that an evaporating demand for base money would make it more difficult for the central bank to control financing conditions in the economy. By contrast, Woodford (2000) argues that the central bank would continue to be able to control short-term interest rates even if the demand for base money evaporated completely, and that monetary policy effectiveness was independent of the size of the central bank's balance sheet.

While the central bank's balance sheet would presumably shrink when commercial banks continue to provide nonbanks with digital money, its evolution is ambiguous if the central bank itself were to open its balance sheet to the general public, motivated, for instance, by a desire to provide a generally accessible legal tender. If consumers were to replace cash with CBDC, the central bank's balance sheet would remain of the same size but its liability structure would change, with reserves increasing at the expense of banknotes (see Meaning et al. 2018). If consumers were to substitute CBDC for bank deposits, the central bank's balance sheet would lengthen, forcing the central bank to acquire more (interest-bearing) assets to counterbalance its increased liabilities. This could raise governance issues, as more credit would be intermediated through the central bank instead of the private sector. Both scenarios imply that if the CRC rate were to remain in place with positive policy rates (as discussed in section 3), a larger part of the central bank's liabilities would be remunerated at the policy rate. Although the central bank would not incur losses as long as interest rates are negative, this could change once interest rates rise.[18]

Seigniorage revenue arises from the difference between the yields on the central bank's assets and its liabilities. The zero interest rate on cash—and in some countries also reserves—is currently the primary source of the central bank's seigniorage revenue. A decrease of cash in circulation or an increase of the share of remunerated reserves could both reduce seigniorage and thus affect the long-term profitability of the central bank. Whereas the average yield spread narrows when interest-bearing reserves replace cash, total seigniorage revenues might increase or decrease, depending on how much the balance sheet expands.

While generating profits is not an objective of a central bank, sufficient seigniorage revenue to cover operating costs is often seen

[18] Some central banks, e.g., the European Central Bank, have always paid interest on reserves whereas other central banks, e.g., the Federal Reserve, recently shifted to paying interest on reserves. Interestingly, proposals on CBDC currently under discussion consider a zero rate on CBDC (Sveriges Riksbank 2018). To effectively remove the lower bound, however, CBDC would need to be interest bearing, at least when the policy rate becomes negative.

as important in ensuring central bank independence and therefore credibility. As important as these considerations are, central bank seigniorage revenue is likely to change in the future, driven by financial innovation and the ensuing increased use of digital payment options at the expense of cash. More research is needed to assess how much seigniorage could decline and whether it could be severe enough to affect the central bank's ability to pursue its price stability target.

4.2 Implications for Monetary Policy Transmission and Financial Stability

In this section, we consider issues relating to the middle panel of figure 4, namely possible implications of the introduction of a dual domestic currency for the banking system. Our reflections are centered on two main questions. First, would interest rate cuts in negative territory transmit in the same way to financial conditions and bank lending as interest rate cuts in positive territory? Second, would more negative interest rate levels—coupled with the introduction of a dual domestic currency system—have adverse implications for financial stability? To answer these questions, we consider how the experience with moderately negative interest rates in some economies to date can inform our thinking about more deeply negative interest rate environments.

In general, the recent experience with negative interest rates suggests that the transmission to money and bond markets of moderate policy rate cuts to negative values works like rate cuts in positive territory. When policy rates were lowered into negative territory, interest rates fell more broadly, as in normal times (Bech and Malkhozov 2016, Christensen 2019). It seems reasonable to assume that money market and bond market rates would keep declining with policy rates if these were to be lowered further. Evidence on lending rates and bank lending is more mixed (Ball et al. 2016) but overall points to a positive impact of negative interest rates on bank lending so far.[19] By contrast, banks to date have not generally imposed

[19]See Albertazzi, Nobili, and Signoretti (2017), Demiralp, Eisenschmidt, and Vlassopoulos (2017), International Monetary Fund (2017), Altavilla et al. (2019), and Eisenschmidt and Smets (2019).

negative interest rates on retail deposits of households (Jackson 2015, Bech and Malkhozov 2016, Jobst and Lin 2016). Although non-deposit funding has become cheaper for banks, their funding costs have not decreased as much as policy rates because of the stickiness of retail deposit rates, narrowing their interest margin.

In a dual domestic currency system, however, a negative yield on cash can be imposed and, therefore, deposit rates could possibly more easily breach the sticky line of zero, eliminating one of the obstacles to the full transmission of negative rates.[20] Banks would not face the threat of a bank run when interest rates are cut on retail deposits, and banks would hence be able to lend at more negative interest rate levels while maintaining their interest margins.

Studies that question the ability of negative interest rates to transmit to bank lending take a zero yield on cash and therefore a lower bound on deposit rates as given.[21] For instance, Brunnermeier and Koby (2019) suggest that due to the interplay between negative interest rates, regulation, and liabilities fixed in nominal terms (such as deposits on sight or savings accounts), interest rate cuts cease to stimulate bank lending if rates become too low. Once recapitalization gains are offset by tighter interest margins, low interest rates reverse their effect. Following similar reasoning, Eggertson, Juelsrud, and Wold (2017) build a model showing that a lower bound on deposit rates limits the extent to which a central bank can stimulate the economy by lowering its policy rate.

Accordingly, it has been argued that negative interest rates have adverse effects on bank profitability with ensuing negative consequences for financial stability. Though periods of low interest rates tend to coincide with lower bank profitability, there is no evidence that negative interest rates are causing low bank profitability (Ball et al. 2016). Typically, bank profitability suffers from weak

[20]Some countries (e.g., France) currently face legal or regulatory constraints to lower deposit rates to negative values. Such regulation or law would have to be changed to make the transmission of negative rates effective. In section 4.3 we discuss other regulation that might need to be adapted to ensure the effectiveness of a dual domestic currency system.

[21]Evidence presented in Ball et al. (2016) suggests that the transmission to bank lending rates in Switzerland following the interest rate cut into negative territory was indeed weaker than usual. This pattern, however, was not seen in other negative interest rate countries where bank lending rates responded normally.

macroeconomic dynamics and central banks respond by lowering policy rates. Moreover, Altavilla, Boucinha, and Peydró (2018) conclude that the adverse effects of monetary policy accommodation on banks' net interest margins in the euro area were largely offset by a positive impact on credit demand and quality as well as capital gains derived from the increase in the value of the securities held by banks. In the recent episode of low interest rates, the use of large-scale asset purchases and the flattening of the yield curve may be more important factors affecting bank profitability. A dual domestic currency system would enable banks to lower deposit rates, perhaps steepen the yield curve, and hence affect interest margins positively.

More fundamental implications for the structure of the financial system could arise if a dual domestic currency system were accompanied by the introduction of CBDC. Demand for traditional bank deposits as well as for physical currency could drop, eroding traditional bank funding models, causing bank disintermediation, and affecting the nature of payments systems. The literature on CBDC is developing quickly (Andolfatto 2018, Bank for International Settlements 2018, International Monetary Fund 2018, Brunnermeier, James, and Landau 2019, Chiu et al. 2019, Keister and Sanches 2019) and raises questions about the role of the central bank in payments and the allocation of credit that are beyond the scope of this paper. These issues are not specific to a dual domestic currency system and should be studied in their own right, as cash increasingly gives way to new digital means of payment.

Another financial stability concern relates to financial institutions' business models that incorporate nominal return targets and might lead to increased risk-taking.[22] Short-term debt is often regarded as "money-like" in the sense that holders consider its value fixed in nominal terms though it earns interest and is subject to credit risk. If nominal values or guaranteed nominal returns are written into contracts or regulation, negative interest rates may lead agents to invest in higher-risk assets and intensify search-for-yield behavior. Such behavior, however, is not specific to negative interest rates but relevant for long periods of below-average interest rates as

[22] Heider, Saidi, and Schepens (2017) found evidence that deposit-funded banks started to lend to riskier borrowers when interest rates became negative.

well. As such, it might be even more of a risk in the present system where countercyclical monetary policy is hampered and cannot contribute to reducing the length of a downturn.[23]

The financial stability implications of more negative interest rates would need to be weighed against the consequences of long periods of low interest rates and below-average economic performance when monetary stimulus cannot be provided forcefully at the lower bound during steep downturns (International Monetary Fund 2016). Being able to remove the constraint on monetary policy resulting from the lower bound would help shorten a downturn and the resulting low-growth period, getting the economy back on track, and thus back into positive interest rate territory, more quickly.

A number of other, less significant issues might arise in the context of negative interest rates, related to the implied change in direction of interest payment flows and the current definition of default (McAndrews 2015). During the recent experience with negative interest rates, negative yields on bonds have been achieved by issuing the bond at a price above par. While this is unproblematic at slightly negative rates, it could become more contentious when interest rates become significantly negative. Other issues are taxes, which often apply to coupon payments but not to capital gains, or the calculation of present values at negative rates. Since negative interest rates would be able to respond forcefully to a downturn, we would expect them to stay in place for only a limited time to help the economy recover more quickly. We hence do not see these issues as fundamental caveats to implementing more negative interest rates.

Summing up, the reason for introducing a dual domestic currency system is to recover monetary policy space in response to strong downturns when nominal interest rates are near zero. We do not see any reason for this to change monetary transmission to the financial sector fundamentally, nor do we see any unmanageable financial

[23]Lucas, Schaumberg, and Schwaab (2017) investigated bank business models at zero interest rates. They found that banks responded to changes in the yield curve. In particular, lower long-term interest rates led to increased size, leverage, complexity, and a less stable funding base for banks. This suggests that monetary policy measures targeted at longer-term rates may also have financial stability implications.

stability implications.[24] By using the CRC rate in a dual domestic currency system to steer the demand for cash, the authorities could even obtain a new tool to strengthen financial stability in situations where bank customers face an incentive to withdraw cash on a large scale.

4.3 Consumers and Retailers

In this section, we discuss whether deeply negative interest rates would successfully stimulate consumption and investment, addressing the right-side panels of figure 4. Moreover, we address issues related to the functions of money as a unit of account and legal tender that may arise for consumers and retailers in a dual domestic currency system.

If interest rates can be turned deeply negative, will economic agents respond by increasing investments and reducing savings or might they save even more to compensate for decreased interest income? From a theoretical perspective the *real*, not the nominal, interest rate should matter for saving and investment behavior. Real interest rates have been negative on many occasions in many countries, as expected inflation has exceeded policy rates substantially, notably in the 1970s, without triggering a discreet behavioral shift toward savings (see also Ball et al. 2016). When appropriately accounting for changes in neutral real interest rates, such episodes seem to be associated with improvements in macroeconomic conditions, as theory and the Euler equation would suggest (Krogstrup 2017; see also Cúrdia 2019).

We cannot exclude, however, that negative *nominal* interest rates might have different effects. Economic agents may partly allow nominal considerations to guide their decisions, at least in the short term, due to money illusion (e.g., Fehr and Tyran 2001). Cliffe (2016) reports that in a survey of bank clients across 15 countries, 11 percent out of 78 percent of respondents that would change their saving

[24]Concerns have been voiced in the context of the Federal Reserve's reverse repurchase (RRP) program, which gives financial institutions access to the Federal Reserve's balance sheet as a counterparty in repo operations; see, for example, Anderson and Kandrac (2016). Observing how the RRP program works during a possible future financial instability event could help inform on this issue.

behavior if they were confronted with negative interest rates indicated that they might save *more* in reaction to negative interest rates. Another 10 percent answered that they would spend more, whereas the rest would either shift into alternative assets or hoard cash. It should be kept in mind that these responses need not necessarily match with actual behavior if such a situation were to occur. Financial education might help reduce the impact of money illusion. Moreover, a shift to a dual domestic currency system could in itself reduce money illusion by confronting economic agents with more than one unit of account.

A dual domestic currency system can only remove the lower bound on nominal interest rates if digital currency rather than cash becomes the relevant unit of account, which may require legal and regulatory reforms. The unit of account is the currency used to value goods, services, assets, liabilities, income, expenses, and so forth. Nominal contracts and invoices are written in the unit of account. To ensure that the dual domestic currency system would work to transmit negative interest rates, citizens would have to measure their wealth and income in terms of units of digital currency, not cash. Prices would have to be quoted predominantly in digital currency; wage contracts and other important nominal contracts should be written in digital currency to ensure this. Mental accounting also should take place in the unit of account, which affects behavior and decisionmaking. The unit of account determines the currency through which monetary policy is transmitted to the economy and, therefore, the currency in which potential frictions apply.[25] As Buiter (2007) explains, if cash instead of digital currency were to become the unit of account in a dual domestic currency regime, economic agents would measure their incomes and assets in units of cash. The depreciation of cash in terms of reserves would instead be perceived as an appreciation of reserves relative to cash. This appreciation could in turn nullify the negative interest on reserves, which consequently would not transmit to the rest of the economy.

To support digital currency as the relevant unit of account, digital currency could be given legal tender status. Throughout history, legal tender served as an economy's main medium of exchange as

[25] See Buiter (2007) for a formal model of this link.

well as the unit of account.[26] Legal tender status implies that a given means of payment is recognized by the legal system to be valid for meeting financial obligations.[27] Agents usually cannot refuse legal tender for settling debts, though the obligation to accept legal tender can be abrogated based on the freedom of contract. In most countries, cash is the legal tender for historical reasons. Central bank reserves are sometimes included as well. In contrast, bank deposits—the main form of digital money that nonbank citizens currently use and have access to—are not. They are accepted in payments only by convention, for convenience, and through trust. Legal tender is usually a liability toward the central bank, whereas other means of payments such as bank transfers or credit cards constitute a liability toward the financial institution that issues them. Holding and accepting other means of payment might be convenient, but is related to incurring some creditor risk. Declaring bank deposits legal tender therefore seems problematic.

Agarwal and Kimball (2015) advocated that the legal tender status of cash be revoked in a dual currency system, leaving ordinary citizens without central bank reserve accounts without access to legal tender. To address this, central bank reserves could be made available to nonbank citizens—for example, in the form of a legal tender CBDC (Niepelt 2015, Bordo and Levin 2017, Ricks, Crawford, and Menaud 2018). Some central banks are discussing the pros and cons of issuing CBDC as a complement to cash to nonbanks; see, e.g., Mersch (2017) and Sveriges Riksbank (2018). Alternatively, banks could be required to offer special deposits that are backed by the bank's holdings of central bank reserves—a type of indirectly issued CBDC.[28] These special deposits could then be granted legal tender

[26]Buiter (2007) discusses some exceptions to this rule. Another example is the Chilean UF (Unidad de Fomento), which was introduced in 1967 and successfully used as an indexed unit of account (indexed to consumer price developments), while the pesos remained the means of exchange, means of storage, and legal tender (Shiller 2002).

[27]Throughout the history of economic thought, the notion can be found that fiat money derives its value from being accepted by the government in paying taxes; see the references given in Forstater (2005).

[28]Relatedly, in June 2018, Swiss citizens voted on whether to introduce a form of narrow banking that would go a long way toward backing bank deposits fully with central bank legal tender. The proposal was rejected; see Assenmacher and Brand (2018).

status. Once a digital, universally accessible legal tender circulates, further measures to ensure that it becomes the unit of account would be unproblematic—and perhaps not even necessary.

Regulation would (ideally) take care of how to interpret existing, or legacy, contracts written before the de-linking of cash and reserves. In this context, keeping the legal tender status of cash in a dual domestic currency regime would raise important issues related to how legacy nominal contracts should be honored when these do not specify whether payment should be made in either cash or digital means of payment. If cash is depreciating but can be used to make good on a contractual agreement, this would create incentives for debtors to repay debts in depreciated units of cash, preventing digital money from becoming the relevant unit of account and leading to an unintended redistribution from creditors to debtors in the new regime. Legal uncertainty with respect to which currency a contract refers to can create legal problems and frictions in the transition to a dual currency system. A shift to a new regime in which cash remains legal tender would therefore have to include amendments to legal frameworks governing contracts and payments on financial obligations. Such transitional issues would need to be planned for to minimize disruptions when introducing a dual domestic currency system.

Another key issue is how the public understands and reacts to the introduction of a dual domestic currency system. Given the lack of precedents for systems allowing for deeply negative interest rates and a de-linking of the value of cash from digital money, the behavioral responses during the transition to a new system would be difficult to predict. In theory, it should be clear that a CRC rate above unity means that cash depreciates over time. In practice, however, if public education efforts about the system are not successful, citizens (especially those that are less financially literate) might initially think that they get more cash value for their deposits when deposits bear negative interest. If cash prices do not increase immediately upon the introduction of the new system, it may add to the illusion that cash retains its purchasing power over time. This could initially lead to a run into cash until it is broadly understood that cash can be redeposited only at a depreciating rate. While such problems should be temporary, they could create practical problems for central banks, e.g., they could temporarily run out of cash. Preparations for a smooth transition should hence include clear communication and large precautionary stocks of cash, as well as taking measures to improve the level of financial education of the population.

Finally, the introduction of a dual domestic currency system could lead people to switch to other forms of currency for their payments, such as foreign currency, gold, or even cryptocurrency. Mechanically, such substitution would lead to a depreciation of the domestic currency and higher inflation, potentially stimulating demand domestically and from abroad. Overall, it seems unlikely that individuals would entirely abandon domestic currency as a means of payments for most transactions. Based on data from five hyperinflations, Barro (1972) showed that even at rates of inflation above 100 percent annually, the domestic currency was still used, though the velocity increased substantially. The introduction of a dual domestic currency system would not change the fact that only domestic currency is legal tender and needed for making good on various obligations. To establish trust in connection with the introduction of a dual domestic currency system, the central bank would have to communicate the system and its merits well and carefully. This is perhaps the most important and also the most challenging part of introducing such a system. Moderately negative nominal rates have been deeply unpopular in some countries. What would successful communication look like? It would create confidence that monetary policy has new and unlimited room to address a downturn. It would thereby reduce any existing crisis sentiments that are likely to be prevalent in a situation where a central bank would want to introduce such a system. Transparency and the quality of communication of the system would be key for building trust.[29]

In conclusion, we see no reasons to anticipate that monetary policy transmission to savings and investment behavior would be hampered with deeply negative interest rates. The main uncertainty

[29] A related observation is that there is no reason that a large-scale capital flight would have different implications in a dual domestic currency system. We would expect to see a depreciation of the domestic currency in a floating rate system or an expansion of central bank reserves through foreign exchange interventions in a fixed exchange rate system (both of which should help bring back inflation faster). The central bank would have to be prepared in terms of foreign exchange reserve adequacy or macroprudential measures to prevent balance sheet vulnerabilities to such episodes.

is the unknown effects of deeply nominal rates due to money illusion. Associated risks should be addressed with ample and appropriate communication and education efforts.

5. Conclusions

The zero lower bound on nominal interest rates is caused by the availability of cash that yields a zero nominal return. De-linking cash from digital currency and making cash depreciate relative to digital currency, as proposed by Buiter (2007) and Agarwal and Kimball (2015), could solve this problem. With such a system in place, a central bank would be able to use conventional monetary policy tools to stabilize the economy without being constrained by the lower bound. In a world of low neutral real interest rates, it would help reduce the length of business cycle downturns and hence the duration of low interest rate episodes. It would reduce the risk of deflationary spirals and the incidence of secular stagnation. It would do so without dispensing with cash. Studies that question the transmission of negative rates to bank lending assume that banks cannot lower their deposit rates, which would not be a constraint in a dual domestic currency system.

Our discussion suggests that the system is technically feasible and would not require fundamental changes to current operating frameworks of central banks. Moreover, in contrast to some other proposals, the system would be fully reversible. After a sufficient normalization of economic conditions, it could be exited if so desired. It would work in economies with fixed or flexible exchange rate systems and could be implemented unilaterally. Communication and financial education would be central for a successful introduction of such a regime and should address any possible risks to the transmission of monetary policy and to financial stability. Further work would be needed to identify, prepare, and implement the necessary legal reforms for ensuring its effective operation.

In our view, the dual domestic currency system should be considered alongside alternative proposals for keeping monetary policy effective at low interest rates, such as phasing out cash all together (Rogoff 2014), a higher inflation target proposed by Blanchard, Dell'Ariccia, and Mauro (2010), or the use of unconventional easing measures such as quantitative easing and forward guidance. All current proposals, including the status quo, have pros and cons which will depend on specific country circumstances. In comparison with alternatives, the dual domestic currency system has the advantage of completely freeing monetary policy from a lower bound while being neutral for banks' business models and their role in monetary policy transmission, allowing for effectively redressing and hence shortening the duration of recessions. Raising the inflation target and unconventional easing measures do not remove the lower bound but shift it downward by some percentage points (Ball et al. 2016). Another advantage is that the dual domestic currency system can be implemented as a crisis measure, ideally with some preparation beforehand, while preserving a role for cash. In contrast, to reap all benefits of raising the inflation target, it would require time for expectations to adjust and credibility to be built around a higher level of inflation.[30] Other advantages include the dual domestic currency system's reversibility, its preservation of a role for cash, and the fact that its introduction would reconfirm the central bank's commitment to the inflation target rather than raise doubts about it. But the dual domestic currency system clearly also has disadvantages. Most importantly, it would be an enormous communicational challenge. It would also require more far-reaching changes to the financial and legal system than simply raising the inflation target or pursuing quantitative easing.

Technological innovation in digital payments systems is proceeding at a rapid pace, without central banks actively promoting this (Casey et al. 2018, Brunnermeier, James, and Landau 2019). Such changes may force a reconsideration of issues around cash and legal tender in the future, irrespective of ZLB considerations. In this context, new developments should also be evaluated in light of their ability to accommodate negative policy rates or even a dual domestic currency system. Pros and cons of a dual domestic currency system as well as alternative solutions should be carefully compared in the

[30]If the inflation target is raised as a crisis measure, it works through expectations of future inflation and is akin to forward guidance. This requires strong credibility, and there are limits to the additional firepower that can be achieved, as evidenced by the Japanese experience; see also Ball et al. (2016).

context of a country's institutional, legal, cultural, and economic situation when considering the future of monetary policy frameworks.

References

Agarwal, R., and M. Kimball. 2015. "Breaking Through the Zero Lower Bound." IMF Working Paper No. 15/224.

———. 2019. "Enabling Deep Negative Rates to Fight Recessions: A Guide." IMF Working Paper No. 19/84.

Albertazzi, U., A. Nobili, and F. M. Signoretti. 2017. "The Bank Lending Channel of Conventional and Unconventional Monetary Policy." Temi di Discussione No. 1094, Banca d'Italia.

Altavilla, C., M. Boucinha, and J.-L. Peydró. 2018. "Monetary Policy and Bank Profitability in a Low Interest Rate Environment." *Economic Policy* 33 (96): 531–86.

Altavilla, C., L. Burlon, M. Giannetti, and S. Holton. 2019. "Is there a Zero Lower Bound? The Effects of Negative Policy Rates on Banks and Firms." ECB Working Paper No. 2289.

Anderson, A. G., and J. Kandrac. 2016. "Monetary Policy Implementation and Private Repo Displacement: Evidence from the Overnight Reverse Repurchase Facility." Finance and Economics Discussion Series No. 2016-096, Board of Governors of the Federal Reserve System.

Andolfatto, D. 2018. "Assessing the Impact of Central Bank Digital Currency on Private Banks." Working Paper No. 2018-026B, Federal Reserve Bank of St. Louis.

Assenmacher, K., and C. Brand. 2018. "The Swiss Sovereign Money Initiative." *Credit and Capital Markets* 51 (4): 621–44.

Ball, L., P. Honohan, J. Gagnon, and S. Krogstrup. 2016. "What Else Can Central Banks Do?" Geneva Report on the World Economy No. 18, ICMB and CEPR.

Bank for International Settlements. 2018. "Central Bank Digital Currencies." CPMI Paper No. 174.

Barro, R. J. 1972. "Inflationary Finance and the Welfare Cost of Inflation." *Journal of Political Economy* 80 (5): 987–1001.

Bech, M., U. Faruqui, F. Ougaard, and C. Picillo. 2018. "Payments Are A-changin' but Cash Still Rules." *BIS Quarterly Review* (March): 67–80.

Bech, M., and A. Malkhozov. 2016. "How Have Central Banks Implemented Negative Policy Rates?" *BIS Quarterly Review* (March): 31–44.

Blanchard, O., G. Dell'Ariccia, and P. Mauro. 2010. "Rethinking Macroeconomic Policy." IMF Staff Position Note No. 2010/03.

Bordo, M. D., and A. T. Levin. 2017. "Central Bank Digital Currency and the Future of Monetary Policy." NBER Working Paper No. 23711.

Brunnermeier, M. K., H. James, and J.-P. Landau. 2019. "The Digitalization of Money." NBER Working Paper No. 26300.

Brunnermeier, M. K., and Y. Koby. 2019. "The Reversal Interest Rate." Mimeo, Princeton University. https://scholar.princeton.edu/sites/default/files/markus/files/25b_reversalrate.pdf.

Buiter, W. H. 2007. "Is Numérairology the Future of Monetary Economics? Unbundling Numéraire and Medium of Exchange through a Virtual Currency and a Shadow Exchange Rate." *Open Economies Review* 18 (2): 127–56.

Casey, M., J. Crane, G. Gensler, S. Johnson, and N. Narula. 2018. "The Impact of Blockchain Technology on Finance: A Catalyst for Change?" Geneva Report on the World Economy No. 21, ICMB and CEPR.

Cheney, J. S. 2006. "The Role of Digital Payments in Disaster Recovery: Providing More than Convenience." Conference Summary, Federal Reserve Bank of Philadelphia.

Chiu, J., M. Davoodalhosseini, J. Jiang, and Y. Zhu. 2019. "Bank Market Power and Central Bank Digital Currency: Theory and Quantitative Assessment." Staff Working Paper No. 2019-20, Bank of Canada.

Christensen, J. H. E. 2019. "Yield Curve Responses to Introducing Negative Policy Rates." FRBSF Economic Letter No. 2019-27, Federal Reserve Bank of San Francisco.

Cliffe, M. 2016. "Negative Rates, Negative Reactions." VoxEU.org, February 26. https://voxeu.org/article/negative-rates-negative-reactions.

Cúrdia, V. 2019. "How Much Could Negative Rates Have Helped the Recovery?" FRBSF Economic Letter No. 2019-04, Federal Reserve Bank of San Francisco.

De Grauwe, P., L. Rinaldi, and P. Van Caysele. 2006. "Issues of Efficiency in the Use of Cash and Cards." University of Leuven. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.629.5066.

Demiralp, S., J. Eisenschmidt, and T. Vlassopoulos. 2019. "Negative Interest Rates, Excess Liquidity and Retail Deposits: Banks' Reaction to Unconventional Monetary Policy in the Euro Area." ECB Working Paper No. 2283.

Eggertson, G. B., R. E. Juelsrud, and E. G. Wold. 2017. "Are Negative Nominal Interest Rates Expansionary?" NBER Working Paper No. 24039.

Eisenschmidt, J., and F. Smets. 2019. "Negative Interest Rates: Lessons from the Euro Area." In *Monetary Policy and Financial Stability: Transmission Mechanisms and Policy Implications,* ed. A. Aguirre, M. Brunnermeier, and D. Saravia, 13–42 (chapter 2). Volume 26 in Central Banking Analysis and Economic Policies Book Series. Santiago: Central Bank of Chile.

Engert, W., B. S. C. Fung, and S. Hendry. 2018. "Is a Cashless Society Problematic?" Staff Discussion Paper No. 2018-12, Bank of Canada.

Esselink, H., and L. Hernández. 2017. "The Use of Cash by Households in the Euro Area." ECB Occasional Paper No. 201.

Fehr, E., and J. R. Tyran. 2001. "Does Money Illusion Matter?" *American Economic Review* 91 (5): 1239–62.

Forstater, M. 2005. "Tax-Driven Money: Additional Evidence from the History of Economic Thought, Economic History and Economic Policy." In *Complexity, Endogenous Money and Macroeconomic Theory*, ed. M. Setterfield, 202–20 (chapter 11). Cheltenham: Edward Elgar.

Friedman, B. M. 1999. "The Future of Monetary Policy: The Central Bank as an Army with Only a Signal Corps?" *International Finance* 2 (3): 321–38.

Gagnon, J. E., and B. Sack. 2014. "Monetary Policy with Abundant Liquidity: A New Operating Framework for the Federal Reserve." Policy Brief No. 14-4, Peterson Institute for International Economics.

Gesell, S. 1916. "Die Natürliche Wirtschaftsordnung Durch Freiland und Freigeld." Selbstverlag, Les Hauts Geneveys 1916; 9. Auflage herausgegeben von Karl Walker: Rudolf Zitzmann Verlag, Lauf 1949.

Goodfriend, M. 2000. "Overcoming the Zero Lower Bound on Interest Rate Policy." *Journal of Money, Credit and Banking* 32 (4, Part 2): 1007–35.

———. 2016. "The Case for Unencumbering Interest Rate Policy at the Zero Bound." Paper presented at the Federal Reserve Bank of Kansas Economic Symposium, Jackson Hole, Wyoming, August 26–27.

Heider, F., F. Saidi, and G. Schepens. 2017. "Life Below Zero: Bank Lending under Negative Policy Rates." SSRN-id2788204.

International Monetary Fund. 2016. "Global Financial Stability Report: Fostering Stability in a Low-Growth, Low-Rate Era." (October).

———. 2017. "Negative Interest Rate Policies — Initial Experiences and Assessments." IMF Policy Paper.

Jackson, H. 2015. "The International Experience with Negative Policy Rates." Staff Discussion Paper No. 2015-13, Bank of Canada.

Jobst, A., and H. Lin. 2016. "Negative Interest Rate Policy (NIRP): Implications for Monetary Transmission and Bank Profitability in the Euro Area." IMF Working Paper No. 16/172.

Keister, T., and D. R. Sanches. 2019. "Should Central Banks Issue Digital Currency?" Working Paper No. 19-26, Federal Reserve Bank of Phaledelphia.

Khiaonarong, T., and D. Humphrey. 2019. "Cash Use Across Countries and the Demand for Central Bank Digital Currency." IMF Working Paper No. 19/46.

Krogstrup, S. 2017. "Monetary Policy Accommodation at the Lower Bound." *Business Economics* 52 (1): 7–14.

Lucas, A., J. Schaumburg, and B. Schwaab. 2017. "Bank Business Models at Zero Interest Rates." ECB Working Paper No. 2084.

Mancini Griffoli, T., M. Soledad Martinez Peria, I. Agur, A. Ari, J. Kiff, A. Popescu, and C. Rochon. 2018. "Casting Light on Central Bank Digital Currency." IMF Staff Discussion Note No. 18/08.

Mankiw, G. 2009. "It May Be Time for the Fed to Go Negative." *New York Times* (April 18).

McAndrews, J. 2015. "Negative Nominal Central Bank Policy Rates: Where Is the Lower Bound?" Remarks at the University of Wisconsin, May 8. http://www.ny.frb.org/newsevents/speeches/2015/mca150508.html.

Meaning, J., B. Dyson, J. Barker, and E. Clayton. 2018. “Broadening Narrow Money: Monetary Policy with a Central Bank Digital Currency.” Staff Working Paper No. 724, Bank of England.

Mersch, Y. 2017. “Digital Base Money: An Assessment from the ECB’s Perspective.” Speech given at the Farewell Ceremony for Pentti Hakkarainen, Deputy Governor of Suomen Pankki, Helsinki, January 16. https://www.ecb.europa.eu/press/key/date/2017/html/sp170116.en.html.

Niepelt, D. 2015. “Reserves for Everyone — Towards a New Monetary Regime?” Vox-EU Column, January 21.

Pfister, C., and N. Valla. 2018. “‘New Normal’ or ‘New Orthodoxy’? Elements of a Central Banking Framework for the After-Crisis.” Working Paper No. 680, Banque de France.

Quinn, S., and W. Roberds. 2014. “How Amsterdam Got Fiat Money.” *Journal of Monetary Economics* 66 (September): 1–12.

Reserve Bank of Zimbabwe. 2015. “Demonetization of the Zimbabwe Dollar.” Press Statement, June 9. https://www.rbz.co.zw/documents/publications/press/demonetisation-press-statement-9-june-2015.pdf.

———. 2016. “The Introduction of Bond Notes.” Press Statement, November 26. https://www.rbz.co.zw/documents/publications/press/press-statement-on-the-introduction-of-bond-notes.pdf.

Ricks, M., J. Crawford, and L. Menand. 2018. “FedAccounts: Digital Dollars.” Vanderbilt Law Research Paper No. 18-33; UC Hastings Research Paper No. 287. https://ssrn.com/abstract=3192162.

Rogoff, K. 2014. “Costs and Benefits to Phasing Out Paper Currency.” NBER Working Paper No. 20126.

Seltmann, T. 2010. “Umlaufsicherung von Banknoten.” *Humane Wirtschaft* 02/2010: 10–16.

Shiller, R. 2002. “Indexed Units of Account: Theory and Assessment of Historical Experience.” In *Indexation, Inflation and Monetary Policy*, ed. F. Lefort and K. Schmidt-Hebbel, 105–34. Santiago: Central Bank of Chile.

Summers, L. 2018. “Monetary Policy in a Low Inflation, Low Interest Rate World.” In *Price and Wage-setting in Advanced Economies*, ed. European Central Bank, 27–31. Conference Proceedings, ECB Forum on Central Banking, Sintra, Portugal, June 16–18.

Svensson, L. E. O. 2003. "Escaping from a Liquidity Trap and Deflation: The Foolproof Way and Others." *Journal of Economic Perspectives* 17 (4): 145–66.

Sveriges Riksbank. 2018. "The Riksbank's e-krona Project: Report 2."

Swiss National Bank. 2017. "Survey on Payment Methods." https://www.snb.ch/en/mmr/reference/paytrans_survey_report_2017/source/paytrans_survey_report_2017.en.pdf.

Woodford, M. 2000. "Monetary Policy in a World without Money." *International Finance* 3 (2): 229–60.

Modeling the Consumption Response to the CARES Act*

Christopher D. Carroll,[a] Edmund Crawley,[b]
Jiri Slacalek,[c] and Matthew N. White[d]
[a]Johns Hopkins University
[b]Federal Reserve Board
[c]European Central Bank
[d]University of Delaware

To predict the effects of the 2020 U.S. CARES Act on consumption, we extend a model that matches responses to past consumption stimulus packages. The extension allows us to account for two novel features of the coronavirus crisis. First, during lockdowns, many types of spending are undesirable or impossible. Second, some of the jobs that disappear during the lockdown will not reappear. We estimate that, if the lockdown is short-lived (the median point of view as we are writing in April 2020), the combination of expanded unemployment insurance benefits and stimulus payments should be sufficient to allow a swift recovery in consumer spending to pre-crisis levels. If the lockdown lasts longer (or there is a "second wave"), an extension of enhanced unemployment benefits will likely be necessary for consumption spending to recover quickly.

JEL Codes: D83, D84, E21, E32.

*The first public version of this paper appeared April 15, 2020. A publicly available preprint appeared in the Centre for Economic Policy Research journal *Covid Economics* on April 27, 2020. Thanks to the Consumer Financial Protection Bureau for funding the original creation of the Econ-ARK toolkit, whose latest version we used to produce all the results in this paper; and to the Sloan Foundation for funding Econ-ARK's extensive further development that brought it to the point where it could be used for this project. We are grateful to Kiichi Tokuoka, who provided valuable feedback and input as this project progressed, to Mridul Seth, who created the dashboard and configurator, and to Luc Laeven, who swiftly handled our submission to IJCB. The views presented in this paper are those of the authors and should not be attributed to the Federal Reserve Board or the European Central Bank. Author contact information: Carroll: http://econ.jhu.edu/people/ccarroll/, ccarroll@jhu.edu. Crawley: http://edmundcrawley.com. edmund.s.crawley@frb.gov. Slacalek: http://www.slacalek.com/, jiri.slacalek@ecb.europa.eu. White: mnwecon@udel.edu.

"Economic booms are all alike; each recession contracts output in its own way." — with apologies to Leo Tolstoy

1. Introduction

In the decade since the Great Recession, macroeconomics has made great progress by insisting that models be consistent with microeconomic evidence (see Krueger, Mitman, and Perri 2016 in the *Handbook of Macroeconomics* for a survey). To predict the effects of the 2020 CARES Act (Coronavirus Aid, Relief, and Economic Security) on consumption, we take, from this new generation, one model that is specifically focused on reconciling apparent conflicts between micro and macro evidence about consumption dynamics,[1] and adapt it to incorporate two aspects of the coronavirus crisis.

First, because the tidal wave of layoffs for employees of shuttered businesses will have a large impact on their income and spending, assumptions must be made about the employment dynamics of laid-off workers. Specifically, the unemployed in our model consist of two categories: normal unemployed and deeply unemployed. Similar to a normal recession, the normal unemployed will be able to quickly return to their old jobs (or similar ones). However, in addition, some people become deeply unemployed, facing a more persistent unemployment shock. This feature reflects the fact that some kinds of jobs will not come back quickly after the lockdown, and that people who worked in these sectors will have more difficulty finding a new job.[2]

On the second count, we model the restricted spending options by assuming that spending during the lockdown is less enjoyable (there is a negative shock to the "marginal utility of consumption.") Based on a tally of sectors that we judge to be substantially shuttered during the "lockdown," we calibrate an 11 percent reduction to spending. Thus households will prefer to defer some of their consumption into the future, when it will yield them greater utility.

[1] This was articulated long ago by Deaton (1992) and documented recently by Havranek, Rusnak, and Sokolova (2017).

[2] The cruise industry, for example, is likely to take a long time to recover. Demand for airline travel is expected to remain depressed, with the International Air Traffic Association projecting that passenger travel will not return to pre-pandemic levels until 2024.

(See Cox et al. 2020, Carvalho et al. 2020, and Andersen et al. 2020 showing a strong effect of this kind in U.S., Spanish, and Danish data, respectively).[3]

Our model captures the two primary features of the CARES Act that aim to bolster consumer spending:

(i) The boost to unemployment insurance benefits, amounting to $7,800 if unemployment lasts for 13 weeks.

(ii) The direct stimulus payments to most households, up to $1,200 per adult.

We estimate that the combination of expanded unemployment insurance benefits and stimulus payments should be sufficient to expect a swift recovery in consumer spending to its pre-crisis levels under our default description of the pandemic, in which the lockdown ends after two quarters on average. Overall, unemployment benefits account for about 30 percent of the total aggregate consumption response, and stimulus payments explain the remainder.

Our analysis partitions households into three groups based on their employment state when the pandemic strikes and the lockdown begins.

First, households in our model who do not lose their jobs initially build up their savings, both because of the lockdown-induced suppression of spending and because most of these households will receive a significant stimulus check, much of which the model says will be saved. Even without the lockdown, we estimate that only about 20 percent of the stimulus money would be spent immediately upon receipt, consistent with evidence from prior stimulus packages about spending on nondurable goods and services. Once the lockdown ends, the spending of the households that remained employed at the onset of the pandemic rebounds strongly thanks to their healthy household finances.

The second category of households is the "normal unemployed," job losers who perceive that it is likely they will be able to resume

[3] A shock to marginal utility may not perfectly capture the essence of what depresses consumption spending, but it accomplishes our purposes and is a kind of shock commonly studied in the literature. Any analysis of the welfare consequences of the lockdown would probably need a richer treatment to be credible.

their old job (or get a similar new job) when the lockdown is over. Our model predicts that the CARES Act will be particularly effective in stimulating their consumption, given the perception that their income shock will be largely transitory. Our model predicts that by the end of 2021, the spending of this group recovers to the level it would have achieved in the absence of the pandemic ("baseline"); without the CARES Act, this recovery would take more than a year longer.

Finally, for households in the "deeply unemployed" category, our model says that the marginal propensity to consume (MPC) from the checks will be considerably smaller, because they know they must stretch that money for longer. Even with the stimulus from the CARES Act, we predict that consumption spending for these households will not fully recover until the middle of 2023. Even so, the Act makes a big difference to their spending, particularly in the first six quarters after the crisis. For both groups of unemployed households, the effect of the stimulus checks is dwarfed by the increased unemployment benefits, which arrive earlier and are much larger (per recipient).

Perhaps surprisingly, we find that the effectiveness of the combined stimulus checks and unemployment benefits package for aggregate consumption is not substantially different from a package that distributed the same quantity of money equally among households. The reason for this is twofold: first, the extra unemployment benefits in the CARES Act are generous enough that many of the "normally unemployed" remain financially sound and can afford to save a good portion of those benefits; second, the deeply unemployed expect their income to remain depressed for some time and therefore save more of the stimulus for the future. In the model, the fact that they do *not* spend immediately is actually a reflection of how desperately they anticipate these funds will be needed to make it through a long period of low income. While unemployment benefits do not strongly stimulate current consumption of the deeply unemployed, they do provide important disaster relief for those who may not be able to return to work for several quarters (see Krugman 2020 for an informal discussion).

In addition to our primary scenario's relatively short lockdown period, we also consider a more severe scenario in which the lockdown is expected to last for four quarters and the unemployment rate increases to 20 percent. In this case, we find that the return of spending toward its baseline path takes roughly three years. Moreover, the spending of deeply unemployed households falls steeply unless the temporary unemployment benefits in the CARES Act are extended for the duration of the lockdown.

Our modeling assumptions—about who will become unemployed, how long it will take them to return to employment, and the direct effect of the lockdown on consumption utility—could prove to be off, in either direction. Reasonable analysts may differ on all of these points and prefer a different calibration. To encourage such exploration, we have made available our modeling and prediction software, with the goal of making it easy for fellow researchers to test alternative assumptions. Instructions for installing and running our code can be found at https://github.com/econ-ark/Pandemic#reproduction-instructions; alternatively, adjustments to our parameterization can be explored with an interactive dashboard at http://econ-ark.org/pandemicdashboard.

There is a potentially important reason our model may underpredict the bounceback in consumer spending when the lockdown ends: "pent-up demand." This term captures the fact that purchases of "durable" goods can be easily postponed, but that when the reason for postponement abates, some portion of the missing demand is made up for.[4] For simplicity, our model does not include durable goods, because modeling spending on durables is a formidable challenge. But it is plausible that, when the lockdown ends, people may want to spend *more* than usual on memorable or durable goods to make up for what they did not spend earlier.

Many papers have recently appeared on the economic effects of the pandemic and policies to manage it. Several papers combine the classic susceptible–infected–recovered (SIR) epidemiology model with dynamic economic models to study the interactions between health and economic policies (Alvarez, Argente, and Lippi 2020 and Eichenbaum, Rebelo, and Trabandt 2020, among others). Guerrieri et al. (2020) shows how an initial supply shock (such as a pandemic) can be amplified by the reaction of aggregate demand. The ongoing work of Kaplan, Moll, and Violante (2020) allows for realistic

[4]We put "durable" in quotes because "memorable" goods (Hai, Krueger, and Postlewaite 2013) have effectively the same characteristics.

household heterogeneity in how household income and consumption are affected by the pandemic. Glover et al. (2020) studies distributional effects of optimal health and economic policies. Closest to our paper is some work analyzing the effects of the fiscal response to the pandemic, including Faria-e-Castro (2020b) in a two-agent dynamic stochastic general equilibrium (DSGE) model, and Bayer et al. (2020) in a HANK (heterogeneous agent New Keynesian) model.

All of this work accounts for general equilibrium effects on consumption and employment, which we omit, but none of it is based on a modeling framework explicitly constructed to match micro and macroeconomic effects of past stimulus policies, as ours is.

A separate strand of work focuses on empirical studies of how the economy reacts to pandemics; see, e.g., Baker et al. (2020), Casado et al. (2020), Chetty et al. (2020), Coibion, Gorodnichenko, and Weber (2020), Correia, Luck, and Verner (2020), Garner, Safir, and Schild (2020), and Jordà, Singh, and Taylor (2020).

2. Modeling Setup

2.1 The Baseline Model

Our model extends a class of models explicitly designed to capture the rich empirical evidence on heterogeneity in the MPC across different types of household (employed, unemployed; young, old; rich, poor). This is motivated by the fact that the act distributes money unevenly across households, particularly targeting unemployed households. A model that does not appropriately capture both the degree to which the stimulus money is targeted and the differentials in responses across differently targeted groups is unlikely to produce believable answers about the spending effects of the stimulus.

Specifically, we use a lifecycle model calibrated to match the income paths of high-school dropouts, high-school graduates, and college graduates.[5] Within each of these groups, we calibrate the distribution of discount factors to match their distribution of liquid assets. Matching the distributions of liquid assets allows us to achieve

[5] The baseline model is very close to the lifecycle model in Carroll et al. (2017).

a realistic distribution of marginal propensities to consume according to education group, age, and unemployment status, and thus to assess the impact of the act for these different groups.[6] Households are subject to permanent and transitory income shocks, as well as unemployment spells.[7]

2.2 Adaptations to Capture the Pandemic

To model the pandemic, we add two new features to the model.

First, our new category of "deeply unemployed" households was created to capture the likelihood that the pandemic will have long-lasting effects on some kinds of businesses and jobs (e.g., the cruise and airline industries), even if the CARES Act manages to successfully cushion much of the initial financial hit to total household income. Moreover, evidence in Yagan (2019) indicates that unemployment shocks from the Great Recession had long-lasting impacts on individuals' employment.

Each quarter, our "deeply unemployed" households have a two-thirds chance of remaining deeply unemployed, and a one-third chance of becoming "normal unemployed." The expected time to reemployment for a "deeply unemployed" household is four-and-a-half quarters, much longer than the historical average length of a typical unemployment spell. Reflecting recent literature on the "scarring effects" of unemployment spells (e.g., Oreopoulos, von Wachter, and Heisz 2012 and Heathcote, Perri, and Violante 2020), permanent income of both "normal" and "deeply" households declines by 0.5 percent each year due to "skill rot" (relative to following the default age profile that would have been followed if the consumer had remained employed).

Second, a temporary negative shock to the marginal utility of consumption captures the idea that, during the period of the pandemic, many forms of consumption are undesirable or even impossible.[8]

---

[6]For a detailed description of the model and its calibration, see appendix A.

[7]Households exit unemployment with a fixed probability each quarter—the expected length of an unemployment spell is one-and-a-half quarters.

[8]For the purposes of our paper, with log utility, modeling lockdowns as a shock to marginal utility is essentially equivalent to not allowing consumers to

The pandemic is modeled as an unexpected (MIT) shock, sending many households into normal or deep unemployment, as well as activating the negative shock to marginal utility. Households understand and respond in a forward-looking way to their new circumstances (according to their beliefs about its duration), but their decisions prior to the pandemic did not account for any probability that it would occur. For simplicity, we assume that each household correctly recognizes whether it is "deeply" or "normal" unemployed and reacts accordingly.

2.2.1 Calibration

The calibration choices for the pandemic scenario are very much open for debate. We have tried to capture something like median expectations from early analyses, but there is considerable variation in points of view around those medians. Section 2.3 below presents a more adverse scenario with a longer lockdown and a larger increase in unemployment.

Unemployment forecasts for 2020:Q2 range widely, from less than 10 percent to more than 30 percent, but all point to an unprecedented sudden increase in unemployment.[9] We choose a total unemployment rate in 2020:Q2 of just over 15 percent, consisting of 5 percent "deeply unemployed" and 10 percent "normal unemployed" households.

Our model assumes that the unemployment shock from the pandemic is a singular event, with no change in the longer-run job separation rate for employed households (calibrated to generate a steady-state unemployment rate of 5 percent). Consequently, agents

buy a subset of goods (which are combined into composite consumption by a Cobb-Douglas aggregator). However, the two approaches would yield different implications for normative evaluations of economic policies.

[9] As of April 16, about 22 million new unemployment claims have been filed in four weeks, representing a loss of over 14 percent of total jobs. JPMorgan Global Research forecast 8.5 percent unemployment (JP Morgan 2020, from March 27); Treasury Secretary Steven Mnuchin predicted unemployment could rise to 20 percent without a significant fiscal response (Bloomberg 2020a); Federal Reserve Bank of St. Louis President James Bullard said the unemployment rate may hit 30 percent (Bloomberg 2020b—see Faria-e-Castro 2020a for the analysis behind this claim). Based on a survey that closely follows the Current Population Survey, Bick and Blandin (2020) calculate a 20.2 percent unemployment rate at the beginning of April.

in our model who remain employed in 2020:Q2 have no additional precautionary saving motive against a heightened risk of unemployment, and any change in their consumption behavior arises from the marginal utility shock.

We calibrate the likelihood of becoming unemployed to match empirical facts about the relationship of unemployment to education level, permanent income, and age, which is likely to matter because the hardest hit sectors skew young and unskilled.[10] Figure 1 shows our assumptions on unemployment along these dimensions. In each education category, the solid or dashed line represents the probability of unemployment type ("normal" or "deep") for a household with the median permanent income at each age, while the dotted lines represent the probability of unemployment type for a household at the 5th and 95th percentile of permanent income at each age; appendix A and table A.2 detail the parameterization and calibration we used.

To calibrate the drop in marginal utility, we estimate that 10.9 percent of the goods that make up the consumer price index become highly undesirable, or simply unavailable, during the pandemic: food away from home, public transportation including airlines, and motor fuel. As we use a coefficient of risk aversion equal to one, we simply multiply utility from consumption during the period of the epidemic by a factor of 0.891.[11] This calibration is in line with recent evidence in Chetty et al. (2020) and Cox et al. (2020). Furthermore, we choose a one-half probability of exiting the period of lower marginal utility each quarter, accounting for the possibility of a "second wave" if restrictions are lifted too early—see Cyranoski (2020).[12]

2.2.2 The CARES Act

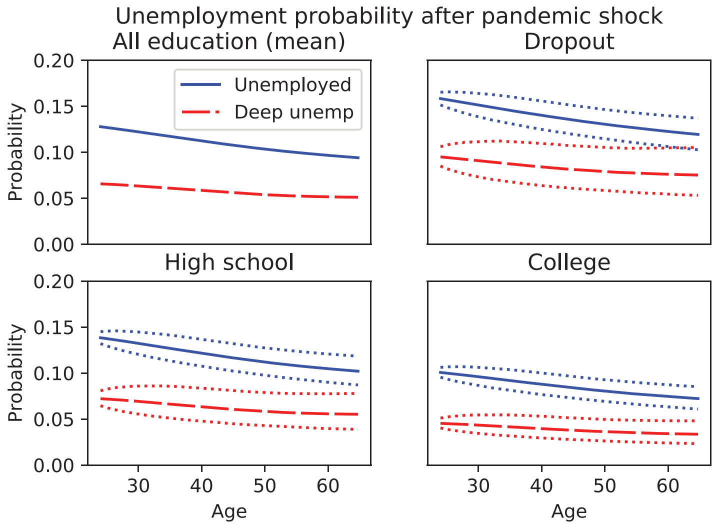
We model the two elements of the CARES Act that directly affect the income of households (see also table A.3):

[10]See Adams-Prassl et al. (2020), Gascon (2020), and Leibovici and Santacreu (2020) for breakdowns of which workers are at most risk of unemployment from the crisis. See additional evidence in Kaplan, Moll, and Violante (2020) and modeling of implications for optimal policies in Glover et al. (2020).

[11]See the Cobb-Douglas interpretation in appendix C.

[12]The Congressional Budget Office expects social distancing to last for three months, and predicts it to have diminished, on average and in line with our calibration, by three-quarters in the second half of the year; see Swagel (2020).

Figure 1. Unemployment Probability in 2020:Q2 by Demographics



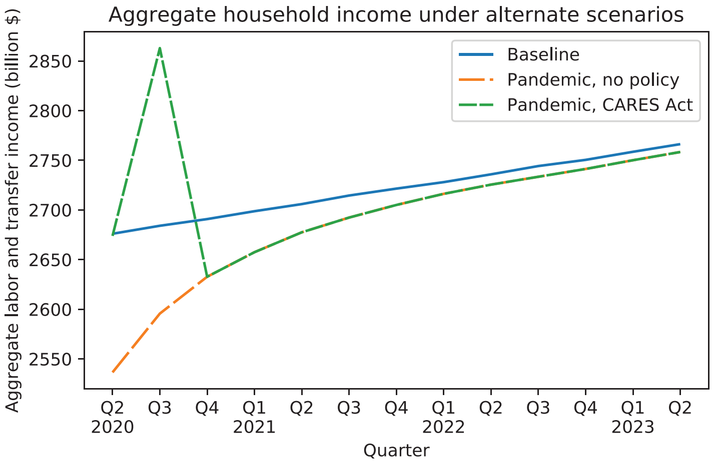
(i) The stimulus check of $1,200 for every adult taxpayer, means tested for previous years' income.[13]

(ii) The extra unemployment benefits of $600 for up to 13 weeks, a total of $7,800. For normal unemployed, we assume they receive only $5,200 to reflect the idea that they may not be unemployed the entire 13 weeks.

We model the stimulus checks as being announced at the same time as the crisis hits. However, only a quarter of households change their behavior immediately at the time of announcement, as calibrated to past experience. The remainder do not respond until their stimulus check arrives, which we assume happens in the following quarter. The households that pay close attention to the announcement of the policy are assumed to be so forward looking that they

[13]The act also includes $500 for every child. In the model, an agent is somewhere between a household and an individual. While we do not model the $500 payments to children, we also do not account for the fact that some adults will not receive a check. In aggregate, we are close to the Joint Committee on Taxation's estimate of the total cost of the stimulus checks.

Figure 2. Labor and Transfer Income



act as though the payment will arrive with certainty next period; the model even allows them to borrow against it if desired.[14]

The extra unemployment benefits are assumed to both be announced and arrive at the beginning of the second quarter of 2020, and we assume that there is no delay in the response of unemployed households' consumption to these benefits.

Figure 2 shows the path of labor income—exogenous in our model—in the baseline and in the pandemic, both with and without the CARES Act. Income in 2020:Q2 and 2020:Q3 is substantially boosted (by around 10 percent) by the extra unemployment benefits and the stimulus checks. After two years, aggregate labor income is almost fully recovered. See below for a brief discussion of analyses that attempt to endogenize labor supply and other equilibrium variables.

[14]See Carroll et al. (2020) for a detailed discussion of the motivations behind this way of modeling stimulus payments, and a demonstration that this model matches the empirical evidence of how and when households have responded to stimulus checks in the past—see Parker et al. (2013), Broda and Parker (2014), and Parker (2017), among others. See also Fagereng, Holm, and Natvik (2017) for a natural experiment measured using national registry data.

3. Results

This section presents our simulation results for the scenario described above. In addition, we then model a more pessimistic scenario with a longer lockdown and higher initial unemployment rate.

3.1 Short-Lived Pandemic

Figure 3 shows three scenarios for quarterly aggregate consumption: (i) the baseline with no pandemic; (ii) the pandemic with no fiscal response; (iii) the pandemic with both the stimulus checks and extended unemployment benefits in the CARES Act. The pandemic reduces consumption by 10 percentage points in 2020:Q2 relative to the baseline.

Without the CARES Act, consumption remains depressed through to the second half of 2021, at which point spending returns to the baseline level as a result of the buildup of liquid assets during the pandemic by households that do not lose their income. We capture the limited spending options during the lockdown period by a reduction in the utility of consumption, which makes households save more during the pandemic than they otherwise would have, with the result that they build up liquid assets. When the lockdown ends, the pent-up savings of the always employed become available to finance a resurgence in their spending, but the depressed spending of the two groups of unemployed people keeps total spending below the baseline until most of them are reemployed, at which point their spending (mostly) recovers while the always employed are still spending down their extra savings built up during the lockdown.

Figure 4 decomposes the effect of the pandemic on aggregate consumption (with no fiscal policy response), separating the drop in marginal utility from the reduction in income due to mass layoffs. The figure illustrates that the constrained consumption choices are quantitatively key in capturing the expected depth in the slump of spending, which is already under way; see Armantier et al. (2020) and Baker et al. (2020) for early evidence. The marginal utility shock hits all households and directly affects their spending decisions in the early quarters after the pandemic; its effect cannot be mitigated by fiscal stimulus. The loss of income from unemployment is large

Figure 3. Consumption Response to the Pandemic and the Fiscal Stimulus

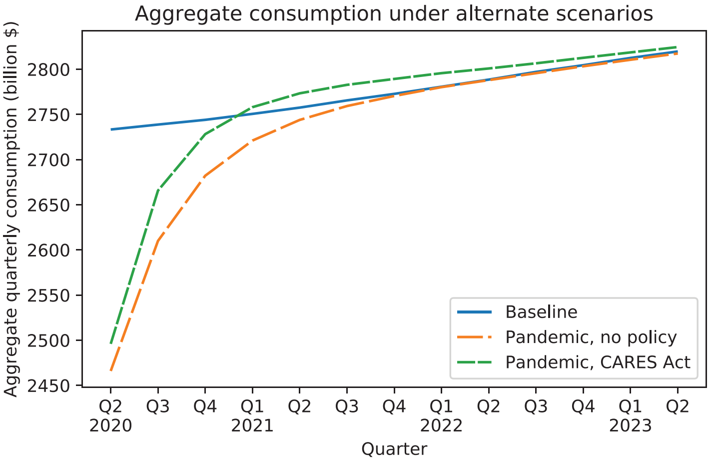
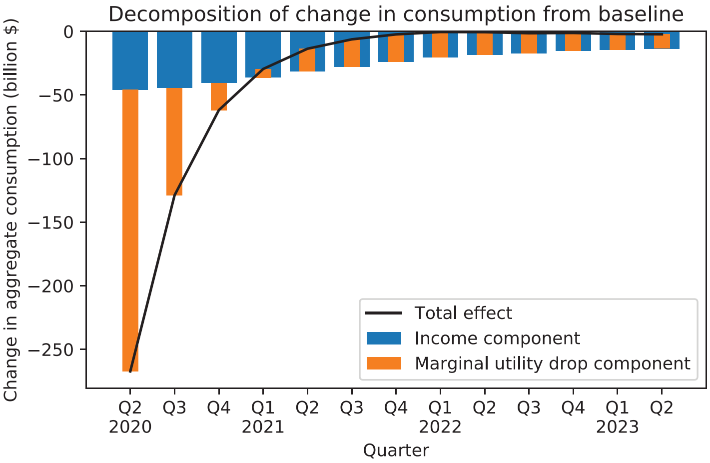


Figure 4. Decomposition of Effect of the Pandemic on Aggregate Consumption (No Policy Response)



but affects only a fraction of households, who are disproportionately low income and thus account for a smaller share of aggregate consumption. Moreover, most households hold at least some liquid assets, allowing them to smooth their consumption drop—the 5 percent decrease in labor income in figure 2 induces only a 1.5 percent decrease in consumption in figure 4.

Figure 5 shows how the consumption response varies depending on the employment status of households in 2020:Q2. For each employment category (employed, unemployed, and deeply unemployed), the figure shows consumption relative to the same households' consumption in the baseline scenario with no pandemic (dotted lines).[15] The upper panel shows consumption without any policy response, while the lower panel includes the CARES Act. The figure illustrates an important feature of the unemployment benefits that is lost at the aggregate level: the response provides the most relief to households whose consumption is most affected by the pandemic. For the unemployed—and especially for the deeply unemployed—the consumption drop when the pandemic hits is much shallower and returns faster toward the baseline when the fiscal stimulus is in place.

Indeed, this targeted response is again seen in figure 6, showing the extra consumption relative to the pandemic scenario without the CARES Act. The short-dashed and dotted lines show the effect of the stimulus check in isolation (for employed workers this is the same as the total fiscal response). For unemployed households, this is dwarfed by the increased unemployment benefits because these benefits both arrive earlier and are much larger. Specifically, in 2020:Q3, when households receive the stimulus checks, the effect of unemployment benefits on consumption makes up about 70 percent and 85 percent of the total effect for the normally and deeply unemployed, respectively.

[15]Households that become unemployed during the pandemic might or might not have been unemployed otherwise. We assume that all households that would have been unemployed otherwise are either unemployed or deeply unemployed in the pandemic scenario. However, there are many more households that are unemployed in the pandemic scenario than in the baseline.

Figure 5. Consumption Response by Employment Status

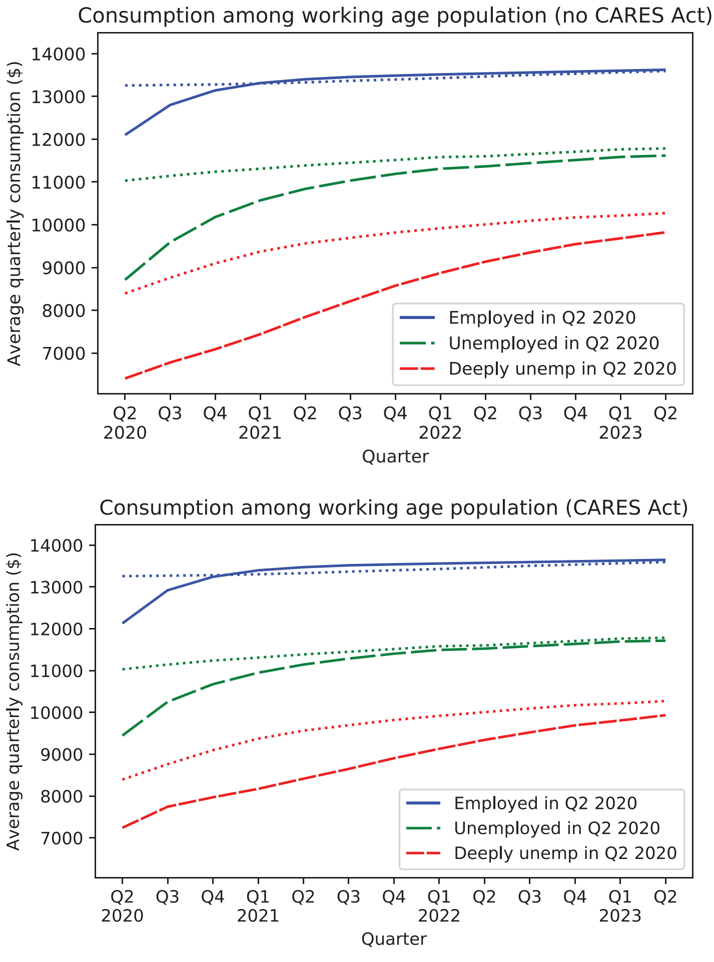


Figure 7 aggregates the decomposition of the CARES Act in figure 6 across all households. In our model economy, the extra unemployment benefits amount to $544 per household, while the stimulus checks amount to $1,054 per household (as means testing reduces or eliminates the stimulus checks for high-income households). Aggregated, stimulus checks amount to $267 billion, while the extended unemployment benefits amount to just over half that,

Figure 6. Effect of CARES Act by Employment Status

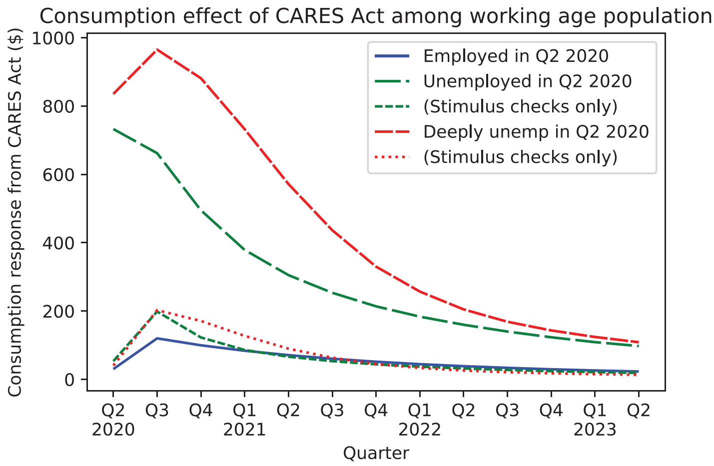
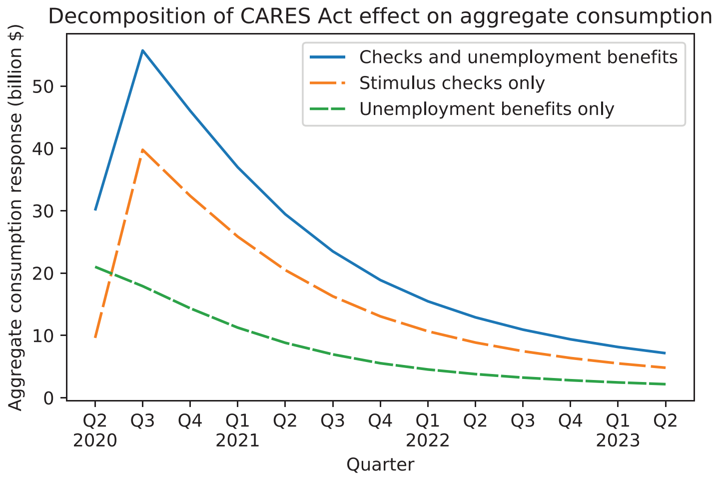


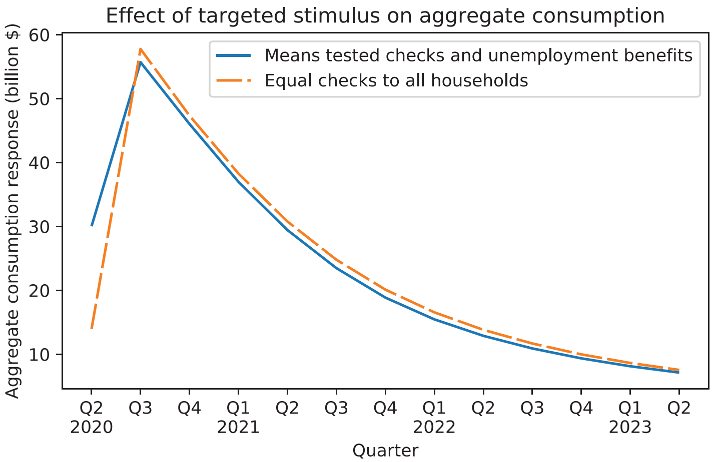
Figure 7. Aggregate Consumption Effect of Stimulus Checks versus Unemployment Benefits



$137 billion.[16] The figure shows that during the peak consumption response in 2020:Q3, the stimulus checks account for about 70 percent of the total effect on consumption for the average household and

[16]See appendix B for details on how we aggregate households.

Figure 8. Effect of Targeting the CARES Act Consumption Stimulus

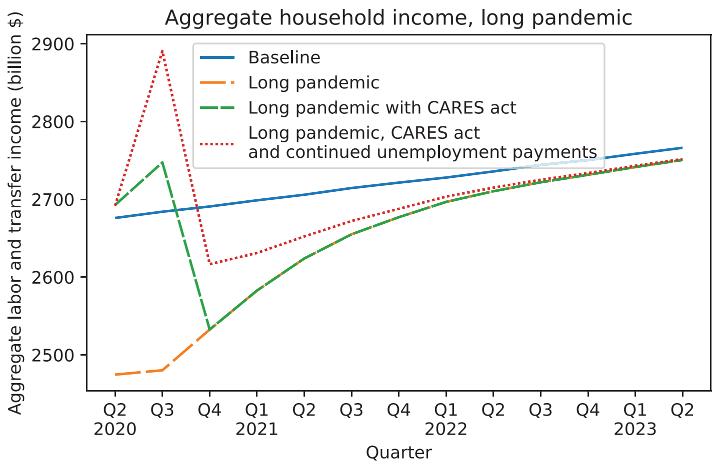


the unemployment benefits for about 30 percent. Thus, although the unemployment benefits make a much larger difference to the spending of the individual recipients than the stimulus checks, a small enough proportion of households becomes unemployed that the total extra spending coming from these people is less than the total extra spending from the more widely distributed stimulus checks.

The previous graphs show the importance of the targeted unemployment benefits at the individual level, but the aggregate effect is less striking. Figure 8 compares the effect of the CARES Act (both unemployment insurance and stimulus checks) to a policy of the same absolute size that distributes checks to everybody. While unemployment benefits arrive sooner, resulting in higher aggregate consumption in 2020:Q2, the untargeted policy leads to higher aggregate consumption in the following quarters.

The interesting conclusion is that, while the net spending response is similar for alternative ways of distributing the funds, the choice to extend unemployment benefits means that much more of the extra spending is coming from the people who will be worst hurt by the crisis. This has obvious implications for the design of any further stimulus packages that might be necessary if the crisis lasts longer than our baseline scenario assumes.

Figure 9. Labor and Transfer Income during the Long, Four-Quarter Pandemic



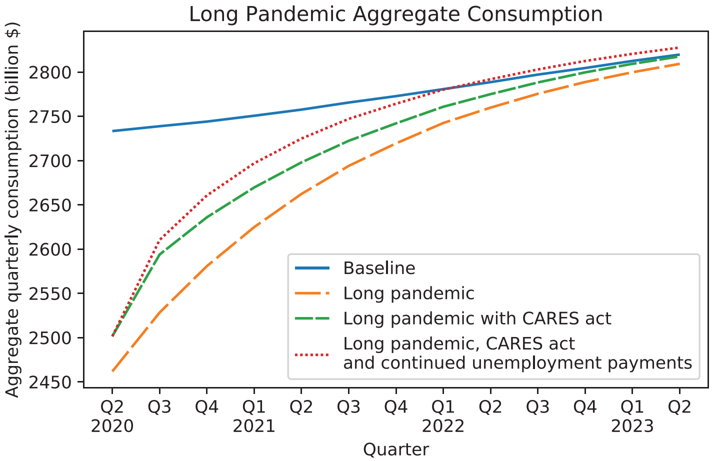
3.2 Alternative Senario: Long, Deep Pandemic

Given the uncertainty about how long and deep the current recession will be, we investigate a more pessimistic scenario in which the lockdown is expected to last for four quarters. In addition, the unemployment rate increases to 20 percent in 2020:Q2, consisting of 15 percent of deeply unemployed and 5 percent of normal unemployed. In this scenario we compare how effectively the CARES package stimulates consumption, also considering a more generous plan in which the unemployment benefits continue until the lockdown is over. We model the receipt of unemployment benefits each quarter as an unexpected shock, representing a series of policy renewals.

Figure 9 compares the effects of the two fiscal stimulus policies on income. The persistently high unemployment results in a substantial and long drop in aggregate income (long-dashed line) as compared to the no-pandemic scenario. The CARES stimulus (medium-dashed line) provides only a short-term support to income for the first two quarters. In contrast, the scenario with unemployment benefits extended as long as the lockdown lasts (dotted line) keeps aggregate income elevated through the recession.

Figure 10 shows the implications of the two stimulus packages for aggregate consumption. The long lockdown causes a much longer

Figure 10. Consumption Response to the Long, Four-Quarter Pandemic



decline in spending than the shorter lockdown in our primary scenario. In the shorter pandemic scenario (figure 3) consumption returns to the baseline path after roughly one year, while in the long lockdown shown here the recovery takes around three years; the CARES stimulus shortens the consumption drop to about two years. The scenario with extended unemployment benefits ensures that aggregate spending returns to near the baseline path after just over one year, and does so by targeting the funds to the people who are worst hurt by the crisis and to whom the cash will make the most difference.

4. Conclusions

Our model suggests that there may be a strong consumption recovery when the social-distancing requirements of the pandemic begin to subside. We invite readers to test the robustness of this conclusion by using the associated software toolkit to choose their own preferred assumptions on the path of the pandemic, and of unemployment, to understand better how consumption will respond.

One important limitation of our analysis is that it does not incorporate Keynesian demand effects or other general equilibrium responses to the consumption fluctuations we predict. In practice, Keynesian effects are likely to cause movements in aggregate income in the same direction as consumption; in that sense, our estimates can be thought of as a "first-round" analysis of the dynamics of the crisis, which will be amplified by any Keynesian response. (See Bayer et al. 2020 for estimates of the multiplier for transfer payments.) These considerations further strengthen the case that the CARES Act will make a substantial difference to the economic outcome. A particularly important consideration is that forward-looking firms that expect consumer demand to return forcefully in the third and fourth quarters of 2020 are more likely to maintain relations with their employees so that they can restart production quickly.

The ability to incorporate Keynesian demand effects is one of the most impressive achievements of the generation of heterogeneous agent macroeconomic models that have been constructed in the last few years. But the technical challenges of constructing those models are such that they cannot yet incorporate realistic treatments of features that our model says are quantitatively important, particularly differing risks of (and types of) unemployment, for different kinds of people (young, old; rich, poor; high and low education). This rich heterogeneity is important both to the overall response to the CARES Act and to making judgments about the extent to which it has been successfully targeted to provide benefits to those who need them most. A fuller analysis that incorporates such heterogeneity, which is of intrinsic interest to policymakers, as well as a satisfying treatment of general equilibrium will have to wait for another day, but that day is likely not far off.

Appendix A. Model Details

The baseline model is adapted and expanded from Carroll et al. (2017). The economy consists of a continuum of expected utility maximizing households with a common CRRA (constant relative risk aversion) utility function over consumption, $\mathrm{u}(\mathbf{c}, \eta) = \eta \mathbf{c}^{1-\rho}/(1-\rho)$, where η is a marginal utility shifter. Households are ex ante heterogeneous: household i has a quarterly time discount factor $\beta_i \leq 1$ and an education level $e_i \in \{D, HS, C\}$ (for dropout, high school, and college, respectively). Each quarter, the household receives (after tax) income, chooses how much of

Table A.1. Parameter Values in the Baseline Model

Description	Parameter	Value
Coefficient of Relative Risk Aversion	ρ	1
Mean Discount Factor, High-School Dropout	$\grave{\beta}_D$	0.9637
Mean Discount Factor, High-School Graduate	$\grave{\beta}_{HS}$	0.9705
Mean Discount Factor, College Graduate	$\grave{\beta}_C$	0.9756
Discount Factor Band (Half Width)	∇	0.0253
Employment Transition Probabilities:		
From Normal Unemployment to Employment	$\Xi_{1,0}$	2/3
From Deep Unemployment to Normal Unemployment	$\Xi_{2,1}$	1/3
From Deep Unemployment to Employment	$\Xi_{2,0}$	0
Proportion of High-School Dropouts	θ_D	0.11
Proportion of High-School graduates	θ_{HS}	0.55
Proportion of College Graduates	θ_C	0.34
Average Initial Permanent Income, Dropout	$\bar{\boldsymbol{p}}_{D0}$	5,000
Average Initial Permanent Income, High School	$\bar{\boldsymbol{p}}_{HS0}$	7,500
Average Initial Permanent Income, College	$\bar{\boldsymbol{p}}_{C0}$	12,000
Steady-State Unemployment Rate	$\mho$	0.05
Unemployment Insurance Replacement Rate	$\underline{\xi}$	0.30
Skill Rot of All Unemployed	χ	0.00125
Quarterly Interest Factor	R	1.01
Population Growth Factor	N	1.0025
Technological Growth Factor	$\daleth$	1.0025

their market resources $\mathbf{m}_{it}$ to consume $\mathbf{c}_{it}$ and how much to retain as assets $\mathbf{a}_{it}$; they then transition to the next quarter by receiving shocks to mortality, income, their employment state, and their marginal utility of consumption.

For each education group e, we assign a uniform distribution of time preference factors between $\grave{\beta}_e - \nabla$ and $\grave{\beta}_e + \nabla$, chosen to match the distribution of liquid wealth and retirement assets. Specifically, the calibrated values in table A.1 fit the ratio of liquid wealth to permanent income in aggregate for each education level, as computed from the 2004 Survey of Consumer Finance. The width of the distribution of discount factors was calibrated to minimize the difference between simulated and empirical Lorenz shares of liquid wealth for the bottom 20 percent, 40 percent, 60 percent, and 80 percent of households, as in Carroll et al. (2017).

When transitioning from one period to the next, a household with education e that has already lived for j periods faces a D_{ej} probability of death. The quarterly mortality probabilities are calculated from the Social Security Administration's actuarial table (for annual mortality probability) and adjusted for education using Brown, Liebman, and Pollett (2002); a household dies with certainty if it (improbably) reaches the age of 120 years. The assets of a household that dies are completely taxed by the government to fund activities outside the model. Households who survive to period $t+1$ experience a return factor of R on their assets, assumed constant.

Household i's state in period t, at the time it makes its consumption–saving decision, is characterized by its age j,[17] a level of market resources $\mathbf{m}_{it} \in \mathbb{R}_+$, a permanent income level $\boldsymbol{p}_{it} \in \mathbb{R}_{++}$, a discrete employment state $\ell_{it} \in \{0, 1, 2\}$ (indicating whether the individual is employed, normal unemployed, or deeply unemployed), and a discrete state $\eta_{it} \in \{1, \underline{\eta}\}$ that represents whether its marginal utility of consumption has been temporarily reduced ($\underline{\eta} < 1$). Denote the joint discrete state as $n_{it} = (\ell_{it}, \eta_{it})$.

Each household inelastically participates in the labor market when it is younger than 65 years ($j < 164$) and retires with certainty at age 65. The transition from working life to retirement is captured in the model by a one-time large decrease in permanent income at age $j = 164$.[18] Retired households face essentially no income risk: they receive Social Security benefits equal to their permanent income with 99.99 percent probability and miss their check otherwise; their permanent income very slowly degrades as they age. The discrete employment state ℓ_{it} is irrelevant for retired households.

Labor income for working-age households is subject to three risks: unemployment, permanent income shocks, and transitory income shocks. Employed ($\ell_{it} = 0$) households' permanent income grows by age-education-conditional factor Γ_{ej} on average, subject to a mean one log-normal permanent income shock ψ_{it} with age-conditional underlying standard deviation of $\sigma_{\psi j}$. The household's

[17] Households enter the model aged 24 years, so model age $j = 0$ corresponds to being 24 years, 0 quarters old.

[18] The size of the decrease depends on education level, very roughly approximating the progressive structure of Social Security: $\Gamma_{D164} \approx 0.56$, $\Gamma_{HS164} \approx 0.44$, $\Gamma_{C164} \approx 0.31$.

labor income $\mathbf{y}_{it}$ is also subject to a mean one log-normal transitory shock ξ_{it} with age-conditional underlying standard deviation of $\sigma_{\xi j}$. The age profiles of permanent and transitory income shock standard deviations are approximated from the results of Sabelhaus and Song (2010), and the expected permanent income growth factors are adapted from Cagetti (2003). Normal unemployed and deeply unemployed households receive unemployment benefits equal to a fraction $\underline{\xi} = 0.3$ of their permanent income, $\mathbf{y}_{it} = \underline{\xi}\boldsymbol{p}_{it}$; they are not subject to permanent nor transitory income risk, but their permanent income grows at rate χ less than if employed, representing "skill rot."[19]

The income process for a household can be represented mathematically as

$$\boldsymbol{p}_{it} = \begin{cases} \psi_{it}\Gamma_{ej}\boldsymbol{p}_{it-1} & \text{if } \ell_{it} = 0,\ j < 164 \ \text{Employed, working age} \\ (\Gamma_{ej} - \chi)\,\boldsymbol{p}_{it-1} & \text{if } \ell_{it} > 0,\ j < 164 \ \text{Unempl., working age} \\ \Gamma_{ret}\boldsymbol{p}_{it-1} & \text{if } j \geq 164 \qquad\qquad \text{Retired,} \end{cases}$$

$$\mathbf{y}_{it} = \begin{cases} \xi_{it}\boldsymbol{p}_{it} & \text{if } \ell_{it} = 0,\ j < 164 \qquad \text{Employed, working age} \\ \underline{\xi}\boldsymbol{p}_{it} & \text{if } \ell_{it} > 0,\ j < 164 \qquad \text{Unempl., working age} \\ \boldsymbol{p}_{it} & \text{if } j \geq 164 \qquad\qquad \text{Retired.} \end{cases}$$

A working-age household's employment state ℓ_{it} evolves as a Markov process described by the matrix Ξ, where element k, k' of Ξ is the probability of transitioning from $\ell_{it} = k$ to $\ell_{it+1} = k'$. During retirement, all households have $\ell_{it} = 0$ (or any other trivializing assumption about the "employment" state of the retired). We assume that households treat $\Xi_{0,2}$ and $\Xi_{1,2}$ as zero: they do not consider the possibility of ever attaining the deep unemployment state $\ell_{it} = 2$ from "normal" employment or unemployment, and thus it does not affect their consumption decision in those employment states.

We specify the unemployment rate during normal times as $\mho = 5\%$, and the expected duration of an unemployment spell as

[19]Unemployment is somewhat persistent in our model, so the utility risk from receiving 15 percent of permanent income for one quarter (as in Carroll et al. 2017) is roughly the same as the risk of receiving 30 percent of permanent income for 1.5 quarters in expectation.

1.5 quarters. The probability of transitioning from unemployment back to employment is thus $\Xi_{1,0} = \frac{2}{3}$, and the probability of becoming unemployed is determined as the flow rate that offsets this to generate 5 percent unemployment (about 3.5 percent). The deeply unemployed expect to be unemployed for *much* longer: we specify $\Xi_{2,0} = 0$ and $\Xi_{2,1} = \frac{1}{3}$, so that a deeply unemployed person remains so for three quarters on average before becoming "normal" unemployed (they cannot transition directly back to employment). Thus the unemployment spell for a deeply unemployed worker is 2 quarters at a minimum and 4.5 quarters on average.[20]

Like the prospect of deep unemployment, the possibility that consumption might become less appealing (via marginal utility scaling factor $\eta_{it} < 1$) does not affect the decisionmaking process of a household in the normal $\eta_{it} = 1$ state. If a household does find itself with $\eta_{it} = \underline{\eta}$, this condition is removed (returning to the normal state) with probability 0.5 each quarter; the evolution of the marginal utility scaling factor is represented by the Markov matrix H. In this way, the consequences of a pandemic are fully unanticipated by households, a so-called MIT shock; households act optimally once in these states but did not account for them in their consumption–saving problem during "normal" times.[21]

The household's permanent income level can be normalized out of the problem, dividing all boldface variables (absolute levels) by the individual's permanent income $\boldsymbol{p}_{it}$, yielding nonbold normalized variables, e.g., $m_{it} = \mathbf{m}_{it}/\boldsymbol{p}_{it}$. Thus the only state variables that affect the choice of optimal consumption are normalized market resources m_{it} and the discrete Markov states n_{it}. After this normalization, the household consumption functions $\mathrm{c}_{e,j}$ satisfy

[20]Our computational model allows for workers' *beliefs* about the average duration of deep unemployment to differ from the *true* probability. However, we do not present results based on this feature and thus will not further clutter the notation by formalizing it here.

[21]Our computational model also allows households' beliefs about the duration of the reduced marginal utility state (via social distancing) to deviate from the true probability. The code also permits the possibility that the reduction in marginal utility is lifted as an aggregate or shared outcome, rather than idiosyncratically. We do not present results utilizing these features here, but invite the reader to investigate their predicted consequences using our public repository.

$$\begin{aligned}
\mathrm{v}_{e,j}(m_{it}, n_{it}) = \max_{\mathrm{c}_{e,j}} \; & (\mathrm{c}_{e,j}(m_{it}, n_{it}), \eta_{it}) \\
& + \beta_i (1 - \mathsf{D}_{e,j}) \mathbb{E}_t \left[\widehat{\Gamma}_{it+1}^{1-\rho} \mathrm{v}_{e,j+1}(m_{it+1}, n_{it+1}) \right] \\
& \text{s.t.} \\
a_{it} &= m_{it} - \mathrm{c}_{e,j}(m_{it}, n_{it}), \\
m_{it+1} &= (\mathsf{R}/\widehat{\Gamma}_{it+1}) a_{it} + y_{it}, \\
n_{it+1} &\sim (\Xi, H), \\
a_{it} &\geq 0,
\end{aligned}$$

where $\widehat{\Gamma}_{it+1} = \boldsymbol{p}_{it+1}/\boldsymbol{p}_{it}$, the realized growth rate of permanent income from period t to $t+1$. Consumption function $\mathrm{c}_{e,j}$ yields optimal *normalized* consumption, the ratio of consumption to the household's permanent income level; the actual consumption level is simply $\mathbf{c}_{it} = \boldsymbol{p}_{it} \mathrm{c}_{e,j}(m_{it}, n_{it})$.

Starting from the terminal model age of $j = 384$, representing being 120 years old (when the optimal choice is to consume all market resources, as death is certain), we solve the model by backward induction using the endogenous grid method, originally presented in Carroll (2006). Substituting the definition of next period's market resources into the maximand, the household's problem can be rewritten as

$$\begin{aligned}
\mathrm{v}_{e,j}(m_{it}, n_{it}) = \max_{c_{it} \in \mathbb{R}_+} \; & \mathrm{u}(c_{it}, \eta_{it}) \\
& + \beta_i (1 - \mathsf{D}_{e,j}) \mathbb{E}_t \left[\widehat{\Gamma}_{it+1}^{1-\rho} \mathrm{v}_{e,j+1}((\mathsf{R}/\widehat{\Gamma}_{it+1}) a_{it} + y_{it}, n_{it+1}) \right] \\
\text{s.t.} \;\; a_{it} = m_{it} - c_{it}, \qquad & a_{it} \geq 0, \qquad n_{it+1} \sim (\Xi, H).
\end{aligned}$$

This problem has one first-order condition, which is both necessary and sufficient for optimality. It can be solved to yield optimal consumption as a function of (normalized) end-of-period assets and the Markov state:

$$\begin{aligned}
& \underbrace{\eta_{it} c_{it}^{-\rho}}_{= \frac{\partial \mathrm{u}}{\partial c}} - \underbrace{\beta_i \mathsf{R} (1 - \mathsf{D}_{e,j}) \mathbb{E}_t \left[\widehat{\Gamma}_{it+1}^{-\rho} \mathrm{v}_{e,j+1}^{m}((\mathsf{R}/\widehat{\Gamma}_{it+1}) a_{it} + y_{it}, n_{it+1}) \right]}_{\equiv \mathfrak{v}_{e,j}^{a}(a_{it}, n_{it})} \\
& = 0 \Longrightarrow c_{it} = \left(\frac{\mathfrak{v}_{e,j}^{a}(a_{it}, n_{it})}{\eta_{it}} \right)^{-\frac{1}{\rho}}.
\end{aligned}$$

To solve the age-j problem numerically, we specify an exogenous grid of end-of-period asset values $a \geq 0$, compute end-of-period marginal value of assets at each gridpoint (and each discrete Markov state), then calculate the unique (normalized) consumption that is consistent with ending the period with this quantity of assets while acting optimally. The beginning-of-period (normalized) market resources from which this consumption was taken is then simply $m_{it} = a_{it} + c_{it}$, the *endogenous gridpoint*. We then linearly interpolate on this set of market resources–consumption pairs, adding an additional bottom gridpoint at $(m_{it}, c_{it}) = (0, 0)$ to represent the liquidity-constrained portion of the consumption function $\mathrm{c}_{e,j}(m_{it}, n_{it})$.

The standard envelope condition applies in this model, so that the marginal value of market resources equals the marginal utility of consumption when consuming optimally:

$$\mathrm{v}^m_{e,j}(m_{it}, n_{it}) = \eta_{it}\mathrm{c}_{e,j}(m_{it}, n_{it})^{-\rho}.$$

The marginal value function for age j can then be used to solve the age $j-1$ problem, iterating backward until the initial age $j = 0$ problem has been solved.

When the pandemic strikes, we draw a new employment state (employed, unemployed, deeply unemployed) for each working-age household using a logistic distribution. For each household i at $t = 0$ (the beginning of the pandemic and lockdown), we compute logistic weights for the employment states as

$$\mathbb{P}_{i,\ell} = \alpha_{\ell,e} + \alpha_{\ell,p}\boldsymbol{p}_{i0} + \alpha_{\ell,j}j_{i0} \quad \text{for} \quad \ell \in \{1, 2\}, \qquad \mathbb{P}_{i,0} = 0,$$

where $e \in \{D, H, C\}$ for dropouts, high-school graduates, and college graduates and j is the household's age. The probability that household i draws employment state $\ell \in \{0, 1, 2\}$ is then calculated as

$$\Pr(\ell_{it} = \ell) = \exp(\mathbb{P}_{i,\ell}) \Bigg/ \sum_{k=0}^{2} \exp(\mathbb{P}_{i,k}).$$

Our chosen logistic parameters are presented in table A.2.

Table A.2. Pandemic Assumptions

Description	Parameter	Value
Short-Lived Pandemic		
Logistic Parameterization of Unemployment Probabilities		
Constant for Dropout, Regular Unemployment	$\alpha_{1,D}$	−1.15
Constant for Dropout, Deep Unemployment	$\alpha_{2,D}$	−1.5
Constant for High School, Regular Unemployment	$\alpha_{1,H}$	−1.3
Constant for High School, Deep Unemployment	$\alpha_{2,H}$	−1.75
Constant for College, Regular Unemployment	$\alpha_{1,C}$	−1.65
Constant for College, Deep Unemployment	$\alpha_{2,C}$	−2.2
Coefficient on Permanent Income, Regular Unemployment	$\alpha_{1,p}$	−0.1
Coefficient on Permanent Income, Deep Unemployment	$\alpha_{2,p}$	−0.2
Coefficient on Age, Regular Unemployment	$\alpha_{1,j}$	−0.01
Coefficient on Age, Deep Unemployment	$\alpha_{2,j}$	−0.01
Marginal Utility Shock		
Pandemic Utility Factor	$\underline{\eta}$	0.891
Prob. Exiting Pandemic Each Quarter	$H_{1,0}$	0.5
Long, Deep Pandemic		
Logistic Parameterization of Unemployment Probabilities		
Constant for Dropout, Regular Unemployment	$\alpha_{1,D}$	−1.45
Constant for Dropout, Deep Unemployment	$\alpha_{2,D}$	−0.3
Constant for High School, Regular Unemployment	$\alpha_{1,H}$	−1.6
Constant for High School, Deep Unemployment	$\alpha_{2,H}$	−0.55
Constant for College, Regular Unemployment	$\alpha_{1,C}$	−1.95
Constant for College, Deep Unemployment	$\alpha_{2,C}$	−1.00
Coefficient on Permanent Income, Regular Unemployment	$\alpha_{1,p}$	−0.2
Coefficient on Permanent Income, Deep Unemployment	$\alpha_{2,p}$	−0.2
Coefficient on Age, Regular Unemployment	$\alpha_{1,j}$	−0.01
Coefficient on Age, Deep Unemployment	$\alpha_{2,j}$	−0.01
Marginal Utility Shock		
Pandemic Utility Factor	$\underline{\eta}$	0.891
Prob. Exiting Pandemic Each Quarter	$H_{1,0}$	0.25

Table A.3. Fiscal Stimulus Assumptions, CARES Act

Description	Value
Stimulus Check	$1,200
Means Test Start (Annual)	$75,000
Means Test End (Annual)	$99,000
Stimulus Check Delay	One Quarter
Fraction that React on Announcement	0.25
Extra Unemployment Benefit for:	
Normal Unemployed	$5,200
Deeply Unemployed	$7,800

Notes: The unemployment benefits are multiplied by 0.8 to account for the fact that 20 percent of the working-age population is out of the labor force. See aggregation details in appendix B.

Appendix B. Aggregation

Households are modeled as individuals and incomes sized accordingly. We completely abstract from family dynamics. To get our aggregate predictions for income and consumption, we take the mean from our simulation and multiply by 253 million, the number of adults (over 18) in the United States in 2019. To size the unemployment benefits correctly, we multiply the benefits per worker by 0.8 to account for the fact that 20 percent of the working-age population is out of the labor force, so the average working-age household consists of 0.8 workers and 0.2 nonworkers. With this adjustment, there are 151 million workers eligible for unemployment benefits in the model. Aggregate consumption in our baseline for 2020 is just over $11 trillion, a little less than total personal consumption expenditure, accounting for the fact that some consumption does not fit in the usual budget constraint.[22] Aggregating in this way underweights the young, as our model excludes those under the age of 24.

[22] Personal consumption expenditures (PCE) consumption in 2019:Q4, from the NIPA (national income and product accounts) tables, was $14.8 trillion. Market-based PCE, a measure that excludes expenditures without an observable price, was $12.9 trillion. Health care, much of which is paid by employers and not in the household's budget constraint, was $2.5 trillion.

Our model estimates the aggregate size of the stimulus checks to be \$267 billion, matching the Joint Committee on Taxation's (JCT's) estimate of disbursements in 2020.[23] This is somewhat of a coincidence: we overestimate the number of adults who will actually receive the stimulus, while excluding the \$500 payment to children.

The aggregate cost of the extra unemployment benefits depends on the expected level of unemployment. Our estimate is \$137 billion, much less than the \$260 billion mentioned in several press reports, but in line with the extent of unemployment in our pandemic scenario.[24] We do not account for the extension of unemployment benefits to the self-employed and gig workers.

Households enter the model at age $j = 0$ with zero liquid assets. A "newborn" household has its initial permanent income drawn log-normally with underlying standard deviation of 0.4 and an education-conditional mean. The initial employment state of households matches the steady-state unemployment rate of 5 percent.[25]

We assume annual population growth of 1 percent, so older simulated households are appropriately downweighted when we aggregate idiosyncratic values. Likewise, each successive cohort is slightly more productive than the last, with aggregate productivity growing at a rate of 1 percent per year. The profile of average income by age in the population at any moment in time thus has more of an inverted-U shape than implied by the permanent income profiles from Cagetti (2003).

Appendix C. Marginal Utility Equivalence

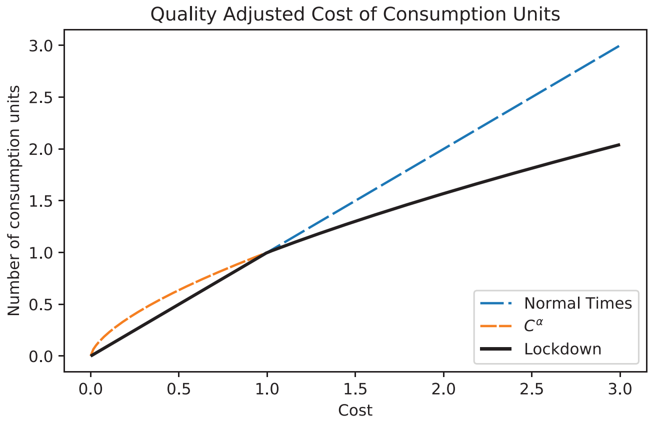
We model the "lockdown" as a reduction in the marginal utility of consumption. This can be interpreted as an increase in the quality-adjusted price of goods, where the quality of basic goods such as

[23]The JCT's March 26, 2020 publication JCX-11-20 predicts disbursements of \$267 billion in 2020, followed by \$24 billion in 2021.

[24]While \$260 billion was widely reported in the press, back-of-the-envelope calculations show this to be an extreme number. Furthermore, the origin of this reported number is unclear.

[25]This is the case even during the pandemic and lockdown, so the death and replacement of simulated agents is a second-order contribution to the profile of the unemployment rate.

Figure C.1. Concave Cost of Consumption Units



shelter and housing has not decreased, but more discretionary goods such as vacations and restaurants have decreased in quality.

Figure C.1 shows how this works. In normal times, the cost of a consumption unit is equal to one, represented by the long-dashed line. During the lockdown, the cost of a unit of consumption is increasing in the number of units bought. As shown here, the number of consumption units that can be bought follows the lower envelope of the long-dashed and short-dashed lines, where the short-dashed line is equal to Cost^{α}. As long as the household is consuming above the kink, their utility is $\log(\text{Cost}^{\alpha}) = \alpha \log(\text{Cost})$, exactly equivalent to the reduction in marginal utility we apply. Taking this interpretation seriously, the drop in marginal utility should not be applied to households with very low levels of consumption, below the kink. Our implementation abstracts from this, taking the marginal utility factor to be the same for all agents.

An alternative interpretation is that consumption is made up of a Cobb-Douglass aggregation of two goods:

$$C = c_1^{\alpha} c_2^{1-\alpha}.$$

During the lockdown, the second good is replaced by home production at a fixed level $\bar{c_2}$. A log-utility function gives $\log(C) = \alpha \log(c_1) + (1-\alpha)\log(\bar{c_2})$, equivalent to our model in which we reduce marginal utility by a constant factor.

References

Adams-Prassl, A., T. Boneva, M. Golin, and C. Rauh. 2020. "Inequality in the Impact of the Coronavirus Shock: New Survey Evidence for the US." Mimeo, Oxford University.

Alvarez, F., D. Argente, and F. Lippi. 2020. "A Simple Planning Problem for COVID-19 Lockdown." BFI Working Paper No. 34, University of Chicago.

Andersen, A. L., E. Toft Hansen, N. Johannesen, and A. Sheridan. 2020. "Consumer Responses to the COVID-19 Crisis: Evidence from Bank Account Transaction Data." *Covid Economics* 7 (April 20): 88–111.

Armantier, O., G. Kosar, R. Pomerantz, D. Skandalis, K. Smith, G. Topa, and W. van der Klaauw. 2020. "Coronavirus Outbreak Sends Consumer Expectations Plummeting." Blog Post, Liberty Street Economics, Federal Reserve Bank of New York. https://libertystreeteconomics.newyorkfed.org/2020/04/coronavirus-outbreak-sends-consumer-expectations-plummeting.html.

Baker, S. R., R. A. Farrokhnia, S. Meyer, M. Pagel, and C. Yannelis. 2020. "How Does Household Spending Respond to an Epidemic? Consumption During the 2020 COVID-19 Pandemic." Working Paper No. 26949, National Bureau of Economic Research.

Bayer, C., B. Born, R. Luetticke, and G. J. Müller. 2020. "The Coronavirus Stimulus Package: How Large Is the Transfer Multiplier?" CEPR Discussion Paper No. 14600.

Bick, A., and A. Blandin. 2020. "Real Time Labor Market Estimates During the 2020 Coronavirus Outbreak." Working Paper, Arizona State University.

Bloomberg. 2020a. "Mnuchin Warns Virus Could Yield 20% Jobless Rate Without Action." Press Article.

———. 2020b. "U.S. Jobless Rate May Soar to 30%, Fed's Bullard Says." Press Article.

Broda, C., and J. A. Parker. 2014. "The Economic Stimulus Payments of 2008 and the Aggregate Demand for Consumption." *Journal of Monetary Economics* 68 (Supplement): S20–S36.

Brown, J., J. B. Liebman, and J. Pollett. 2002. "Estimating Life Tables That Reflect Socioeconomic Differences in Mortality." In *The Distributional Aspects of Social Security and Social Security*

Reform, ed. M. Feldstein and J. B. Liebman, 447–57. University of Chicago Press.

Cagetti, M. 2003. “Wealth Accumulation Over the Life Cycle and Precautionary Savings.” *Journal of Business and Economic Statistics* 21 (3): 339–53.

Carroll, C. D. 2006. “The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems.” *Economics Letters* 91 (3): 312–20.

Carroll, C. D., E. Crawley, J. Slacalek, K. Tokuoka, and M. N. White. 2020. “Sticky Expectations and Consumption Dynamics.” *American Economic Journal: Macroeconomics* 12 (3): 40–76.

Carroll, C. D., A. M. Kaufman, J. L. Kazil, N. M. Palmer, and M. N. White. 2018. “The Econ-ARK and HARK: Open Source Tools for Computational Economics.” In *Proceedings of the 17th Python in Science Conference*, ed. F. Akici, D. Lippa, D. Niederhut, and M. Pacer, 25–30. https://doi.org/10.5281/zenodo.1001067.

Carroll, C. D., J. Slacalek, K. Tokuoka, and M. N. White. 2017. “The Distribution of Wealth and the Marginal Propensity to Consume.” *Quantitative Economics* 8 (3): 977–1020. https://econ2.jhu.edu/people/ccarroll/papers/cstwMPC.pdf.

Carvalho, V. M, J. R. Garcia, S. Hansen, A. Ortiz, T. Rodrigo, J. V. Mora Rodriguez, and J. Ruiz. 2020. “Tracking the COVID-19 Crisis with High-Resolution Transaction Data.” Discussion Paper, Cambridge University.

Casado, M. G., B. Glennon, J. Lane, D. McQuown, D. Rich, and B. A. Weinberg. 2020. “The Effect of Fiscal Stimulus: Evidence from COVID-19.” Working Paper No. 27576, National Bureau of Economic Research.

Chetty, R., J. Friedman, N. Hendren, M. Stepner, and The Opportunity Insights Team. 2020. “How Did COVID-19 and Stabilization Policies Affect Spending and Employment? A New Real-Time Economic Tracker Based on Private Sector Data.” Working Paper, Harvard University.

Coibion, O., Y. Gorodnichenko, and M. Weber. 2020. “How Did U.S. Consumers Use Their Stimulus Payments?” Working Paper No. 27693, National Bureau of Economic Research.

Correia, S., S. Luck, and E. Verner. 2020. "Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu." Mimeo, MIT Sloan.

Cox, N., P. Ganong, P. Noel, J. Vavra, A. Wong, D. Farrell, and F. Greig. 2020. "Initial Impacts of the Pandemic on Consumer Behavior: Evidence from Linked Income, Spending, and Savings Data." *Brookings Papers on Economic Activity*, forthcoming.

Cyranoski, D. 2020. "'We Need to be Alert': Scientists Fear Second Coronavirus Wave as China's Lockdowns Ease." *Nature*.

Deaton, A. S. 1992. *Understanding Consumption*. New York: Oxford University Press.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt. 2020. "The Macroeconomics of Epidemics." Working Paper No. 26882, National Bureau of Economic Research.

Fagereng, A., M. B. Holm, and G. J. Natvik. 2017. "MPC Heterogeneity and Household Balance Sheets." Discussion Paper, Statistics Norway.

Faria-e-Castro, M. 2020a. "Back-of-the-Envelope Estimates of Next Quarter's Unemployment Rate." Blog Post, Federal Reserve Bank of St. Louis, March 24.

———. 2020b. "Fiscal Policy during a Pandemic." *Covid Economics* 2 (April 8): 67–101.

Garner, T. I., A. Safir, and J. Schild. 2020. "Receipt and Use of Stimulus Payments in the Time of the Covid-19 Pandemic." *Beyond the Numbers: Prices and Spending* (U.S. Bureau of Labor Statistics) 9 (10).

Gascon, C. 2020. "COVID-19: Which Workers Face the Highest Unemployment Risk?" Blog Post, Federal Reserve Bank of St. Louis.

Glover, A., J. Heathcote, D. Krueger, and J.-V. Ríos-Rull. 2020. "Health versus Wealth: On the Distributional Effects of Controlling a Pandemic." *Covid Economics* 6 (April 17): 22–64.

Guerrieri, V., G. Lorenzoni, L. Straub, and I. Werning. 2020. "Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages?" Working Paper No. 26918, National Bureau of Economic Research.

Hai, R., D. Krueger, and A. Postlewaite. 2013. "On the Welfare Cost of Consumption Fluctuations in the Presence of Memorable

Goods." Working Paper No. 19386, National Bureau of Economic Research.

Havranek, T., M. Rusnak, and A. Sokolova. 2017. "Habit Formation in Consumption: A Meta-analysis." *European Economic Review* 95 (June): 142–67.

Heathcote, J., F. Perri, and G. L. Violante. 2020. "The Rise of US Earnings Inequality: Does the Cycle Drive the Trend?" *Review of Economic Dynamics* 37 (Supplement 1): S181–S204.

Jordà, Ò., S. R. Singh, and A. M. Taylor. 2020. "Longer-run Economic Consequences of Pandemics." *Covid Economics* 1 (April 3): 1–15.

JPMorgan. 2020. "Fallout from COVID-19: Global Recession, Zero Interest Rates and Emergency Policy Actions." Blog Post, JPMorgan.

Kaplan, G., B. Moll, and G. L. Violante. 2020. "The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the U.S." Mimeo, Princeton University.

Krueger, D., K. Mitman, and F. Perri. 2016. "Macroeconomics and Household Heterogeneity." *Handbook of Macroeconomics,* Vol. 2A, ed. J. B. Taylor and H. Uhlig, 843–921 (chapter 11). Elsevier.

Krugman, P. 2020. "Notes on Coronacoma Economics." Mimeo, City University of New York.

Leibovici, F., and A. M. Santacreu. 2020. "Social Distancing and Contact-Intensive Occupations." Blog Post, Federal Reserve Bank of St. Louis, March 24.

Oreopoulos, P., T. von Wachter, and A. Heisz. 2012. "The Short- and Long-Term Career Effects of Graduating in a Recession." *American Economic Journal: Applied Economics* 4 (1): 1–29.

Parker, J. A. 2017. "Why Don't Households Smooth Consumption? Evidence from a $25 Million Experiment." *American Economic Journal: Macroeconomics* 9 (4): 153–83.

Parker, J. A, N. S. Souleles, D. S. Johnson, and R. McClelland. 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review* 103 (6): 2530–53.

Sabelhaus, J., and J. Song. 2010. "The Great Moderation in Micro Labor Earnings." *Journal of Monetary Economics* 57 (4): 391–403.

Swagel, P. 2020. "Updating CBO's Economic Forecast to Account for the Pandemic." Blog Post, Congressional Budget Office, April 2.

Yagan, D. 2019. "Employment Hysteresis from the Great Recession." *Journal of Political Economy* 127 (5): 2505–58.

The First Line of Defense: The Discount Window during the Early Stages of the Financial Crisis*

Elizabeth Klee
Board of Governors of the Federal Reserve System

Differences in stigma of borrowing from the discount window across banks caused federal funds rates to rise early in the 2007–09 financial crisis, even as the spread between the discount rate and the target rate narrowed. Low-stigma banks went to the discount window, leaving only high-stigma banks in the market, creating a separating equilibrium. A simple theoretical model illustrates this point, and its implications are evaluated using an empirical selection model. The results suggest the selection effect became stronger as the crisis intensified pre-Lehman, but faded once reserves ballooned.

JEL Codes: E52, E58, G28.

1. Introduction

The discount window's "lender-of-last-resort" function is one of the Federal Reserve's oldest tools to combat financial crises. It was also one of the first tools the Federal Reserve used at the start of the financial crisis in August 2007. About two weeks into the financial crisis, the Federal Reserve Board narrowed the spread between the rate on discount window loans (the "discount rate" or the "primary credit rate") and the Federal Open Market Committee's (FOMC's)

*The views expressed in this paper are those of the author and not necessarily those of the Board of Governors, other members of its staff, or the Federal Reserve System. Thanks to the editors and an anonymous referee for guidance, as well as Seth Carpenter, Huberto Ennis, Jane Ihrig, Ruth Judson, Ed Nelson, Viktors Stebunovs, and Shangwen Wang for helpful comments. Peter Andersen, Amanda McLean, and Ari Morse provided excellent research assistance. Author contact: Board of Governors of the Federal Reserve System, 20th and C Streets, Washington, DC 20551. Tel: (202) 721-4501. E-mail: elizabeth.c.klee@frb.gov.

policy rate (the target federal funds rate, or the "target rate") to promote the restoration of orderly conditions in financial markets.[1,2] On March 16, 2008, the spread was narrowed again, to "bolster market liquidity."[3] By narrowing the spread between the target rate and the primary credit rate,[4] the Federal Reserve aimed to provide ample liquidity to the federal funds market, the overnight U.S. interbank market for funds held in accounts by depository institutions at the Federal Reserve, and to keep rates in the federal funds market trading near the target rate.[5]

These actions were successful, as shown in figure 1: Lending in the Federal Reserve's main discount window program, the primary credit program, stepped up with each narrowing of the spread. (To see the figures in color, where the lines can more easily be differentiated, see the online version of the paper at http://www.ijcb.org.) However, volatility in the federal funds market picked up, and the spread between the highest federal funds rates banks paid and the target rate widened with each narrowing of the spread between the discount rate and the target rate.[6] In fact, on many days, the highest brokered rate was often above the discount rate. This was a puzzle, given that a bank could borrow directly from the Fed at the discount rate, and so the discount rate should have been a ceiling for rates in the federal funds market.[7]

Why did some federal funds trades occur at higher rates, even as the discount rate fell? One possible explanation is that banks borrowing federal funds differ according to their internal costs of using

[1]Federal Reserve Board Press Release, August 17, 2007.

[2]The primary credit program is the name of the principal Federal Reserve discount window program, and the primary credit rate is the rate at which funds are lent to sound depository institutions in that program. Since 2003, primary credit has been offered at a rate above the target federal funds rate.

[3]Federal Reserve Board Press Release, March 16, 2008.

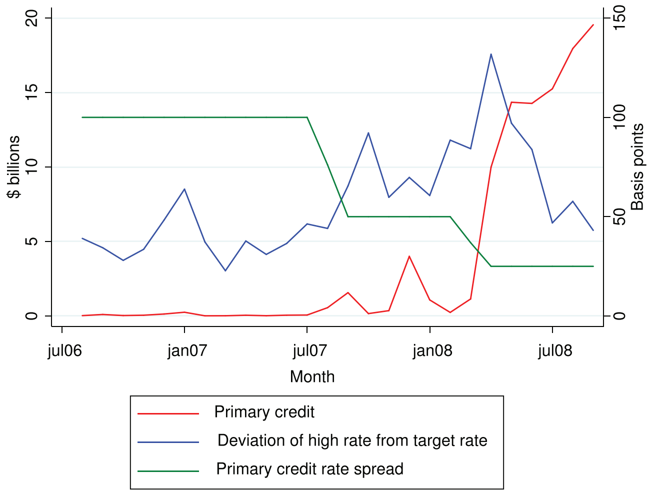
[4]The spread was narrowed by lowering the discount rate, not by raising the target rate.

[5]At that time, the Federal Reserve implemented monetary policy largely by influencing conditions in the federal funds market so that the average rate in that market (the "effective" federal funds rate) trades close to the target rate. For the rate definitions used in this paper, refer to table 1.

[6]In this paper, we use the term "banks" even in instances when the broader term "depository institutions" may apply.

[7]Although there were some instances of trading above the primary credit rate before the beginning of the financial crisis, the incidence was much less frequent.

Figure 1. Primary Credit and the Federal Funds Rate, Monthly Averages



Sources: H.4.1 Statistical Release, Federal Reserve Board; Federal Reserve Bank of New York.
Notes: This figure presents monthly average information on primary credit extensions ($billions), the deviation of the highest brokered federal funds rate from the target rate (basis points), and the spread between the primary credit rate and the target federal funds rate (basis points). The panel reflects data from August 2006 to September 2008.

the discount window as a funding source, over and above the rate charged by the Federal Reserve. These costs can be interpreted as a stigma of discount window borrowing.[8] Banks lending federal funds recognize that some borrowers might have an additional stigma cost of going to the discount window. Consequently, lenders charge borrowers higher rates than would be predicted simply by using the

[8] As described by Bernanke (2008):

> The efficacy of the discount window has been limited by the reluctance of depository institutions to use the window as a source of funding. The "stigma" associated with the discount window, which if anything intensifies during periods of crisis, arises primarily from banks' concerns that market participants will draw adverse inferences about their financial condition if their borrowing from the Federal Reserve were to become known.

spread between the primary credit rate and the target rate as a guide for the maximum rate in the market. In addition, federal funds market trading drops, as banks with lower stigma costs borrow from the discount window instead of paying high rates in the market.

This paper provides an analytical framework that captures characteristics of discount window borrowing and the federal funds market during the first year of the financial crisis, including (i) the narrowing of the spread between the discount rate and the target rate; (ii) the increased incidence of high-rate trading; and (iii) the decline in participation in the federal funds market. Lenders have imperfect information on the stigma costs of borrowers. These stigma costs can be interpreted as borrowers having different private costs of using the discount window as a funding source. The source of these costs could be something as simple as a manager of funding operations not wanting to fill out the necessary paperwork to execute a discount window loan, to a broader sentiment that banks do not want to be observed borrowing funds from the discount window during a financial crisis.[9] Differences in stigma across banks can cause both the federal funds rate to rise and discount window borrowing to increase when the spread between the discount rate and the target rate narrows. When the discount rate is high relative to the target rate, all banks stay in the funds market and few borrow from the discount window. Lenders cannot distinguish between different types of banks, and therefore all banks pay the same rate. By contrast, after the spread between the discount rate and the target rate narrows, banks that perceive a relatively lower stigma of going to the discount window ("lower-stigma" banks) do so, and exit the federal funds market. Concurrently, banks that perceive a higher stigma of going to the discount window ("higher-stigma" banks) refuse to borrow, and remain in the federal funds market. Lenders recognize that only high-stigma banks are left in the market, and so lenders can charge these banks high rates. This selection mechanism results in higher traded federal funds rates, lower federal funds market volume, and higher discount window borrowing. Moreover, any increases in

[9]Like other authors (Armantier and Copeland 2015), this paper is agnostic on the source and nature of this cost, but does suggest that some internal costs exist.

discount window stigma, which possibly could have occurred over the first year of the crisis, magnify these outcomes.[10]

After developing the framework, the paper explores its implications using federal funds market data. Both aggregate and bank-level data show how a possible increase in stigma and the resulting selection mechanism could contribute to higher observed rates in the federal funds market. The data suggest that, in aggregate, both federal funds volume brokered at rates above the primary credit rate and discount window increased during the first stages of the crisis. The empirical model results suggest that funds rates were correlated with some indicators of credit risk during the crisis in ways not evident during normal times. These indicators could be correlated with the stigma of going to the discount window, or be a proxy for the intensification of stigma as the crisis progressed. Bank-level data suggest some selection in the federal funds market, as banks that did not borrow from the discount window paid higher rates in the federal funds market than banks that did both. This selection became stronger as the spread between the primary credit rate and the target rate narrowed, coincident with the intensification of the financial crisis.

This paper is part of a long literature on discount window stigma. The literature suggests that there is a stigma associated with borrowing from the discount window that becomes more pronounced during financial crises. Friedman and Schwartz (1963) noted that such a stigma existed in the Great Depression, which may have impeded the Federal Reserve's ability to ease financial market conditions. Other stigma episodes stem from strains in the banking industry; Peristiani (1998) explored the rise in discount window stigma during the 1980s, which he attributed to worsening bank conditions. Similar to the analysis here, Ennis and Weinberg (2013) also model the effects of stigma on discount window borrowing during the recent financial crisis.[11] Finally, in recent empirical work, Armantier et al.

[10]Stigma is not the same as riskiness, and buying banks can experience a rise in stigma costs without an increase in riskiness. Still, this increase in stigma may be correlated with overall indicators of financial risk, as borrowers would be concerned that lenders would perceive banks as risky if they did go to the discount window.

[11]Calomiris (1994) empirically examines a related issue, the spreads on commercial paper as a result of the Fed's discount window lending during the Penn

(2015) also show that discount window stigma existed during the financial crisis, and banks substituted Term Auction Facility (TAF) borrowings as a result.

Still, other studies suggest a discount window stigma was present even in relatively normal times. In a theoretical model, Clouse and Dow (1999) pointed out that discount window stigma can lead to high rates in the federal funds market. Furfine (2003) concluded that stigma from borrowing at the discount window still existed even after the introduction of the primary credit program in 2003; by contrast to the previous discount window program, there was no "administration" from bank regulators in case of a borrowing.

This paper contributes to the literature in three ways. First, it provides a simple framework to illustrate how changes in the discount rate and increases in stigma can lead to selection in the federal funds market. Second, it evaluates how this stigma may have increased in aggregate in the federal funds market during the recent crisis by examining the correlation of trading at high rates and various indicators of market risk. And third, it confirms the existence of selection in the federal funds market during the financial crisis using bank-level data and panel estimation techniques to control for selection bias. Although previous literature has addressed different parts of the overall question, few studies have tied together both the theoretical implications of a simple model of a stigma with an illustration of its existence in the data.

2. Background

2.1 Monetary Policy Implementation

For many years, the discount window was one of the Federal Reserve's three main tools to implement monetary policy; the other two were open market operations and reserve requirements. Traditionally, the Federal Reserve implemented monetary policy by providing an appropriate level of reserve balances so that the federal funds rate would trade close to the target federal funds rate set

Central crisis. The lending in question, however, was to a nonbank and, in today's parlance, would have likely fallen under the auspices of the Commercial Paper Funding Facility (CPFF) or a direct loan to a nonbank counterparty, rather than the investigation of Fed lending to banks that is examined here.

by the FOMC. One way reserve balances are supplied is through open market operations. The other way that reserve balances can be supplied is through discount window borrowing.

As a result, in normal times, a bank had two ways to obtain funds to satisfy its reserve requirement, defined as an average level of funds required to be held in a bank's account at the Federal Reserve, and calculated as a percentage of a bank's total deposits. A bank could either buy funds in the federal funds market or borrow funds directly from the Federal Reserve at the discount window. Federal funds loans are unsecured advances of another bank's excess balances held in its account at the Federal Reserve. Federal funds loans are usually overnight, although some are for longer terms.

In some periods, discount window borrowing has been an integral part of monetary policy implementation, while in other times, its role has been less direct. For example, under the implementation regime in effect during the 1970s and 1980s, the FOMC declared a target for "borrowed" reserves, or those obtained from the discount window. The appropriate level of open market operations would be determined so that the level of "nonborrowed reserves" would induce the right amount of borrowing of "borrowed reserves." In turn, the level of "total reserves" would be such that funds would trade near the target federal funds rate. By contrast, through the 1990s and 2000s, discount window borrowing was not forecasted and was not an active part of the FOMC policy directive. This was the regime in place for the period studied in this paper as well.

2.2 The Discount Window

At its inception, one of the goals of the Federal Reserve System was to moderate the swings in deposits experienced by banks outside of the country's major banking centers. Loans outstanding would increase at the beginning of the growing season, while deposits would decline markedly, and after the harvest, loans would be repaid and deposits would increase. This led to a mismatch in timing between assets and liabilities for smaller banks outside of the major cities. While larger banks could provide funds to smaller ones, there were still banks with limited access to broader funding markets.

The discount window and the associated seasonal credit program were established in part because of this mismatch. In particular, the discount window was viewed as a backstop funding facility to institutions with limited access to funds through other channels that would experience these swings in assets and liabilities. Although through the second half of the 20th century fewer banks were dependent strictly on an agrarian economy, the discount window remained available for institutions that lacked other access to funding.

While the function of the discount window has remained fairly constant over its history, its administration has not. According to Madigan and Nelson (2002), from the start of the Federal Reserve System through the mid-1960s, discount window loans were extended at rates equal to or higher than short-term market interest rates. This framework is known as a "penalty rate" regime. However, the regime changed subsequently, and from the mid-1960s through 2002, the rate paid on discount window loans was pegged 25 to 50 basis points below the target federal funds rate. The amount of funds lent through the discount window was controlled through Federal Reserve requirements that banks borrow only for short-term needs, exhaust other sources of funds, and refrain from arbitrage using funds borrowed from the discount window.[12] There were two major discount window programs. The first, adjustment credit, was for banks in sound financial condition, while the second, extended credit, was available for banks with lower credit ratings. In both cases, funds were offered at a below-market rate; however, there were restrictions on the use of the funds and there was significant administration attached to these borrowings. Limits on lending to at-risk institutions were established by the Federal Deposit Insurance Corporation (FDIC) so that discount window credit would not prop up a failing institution.

On January 9, 2003, the Federal Reserve returned to a penalty-rate regime for discount window loans. Two programs were established—primary credit and secondary credit. Primary credit is the principal safety valve for ensuring adequate liquidity in the banking system; it is a backup source of short-term funds for banks in sound financial condition. Normally, primary credit is granted on a "no-questions-asked" basis, with minimal administration and no

[12]The discouragement of longer-term borrowing has been evident for much of the discount window's history; this characteristic was not unique to the period discussed here.

restrictions on its use, including for arbitraging the federal funds market. Secondary credit is available to banks not eligible for primary credit, and entails a higher level of administration.[13] At the outset, the primary credit rate was 100 basis points above the target federal funds rate and secondary credit was 150 basis points above. Artuc and Demiralp (2010) find that this change in regime reduced discount window stigma in normal times.

2.3 The Crisis

The discount window changed quickly during the first year of the crisis. The Federal Reserve lowered the relative cost of borrowing at the discount window and increased the length of the term of borrowing on two separate occasions from its usual price of 100 basis points above the target federal funds rate for typically overnight loans. On August 17, 2007, a week or so after the suspension of redemptions from two mutual funds associated with BNP Paribas, the Federal Reserve Board voted to narrow the spread between the primary credit rate and the target rate to 50 basis points from 100 basis points, the spread that had been in effect since the start of the primary credit program in January 2003. At the same time, the allowable term for primary credit borrowing was increased to 30 days. Approximately seven months later, in the wake of the takeover of Bear Stearns by JPMorgan Chase, the Board narrowed the spread another 25 basis points.

Stigma for borrowing primary credit was reportedly a concern for some banks. To address this issue, at the end of August 2007, several large banks, including Bank of America, Citibank, JPMorgan Chase, and Wachovia, borrowed from the discount window in concert in an attempt to override any discount window stigma that could possibly exist (Sidel, Ip, and Bauerlein 2007). Nevertheless, total borrowing remained low and only a moderate additional amount in loans was extended.

Still, some stigma appeared to persist. As shown in table 1, the spread between the highest brokered rate and the target rate was typically 38 basis points before August 2007. This average spread

[13] For more details, refer to http://www.frbdiscountwindow.org/programs.cfm?hdrID=14.

Table 1. Rate Definitions

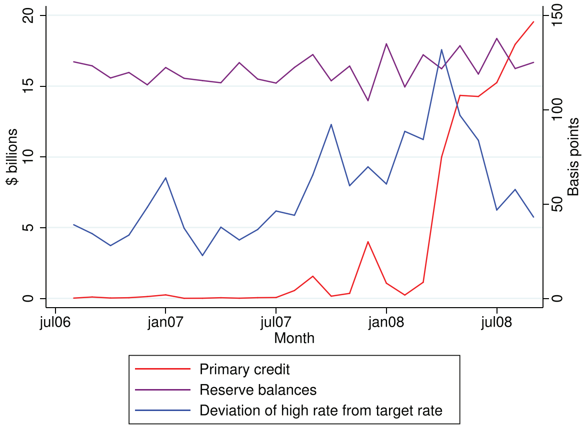
Effective Federal Funds Rate	Volume-weighted average of rates on federal funds transactions
Target Federal Funds Rate	Target rate set for trading in the federal funds market by the FOMC
High-Rate Trades	Trades on transactions late in the trading day well in excess of target rate
Primary Credit Rate	Rate on discount window borrowings from the Federal Reserve
r^{alt}	Alternative rate for federal funds lenders; usually zero

jumped to 69 basis points from August 2007 to March 14, 2008, and rose further to 82 basis points from March 17, 2008 to September 10, 2008. Moreover, the relative frequency of observing trades at wide spreads to the primary credit rate increased over the same period, as did the share of volume at high rates. During the baseline period from August 2006 to August 2007, trading occurred at rates 100 basis points above the target rate on 8 percent of the days. This share increased to 12 percent with the advent of the crisis. The share of days with trades in moderately high ranges is perhaps more striking: there were trades brokered at rates 25 to 50 basis points above the target rate on only 10 percent of the days in 2006 and 2007; this figure jumped to nearly half of the days with the beginning of the financial crisis.

Only once primary credit borrowing reached a threshold value of about $15 billion outstanding did the federal funds rate begin to fall. Notably, as shown in figure 2, this occurred around the time that primary credit equaled total Fed balances. The result was that, as shown in figure 3, federal funds market volume started to drop, and at the end of the sample period in September 2008, volume was considerably lower than it had been in March. Concurrently, the number of borrowers and lenders also fell, as shown in figure 4.

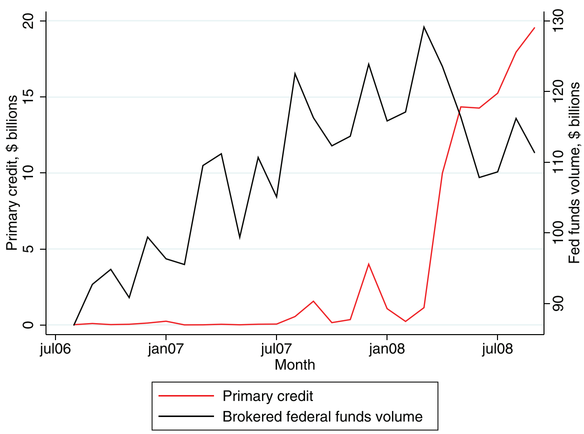
The summary statistics and distributions explored above present a few salient facts about discount window borrowing and the federal

Figure 2. Primary Credit, Reserve Balances, and the Federal Funds Rate, Monthly Averages



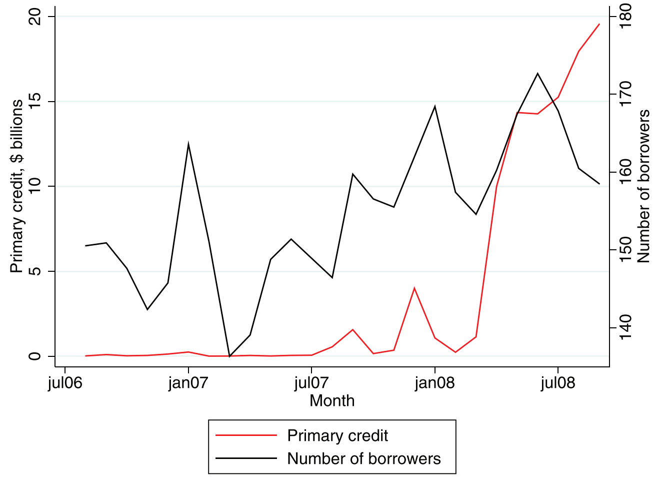
Sources: H.4.1 Statistical Release, Federal Reserve Board; Federal Reserve Bank of New York.
Notes: This figure presents monthly average information on primary credit extensions ($billions), reserve balances ($billions), and the deviation of the highest brokered federal funds rate from the target rate (basis points). The panel reflects data from August 2006 to September 2008.

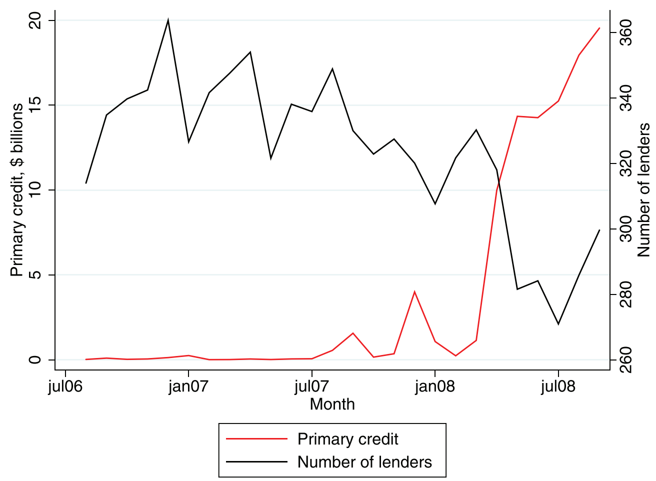
Figure 3. Primary Credit and the Brokered Federal Funds Volume, Monthly Averages



Sources: H.4.1 Statistical Release, Federal Reserve Board; Federal Reserve Bank of New York.
Notes: This figure presents monthly average information on primary credit extensions ($billions) and brokered federal funds volumes ($billions). The panel reflects data from August 2006 to September 2008.

Figure 4. Primary Credit and the Federal Funds Market Participation





Sources: H.4.1 Statistical Release, Federal Reserve Board; Federal Reserve Bank of New York.
Notes: This figure presents monthly average information on primary credit extensions ($billions) and the number of borrowers and lenders in the brokered federal funds market (number). The panel reflects data from August 2006 to September 2008.

funds market over the first year of the crisis. As the spread between the primary credit rate and the target rate narrowed, (i) the overall distribution of rates on brokered federal funds trades shifted to the right; (ii) the level of primary credit borrowing increased substantially; and (iii) federal funds volume trended down. The framework described below illustrates how this might happen.

3. Framework

This section highlights the key determinants of trading above the primary credit rate. The model illustrates how reluctance to borrow from the discount window, or "stigma," can generate a selection bias in the federal funds market. The implications of the framework will be tested in the empirical sections that follow.

The model presented here uses methodology presented in Bech and Klee (2011), Ennis and Weinberg (2013), and Afonso and Lagos (2015) by assuming a search-and-bargaining structure for the federal funds market. The methodology assumes that the borrower and lender negotiate a rate for the federal funds transaction through a Nash bargaining framework. It focuses on the decision between borrowing in the federal funds market and borrowing from the discount window in a static setting. In particular, decisions to borrow from the discount window are usually made very late in the trading session. In most situations where borrowing from the discount window is considered as an option, it is unlikely that if the negotiation fails, either party would meet either each other or another counterparty from which to borrow or lend. As such, the most reasonable outside options are, for the borrower, the discount window, and, for the lender, to leave funds in an alternative instrument, most likely in its Federal Reserve account overnight. By contrast, if a lender and a borrower disagree earlier in the trading session, each could later meet other counterparties from which to borrow and lend. In this dynamic situation, the discount window would likely not need to be considered as an option. Because the analysis is restricted to the implications of the discount window for the federal funds market, the focus is on the late-day decision.

Against this backdrop, a Nash bargaining problem consists of a disagreement point $d = (d_b, d_l)$, where d is the payoff to the borrower or lender in the case of a disagreement. In this problem, lenders have the option of lending federal funds in the market or investing in an alternative instrument. The rate earned on this alternative instrument is denoted by r^{alt}, which constitutes the disagreement point for the lender.

In addition to the posted rate paid for going to the discount window, there are other, possibly nonpecuniary costs of borrowing at the discount window.[14] We call these costs "stigma," and denote them by θ. This parameter is private information to the borrower. Moreover, while these costs may be nonpecuniary, we assume that there is a one-to-one mapping from the distribution of these costs to the real line and, as a result, we can model these costs as pecuniary.[15] This stigma cost could represent private information regarding the low quality of a bank's assets, or it could be concern that a bank's assets could be perceived as low quality if discount window borrowing is observed, as in Ennis and Weinberg (2013). More generally, it may be the case that these costs increase as general risk in financial markets climbs, as well as during a financial crisis. Taken together, these assumptions imply that $r^{dw}+\theta_k$ denotes the all-in cost of a type k bank in going to the discount window.

A solution to a Nash bargaining problem also has an agreement set A, a closed convex subset of $\mathbb{R}^2$. In our case, the agreement is an interest rate r, and an agreement is a pair $A = (-r, r)$, reflecting the fact that the trade involves a payment from the borrower to the lender. Characterizing this agreement point needs to reflect the environment surrounding the trade. The information setting for this problem is one of incomplete information. This information setting is consistent with brokered federal funds transactions, where both the borrower and the lender use a third-party intermediary to conduct the transaction. Although the lender knows its potential borrowers, it does not know exactly which one.[16]

Effectively, then, the reservation price of the borrower is unknown to the lender when bargaining. Results from Livne (1988) and others

---

[14]Goodfriend (1983) explored nonprice rationing at the discount window as a possible cost over and above the rate paid for borrowing.

[15]Other work, including Armantier et al. (2015), takes the view that rate spreads over the discount window rate can be viewed as a proxy for stigma; the analysis here also follows this reasoning.

[16]In particular, if a lender uses a broker, the lender will identify the names of potential borrowers and also the size of the loan that the lender is willing to

suggest that bargaining takes place with the disagreement point as the expected cost of borrowing from the discount window.

The setup of the game is as follows. There are two periods, with no discounting between. In period 1, a bank decides to borrow from the discount window or from the federal funds market. If the bank decides to borrow from the window, in period 2, it exits the federal funds market. If the bank decides to borrow in the federal funds market, in period 2, it meets a lender, and they bargain over the terms of trade.

This problem is solved backwards, starting in period 2. When discussing the trade with its broker, the lender attaches a probability of p_k that the borrower is of type θ_k, with p_k equal to zero if the bank does not participate in the market. Therefore, the expected disagreement point of the borrower is $-\left(r^{dw} + \sum_k p_k\theta_k\right)$. The disagreement point of the lender is the rate of return it would receive on its next best option, denoted by r^{alt}.

In addition, we posit that, given the level of funds left by the Desk in its morning open market operation, there is some bargaining power q that the borrower enjoys when bargaining with the lender. q is presumably increasing in the level of reserve balances; that is, banks pay less to borrow funds the more plentiful they are.

These assumptions imply the following form for the bargaining game:

$$\max_r \left(r - r^{alt}\right)^{1-q} \left(-r - \left(-r^{dw} - \sum_k p_k\theta_k\right)\right)^q. \tag{1}$$

During the first stages of the financial crisis, banks did not earn interest on funds kept overnight in their account at the Federal Reserve. As a result, r_{alt} may have been close to zero, particularly late in the

extend to a particular borrower. One this decision is made, the lender is generally required to accept trades with the pre-approved borrowers. As discussed in Stigum and Credenzi (2007, p. 516):

> In the fed funds market, whenever a buyer takes a seller's offering, the broker has to go back to the seller and tell her the name of the buyer and ask her if she will do the trade. The ethics of the game are such that the seller is supposed to do the trade unless she does not have a line to the buyer or her line to the buyer is filled.

day when most of these trades would have occurred.[17] q is the bargaining power of the borrower which is exogenous and, as explained above, likely a function of the level of reserve balances.[18] Finally, $\left(-r^{dw} - \theta_k\right)$ is the all-in cost of going to the discount window, the disagreement point for the borrower, should the negotiation fail.

Evaluating the Nash product gives the following expression:

$$r^* = qr^{alt} + (1-q)\left(r^{dw} + \sum_k p_k \theta_k\right), \tag{2}$$

where r^* indicates the equilibrium interest rate.

The solution to the second stage informs the participation decision in the first stage. A bank j exits the federal funds market and borrows from the discount window in period 1 if

$$r^{dw} + \theta_j \leq qr^{alt} + (1-q)\left(r^{dw} + \sum_k p_k \theta_k\right). \tag{3}$$

Rearranging a bit shows that for this to be true,

$$\theta_j \leq \frac{q\left(r^{alt} - r^{dw}\right) + (1-q)\sum_{k \neq j} p_k \theta_k}{1 - p_j(1-q)}. \tag{4}$$

Let θ^* denote the critical value of θ such that (4) holds with equality. Intuitively, this says that a bank will exit the federal funds market and go to the discount window if its cost of going to the discount window is sufficiently below that of other borrowing banks in the market, controlling for the amount of surplus captured by the borrower and lender from the bargaining problem. As a result, the critical value θ^* is a function of r^{alt}, r^{dw}, and the weighted average stigma of all other borrowing banks, $\sum_{k \neq j} p_k \theta_k$.

[17] Federal funds trading often took place after the close of other financial markets.

[18] Extensions to this model might plausibly make this an endogenous parameter that depended on the level of Fed balances.

It is fairly easy to see that this critical value increases with a step-up in stigma. For illustrative purposes, let θ_h denote a high-stigma bank. We then see that

$$\frac{d\theta^*}{d\theta_h} = \frac{p_h\,(1-q)}{1-(1-p_h)\,(1-q)} > 0. \tag{5}$$

It is then possible to see how the decision to go to the discount window changes with respect to the discount window rate. Interestingly, we see that

$$\frac{d\theta^*}{dr^{dw}} = -\frac{q}{1-(1-p_h)\,(1-q)} < 0. \tag{6}$$

If the discount window rate moves up, the critical value for staying in the market goes down. Note also that the absolute value of this effect ranges between 0 and 1, and moreover, it equals 0 only if lenders have all of the market power and equals 1 only if borrowers hold all of the bargaining power, or $q = 1$. In this way, a change in the discount window rate can generate selection among banks, where if there is a high discount window rate relative to general market rates, then all banks stay in the federal funds market; but if the discount window rate falls, then those banks with the lowest stigma cost of going to the discount window do so, and only those with higher stigma costs stay in the market. Taken together, a fall in the discount window rate leads to a decline in the number of federal funds market participants and an increase in discount window borrowing.

There are a couple of points worth discussing that are not explicitly modeled. First, discount window borrowing increases reserve balances. This change in the level of reserve balances likely affects bargaining power, q. If low-cost types borrow from the discount window and drop out of the market, then there are two opposing effects on the bargaining power of the remaining banks. At first, the increase in balances due to the discount window borrowing lowers the bargaining power of the sellers. However, at the same time, the existence of only the high-cost types in the market raises the bargaining power of the lenders. Still, it can be shown that, with reasonable parameter values, even though the cost of borrowing increases for the high-stigma types, the overall cost of funding goes down with the decrease in the discount window rate. As such, it is likely that lowering the primary credit rate is the correct policy response, in terms of providing liquidity at the least cost.

Second, the model does not explicitly account for TAF borrowing, which in principle can affect bargaining power, as it is a complementary source of liquidity. However, because TAF borrowing is for a fixed, forward-settling term, discount window borrowing and TAF borrowing cannot be perfect substitutes. That said, this imperfect substitutability is put to good use in the empirical sections that follow.

4. Empirical Findings—Aggregate Data

As presented above, there are likely two factors that boost federal funds trading volume above the primary credit rate. The first factor is increased stigma. Theoretical models and casual observation suggest that stigma could climb if bank health deteriorates, leading banks to become more reluctant to borrow from the discount window. As a result, for any given spread of trading to the target rate, as stigma increases, one would expect to see less discount window borrowing and more federal funds purchases. The second factor is selection. Holding the distribution of stigma costs constant, one would still expect to see increased trading above the primary credit rate as the spread between the primary credit rate and the target rate is narrowed, if some portion of the distribution of costs is above the primary credit rate.

In order to explore increased stigma and selection more closely, this section investigates the daily distributions of rates on brokered federal funds trades to determine whether trading at relatively higher rates is correlated with indicators of aggregate credit risk, including a bank-based credit default swap (CDS) index and the LIBOR-overnight indexed swap (OIS) spread. While these results cannot identify directly the factors modeled above, they can document correlations implied by the model.

The data used are aggregate data on federal funds trading that the Federal Reserve Bank of New York collected from federal funds brokers to construct the effective federal funds rate. The daily data cover 2006 to 2008 and consist of the rates at which trades were brokered and the volumes of trades at those rates. The analysis focuses on three ranges for federal funds trading: trading that occurs at rates more than 100 basis points above the target rate; greater than 50 basis points up to 100 basis points above the target rate; and greater than 25 basis points up to 50 basis points. These series are plotted as a share of daily volume in figure 5. As is evident from the figure, high-rate trading at all spreads to the primary credit rate increased as the crisis intensified.

The specification tests whether the effects of aggregate risk indicators change with the spread of the primary credit rate to the target rate. To this end, the sample is split into three periods: (i) the 100 basis point regime from August 2006 to August 14, 2007, the day before the narrowing of the spread between the target and the primary credit rate; (ii) the 50 basis point regime from August 17 and March 14, 2008, the day before the second spread narrowing; and (iii) the 25 basis point regime March 17, 2008 to September 10, 2008. See table 2. The sample has 529 daily observations.

4.1 Specification

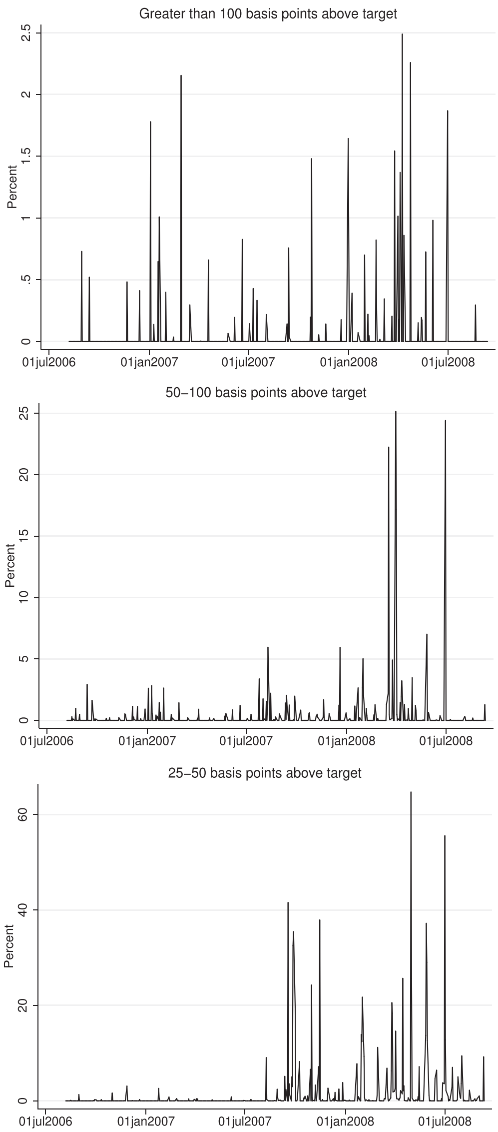
To investigate the determinants of trading at selected spreads to the target rate, let V_{it} represent volume brokered at selected ranges to the target rate, denoted by i. Furthermore, let $E\ (V_{it}) = \mu_{it}$, the mean volume brokered at a particular spread to the target federal funds rate. μ_{it} is specified as

$$\mu_{it} = x_t \beta_i = \beta_0 + \beta_{1ij} risk_t + \beta_{2i} calendar_t. \tag{7}$$

The mean volume brokered at a particular spread to the target rate depends on a number of factors, including $risk_t$, which is a vector of indicators of general financial risk; $fftar_t$ is the target federal funds rate, and $calendar_t$ is a vector of calendar effects. The β coefficients are allowed to vary both by the spread to the target for the trading volume, i, as well as the primary credit spread in effect, j, at time t.

There are three characteristics of the dependent variable that influence the chosen functional form. First, the dependent variable is strictly non-negative, suggesting that a transformation of the variable is appropriate. Second, there are a number of observations with the value of zero for which there is significant economic meaning, ruling out the usual log transformation. And third, the observation of

Figure 5. Trading above the Target Rate



Source: Federal Reserve Bank of New York.
Note: The graphs display the share of federal funds volume brokered at selected spreads to the target federal funds rate.

Table 2. Trading at High Rates in the Federal Funds Market (percent)

	Average Spread between High and Target Rate	Percent of Days			Percent of Brokered Volume		
		> 100	50–100	25–50	> 100	50–100	25–50
Aug. 1, 2006 to Aug. 14, 2007	38	8	19	10	0.04	0.14	0.06
Aug. 17, 2007 to March 14, 2008	69	12	29	42	0.05	0.30	1.76
March 17, 2008 to Sep. 10, 2008	82	12	25	56	0.12	0.79	3.10

Source: Federal Reserve Bank of New York.
Notes: This table presents summary statistics for trading at high rates in the federal funds market over three daily sample periods. Percent of days is the percent of trading days on which nonzero volume was brokered at rates within or above selected spreads to the target federal funds rate. Percent of volume is percent of dollar volume brokered at rates within or above selected spreads to the target federal funds rate.

significant volume above the primary credit rate is generally rare. To address these three issues, we use a generalized linear model with the functional form $E\left(V_{it}|x_t, V_t\right) = V_t \Pr\left(x_t\beta_i\right)$, where V_t is total daily transaction volume and Pr() is the probability of a transaction occurring in rate range i.

The functional form of Pr() is chosen to be a Gompertz distribution, appropriate for modeling an infrequent event at the higher ranges of the support of a distribution. As federal funds trades at relatively high spreads to the primary credit rate are infrequent, this distribution is appropriate. It has the functional form

$$\Pr(x_t\beta_i) = 1 - \exp(-\exp(x_t\beta_i - 1)). \tag{8}$$

Combining this distributional assumption with the mean specification outlined above, the final specification is

$$E\left(V_{it}|x_t, V_t\right) = V_t \Pr\left(x_t\beta_i\right) = V_t\left(1 - \exp(-\exp(x_t\beta_i - 1))\right). \tag{9}$$

In addition, Newey-West errors that are corrected for heteroskedasticity and autocorrelation are used, as the dependent variable is nonlinear and these are time-series data.

4.2 Results

Table 3 shows three sets of results, each reflecting how various factors are correlated with federal funds volume brokered at rates at selected spreads above the target federal funds rate ("high-rate funds market trading"). The first set of results displays information for rates 100 basis points or more above the target rate (columns 1 and 2); the second, 50 to 100 basis points above (columns 3 and 4); and the third, 25 to 50 basis points above (columns 5 and 6). Two specifications are highlighted in each set, and the results are presented in terms of marginal effects for ease of interpretation.

Overall, high-rate funds market trading and measures of perceived bank risk are significantly correlated, and this correlation deepened as the crisis wore on. Moreover, the sign of the correlation is perhaps consistent with some form of selection in the federal funds market. These results are evident in two specifications that use different indicators of bank risk: the CDS index (labeled A) and the LIBOR-OIS spread (labeled B). In the 100 basis point discount

Table 3. Federal Funds Trading at Selected Spreads above Target Rate

	Above 100 bp		50–100 bp		25–50 bp	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
A. CDS Index						
100 bp Regime	−0.836**		4.759**		26.260**	
	(0.27)		(0.42)		(1.00)	
50 bp Regime	0.719**		−5.448**		−32.668**	
	(0.27)		(0.44)		(1.41)	
25 bp Regime	0.734**		−3.629**		−28.614**	
	(0.27)		(0.59)		(1.02)	
B. LIBOR-OIS Spread						
100 bp Regime		−1.972**		3.611**		19.004**
		(0.49)		(0.21)		(0.71)
50 bp Regime		1.810**		−3.766**		−23.748**
		(0.49)		(0.31)		(0.87)
25 bp Regime		2.024**		−10.494**		−20.330**
		(0.49)		(0.85)		(1.45)
C. Spread of Repo Rate to Target Rate						
100 bp Regime	0.063	0.222**	2.233**	1.388**	46.912**	30.200**
	(0.06)	(0.09)	(0.30)	(0.20)	(8.31)	(5.69)
50 bp Regime	−0.102	−0.275**	−2.242**	−1.396**	−45.750**	−29.693**
	(0.06)	(0.09)	(0.28)	(0.17)	(8.30)	(5.68)
25 bp Regime	−0.060	−0.200*	−2.074**	−0.507*	−46.717**	−28.974**
	(0.07)	(0.10)	(0.36)	(0.20)	(8.32)	(5.73)

(*continued*)

Table 3. (Continued)

	Above 100 bp		**50–100 bp**		**25–50 bp**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
D. TAF						
50 bp Regime	0.001**	−0.000	0.010**	0.004**	0.045**	−0.025**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
25 bp Regime	−0.002**	−0.002**	−0.019**	−0.042**	−0.021**	−0.038**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
E. Calendar Effects						
Month End	0.110**	0.117**	1.190**	1.405**	3.649**	3.633**
	(0.01)	(0.01)	(0.12)	(0.19)	(0.22)	(0.22)
P&I Day	0.026	0.021	0.076	0.213**	1.427**	1.338**
	(0.02)	(0.02)	(0.15)	(0.07)	(0.20)	(0.21)
50 bp Regime	−0.194**	−0.115**	1.670**	0.484**	15.402**	9.653**
	(0.07)	(0.04)	(0.16)	(0.11)	(0.80)	(0.48)
25 bp Regime	0.226**	0.154**	2.949**	10.031**	18.104**	13.609**
	(0.07)	(0.06)	(0.30)	(0.89)	(0.64)	(2.13)
Observations	529	528	529	528	529	528

Notes: This table presents estimated marginal effects for the regression $E(V_{it}|x_t,\ V_t) = V_t \mathrm{Pr}(x_t\beta_i)$, where $\mathrm{Pr}(x_t\beta_i) = 1 - \exp(-\exp(x_t\beta_i - 1))$ and $x_t\beta_i = \beta_0 + \beta_{1ij}\,risk_t + \beta_{2i}\,calendar_t$. Dependent variable is daily brokered volume in \$billions. Financial series are in percentage points. TAF is in \$billion. Calendar effects are 0/1 indicators. Newey-West heteroskedastic-autocorrelated corrected errors are in parentheses. ** denotes significance at the 5 percent level.

window regime, as these measures of bank risk increased, volumes at the highest rates actually fell, while trading at more moderate spreads to the target rose. Evaluated at the means of the variables, the coefficients imply that a 1 percentage point increase in the CDS index is associated with an 83 basis point decrease in the volume of trades 100 basis points above the target rate, and a 26 percentage point increase in the volume of federal funds trades brokered at rates 25 to 50 basis points above the target rate. Results using the LIBOR-OIS spread instead of the CDS index are qualitatively similar, suggesting a reasonably robust result. Even though market strains may have been appearing, borrowing banks with low stigma pooled with borrowing banks with high stigma, leading to some increase in rates, but not to extremes.

Once the crisis began and the primary credit spread narrowed, high-rate funds market trading at rates more than 100 basis points above the target rate increased as these measures of bank risk climbed. At the same time, high-rate trading at rates between 25 and 100 basis points above the target fell somewhat. The differential effects according to selected ranges above the target rate suggest that there could be some selection in observed trades. That is, banks with low stigma costs went to the discount window, but banks that had high stigma costs remained in the market and were forced to borrow funds at higher rates. Taken together, these results suggest that as banks were perceived as more risky, and the discount window spread narrowed, rates in the federal funds market went up, not down. These empirical results are consistent with the Bernanke (2008) observation discussed in the introduction; namely, banks may become reluctant to borrow at the discount window as financial strains intensify.

One caveat to the estimated coefficients for these bank-risk results is that there may be some endogeneity issues. That is, the high-rate trading in the federal funds market may boost measures of bank risk, including the LIBOR-OIS spread as well as the CDS index. As a result, lenders may be unwilling to extend federal funds loans to banks that exhibit high risk. This is, of course, a weakness of this type of aggregate specification. That said, as discussed in the model section, the institutional norm in the market was that if a credit line was available, lenders were obligated to extend loans to borrowers. Moreover, even if lenders did want to cut credit lines, they usually did so on a case-by-case basis, and did not do so very frequently. Taken together, these results suggest that selection on the borrowing side could be driving the results. However, this issue will be investigated more completely in the bank-level analysis below.

The next set of results (labeled C) explores the correlation of high-rate funds market trading with the demand for safe collateral. While the demand for safe collateral suggests overall market risk, it is not specific to banks. During the 100 basis point primary credit regime, the Treasury GC repo rate and high-rate funds trading tended to co-move, reflecting the more general behavior of short-term money market rates in normal times. As the first year of the crisis wore on, however, the correlation between high-rate trading and the Treasury GC repo rate became negative. The Treasury GC repo rate tends to fall with heightened demand for safe collateral, which occurs during periods of market stress. Evaluating the marginal effects at the mean of the variables suggest that for every percentage point decrease in the Treasury GC repo rate below the target rate, the share of funds brokered within the 50 to 100 basis point range above the target increases by about 2 percentage points relative to normal times. The effect for more modest ranges above the target is larger: for every percentage point decrease in the repo rate below the target rate, the share of volume brokered in the 25 to 50 basis point range above the target increases by about 35 percentage points relative to the baseline period. Taken together, then, the coefficient suggests that high-rate federal funds trading increased concurrently with elevated demand for safe collateral, with no specific selection effects related to widespread collateral demand.

Correlations of high-rate federal funds market trading with TAF borrowing are displayed in the rows labeled D. While there is some variation, the results generally suggest that increased TAF borrowing was associated with more higher-rate trading under the 50 basis point primary credit regime, but with less under the 25 basis point regime. Dollar for dollar, the effect appears to be largest on volume brokered at rates 25 to 100 basis points above the target rate. TAF borrowings can provide some certainty regarding a bank's level of reserve balances. Because TAF funds were auctioned two days before settlement, endogeneity concerns are likely minimal.

The final set of rows control for various calendar effects (labeled E). High-rate trading was somewhat more evident on month-end and on days when Fannie Mae and Freddie Mac made principal and interest (P&I) payments on their mortgage-backed securities (MBS). The marginal effects for these suggest the share of trading at relatively elevated ranges to the target rate increased a percentage point or so on those days. The last couple of lines of the table show the level of high-rate trading during the 50 basis point and 25 basis point primary credit regimes, independent of bank credit risk or demand for safe collateral. Overall, high-rate funds market trading became more prevalent as the crisis intensified.

To summarize, the results presented here suggest that some pickup in high-rate trading was due to overall increases in financial risk. At the same time, the effects of increases in bank-specific financial risk appear consistent with selection in the funds market. In addition, TAF borrowings appear to dampen high-rate trading in the federal funds market. These results inform the bank-level results, which are discussed below.

5. Empirical Findings—Bank-Level Data

This section uses bank-level data to evaluate the connection between high-rate funds market trading, primary credit, and various bank characteristics. However, a simple panel regression will not be appropriate, as the model illustrates that rates and borrowing are endogenously determined. Also, the aggregate empirical results suggest that funds market participation exhibits selection that is correlated with overall bank risk. Against this backdrop, the empirical strategy follows a framework outlined by Semykina and Wooldridge (2010) to control for endogeneity of primary credit borrowing and selection bias in the federal funds market.

5.1 Data Construction

The data are constructed by combining bank-level daily data, market-level daily data, and bank-level quarterly data.

The bank-level daily data on federal funds rates is constructed using proprietary transaction-level data from the Fedwire Funds Service, using an algorithm pioneered by Furfine (1999) to match and form plausible overnight funding transactions, likely related to the federal funds market.[19] However, because there is no independent way to verify if these are actual federal funds transactions, identified trades and characteristics of these trades are subject to error.[20] These errors potentially include correspondent transactions (transactions done by one bank on behalf of another bank), capturing funding transactions outside of the federal funds market, or coincidence. Although these are possible weaknesses, they may not be critical for this analysis. If the high-rate trade reflects a correspondent relationship, it still reflects an unwillingness to go to the discount window. In addition, if banks are obtaining funding outside of the federal funds market at rates higher than the primary credit rate, then there is still aversion to using the discount window; the actual funding source is less critical. And finally, while trades could be simply coincidence, this likelihood is minimized by using trades that match rates observed in the brokered data. Specifically, the high-rate data critical to the analysis below were cross-validated with brokered data for many sample days.

The data cover August 1, 2006 to September 11, 2008. The transaction data contain information on the amount of the transaction, the implied interest rate of the identified transaction, and the lender and borrower in the trade. From there, these data are summarized on a daily basis, and the high rate on the day and the total funds bought are calculated.

Other bank-account activity variables are also generated from proprietary Federal Reserve databases. The data on reserve account balances are constructed from the Federal Reserve's database of banks that report reserve balance and related information on the weekly "Report of Transaction Accounts, Other Deposits, and Vault

[19]The algorithm matches an outgoing Fedwire funds transfer sent from one account and received by another with a corresponding incoming transfer on the next business day sent by the previous day's receiver and received by the previous day's sender. This pair of transfers is considered a federal funds transaction if the amount of the incoming transfer is equal to the amount of the outgoing transfer plus interest at a rate consistent with the rates reported by major federal funds brokers. Similar data were used by Demiralp, Preslopsky, and Whitesell (2006), Bartolini, Hilton, and McAndrews (2010), and Afonso, Kovner, and Schoar (2011).

[20]Armantier and Copeland (2015) discuss some of the important shortcomings of the data.

Cash."[21] Information on primary credit and TAF borrowings is from internal databases; current borrowings are available on the Board's public website.[22] The daylight overdraft information is calculated from the same database used to construct the federal funds transactions. Peak daylight overdrafts is the maximum amount a bank overdrafts its Fed account on a particular day.

The daily data are then paired with bank-level information from the Call Report and other regulatory reports that are issued on a less frequent basis, to capture key balance sheet and reserves-related items. Also included are some of the daily financial market indicators studied in section 4 to control for overall market conditions on the day.

After the combination of all data sets, the data are summarized by week. For the purposes of testing for selection, the sample is split into the three regimes described above: August 2006 to August 2007, the 100 basis point primary credit regime; August 2007 to March 2008, the 50 basis point regime; and March 2008 to September 2008, the 25 basis point regime. The sample ends at September 11, 2008, immediately before the failure of Lehman Brothers.

5.2 Estimation Framework for Testing and Correction

As described in the model and illustrated in the aggregate empirical results, a regression that explores the dependence of high-rate trading on discount window borrowing likely suffers from both selection and endogeneity problems.

Semykina and Wooldridge (2010) develop a panel data estimator that controls for both selection bias and endogenous regressors. In the spirit of a traditional Heckman selection model, the diagnostic and estimation procedure is in three steps. Using the notation in Semykina and Wooldridge, in the first step, a probit model is estimated for each time period:

$$Pr\left(s_{it} = 1 | z_i\right) = \Phi\left(z_{it}\delta_t^a + \bar{z}_i\xi_t^a\right) \tag{10}$$

[21] Reporting form FR2900, http://www.federalreserve.gov/apps/reportforms/default.aspx.

[22] http://www.federalreserve.gov/newsevents/reform_discount_window.htm.

for each bank i during week t. s_{it} equals 1 if bank i borrowed from the discount window in week t, z_{it} is a vector of exogenous variables, and $\bar{z}_i$ is the mean of these variables for each i over all t. Importantly, the z_{it} should be observed for all i, regardless of whether the bank borrowed from the discount window. The means of exogenous variables $\bar{z}_i$ control for unobserved fixed effects and are used to correct for possible selection bias.

The inverse Mills ratios are then calculated for each t, which take the form

$$\hat{\lambda}_{it} = \lambda\left(z_{it}\hat{\delta}_t^a + \bar{z}_i\hat{\xi}_t^a\right), \tag{11}$$

where $\hat{\delta}_t^a$ and $\hat{\xi}_t^a$ are the estimated coefficients from equation (10). With these Mills ratios, the selection bias test can now be executed. A fixed-effects two-stage least squares model is estimated only on the sample of institutions that borrowed from the discount window,

$$\begin{aligned} ffhighdev_{it} &= c_i + \alpha primary_{it} + x_{it}\beta + \gamma pcspread_t \\ &\quad + \rho_{spread}\hat{\lambda}_{it} + \epsilon_{it1}. \end{aligned} \tag{12}$$

The coefficients ρ differ according to primary credit regime to capture changes in selection as the spread between the primary credit rate and the target rate narrowed. Selection bias is indicated by significant coefficients on the $\hat{\lambda}_{it}$ terms.

To control for endogeneity, a subset $z_{it1} \subset z_{it}$ is used as instruments in the estimation, and to control for the selection, one variable is excluded from z_{it1} but included in z_{it}. This construct conforms to the Semykina and Wooldridge (2010) requirements that one instrumental variable is necessary to control for the endogenous regressor, and another is necessary to control for the selection.

If selection is present and endogeneity is suspected, pooled two-stage least squares is run on the following specification, which controls for both selection and endogeneity:

$$\begin{aligned} ffhighdev_{it} &= c_i + \alpha\widehat{primary}_{it} + x_{it}\beta + \gamma pcspread_t \\ &\quad + \bar{z}_i\eta + \rho_{spread}\hat{\lambda}_{it} + \epsilon_{it1}, \end{aligned} \tag{13}$$

where $\widehat{primary}$ indicates that primary credit is instrumented. With this construction, the α coefficient on the *primary* term should indicate the true relationship between rate paid on federal funds and the level of discount window borrowing, while the ρ_{spread} coefficients indicate the degree of selection according to time period.

$$primary_{it} = x_{it}\beta + z_{it1}\gamma + \bar{z}_i\eta + \rho_{spread}\hat{\lambda}_{it} + \mu_{it} \tag{14}$$

A key factor in the success of this approach is identifying appropriate instruments. The specification uses TAF borrowing to control for selection bias, and daylight overdrafts to instrument for primary credit. The operational framework surrounding each of these forms of Federal Reserve credit make them good candidates to help identify the effects of discount window stigma on high-rate federal funds trades.

Turning first to correcting for selection bias, it is important to identify a factor that is correlated with the probability of borrowing from the discount window, but not correlated with high-rate federal funds market trading that is a function of stigma. TAF borrowing likely affects the probability of borrowing from the discount window in a given week, but is uncorrelated with unexpected account shortfalls and was generally free from stigma. In particular, if a bank has sufficient funds in its account from TAF borrowing, it may not need to borrow primary credit. Alternatively, if a bank had a general need for funds, it may choose to borrow at either facility. However, the two forms of reserve bank credit were not perfect substitutes. Because TAF auctions occurred at predetermined intervals (funds were usually auctioned on Tuesdays, and settled on Thursdays) banks could not borrow from the TAF to cover unexpected daily funding needs. Moreover, in part because of this settlement structure, TAF borrowing was generally free from stigma.

Turning next to correcting for endogeneity, it is important to identify a factor uncorrelated with unobserved stigma and indirectly related to federal funds rates. Daylight overdrafts likely satisfy these requirements. Specifically, they are a byproduct of the thousands of payments banks make each day. These payments use funds in a bank's Fed account. During this sample period, banks with insufficient funds could still make payments, but would incur a "daylight overdraft," to be repaid before the end of the banking day. If a bank

Table 4. Bank-Level TAF and Daylight Overdraft Summary Statistics, $Billions

	Mean	Median	Std. Dev.	Obs.
TAF Borrowings				
Borrowed Primary Credit	1.053	0	6.599	2,421
Did Not Borrow Primary Credit	0.114	0	1.571	123,640
Peak Daylight Overdrafts				
Borrowed Primary Credit	1.213	0.004	6.409	2,421
Did Not Borrow Primary Credit	0.273	0.002	2.612	123,640

Source: Federal Reserve Board.
Note: This table reports summary statistics for the variables used to identify selection and to control for endogeneity.

was short, they would need to borrow funds, either in the federal funds market or from the discount window. Failure to do so meant a bank incurred an overnight overdraft, with a hefty fee of 400 basis points above the effective federal funds rate.

Daylight overdrafts had two important qualities that make them suitable as an instrument. First, any individual bank's daylight overdraft was presumably private information: while a counterparty would likely have some idea of the amount of payments or loans it sent to any one bank, it would likely not have information on that bank's payment activity with other institutions. Second, demand for federal funds as a result of daylight overdrafts was likely independent of credit risk, market stress, or stigma. Rather, daylight overdrafts independently shift a bank's demand curve for reserve balances; these balances could be obtained in the federal funds market or at the discount window. Presumably, banks in need of funds late in the day in order to cover a daylight overdraft were willing to pay high rates in the federal funds market or take out a discount window loan, as both of these options had lower costs than the 400 basis point overnight overdraft fee.

Summary statistics on TAF borrowing and daylight overdrafts corroborate the usefulness of these indicators as instrumental variables. Table 4 displays basic summary statistics on TAF borrowings and peak daylight overdrafts at an institution-week level according to whether primary credit borrowing was also observed. TAF borrowings are substantially higher for those institutions that borrowed primary credit relative to those that did not. While most institutions did not borrow from the TAF, the standard deviation of borrowings for those that borrowed primary credit is much higher than that for those that did not.

Turning to daylight credit, mean peak daylight credit is substantially higher for those banks that borrowed primary credit, by about $1 billion on average. In addition, the median peak overdraft is substantially higher for banks that also borrowed primary credit. Similar to TAF borrowings, the standard deviation of peak daylight overdrafts is also higher for the banks that borrowed primary credit than those that did not.

5.3 Baseline Panel Estimates and Results

Before proceeding to the selection tests, it is instructive to have baseline panel regression results as a comparison. This is specified as

$$\begin{aligned} ffhighdev_{it} = c_i + \alpha_t^1 primary_{it} + \alpha^2 daylight_{it} \\ + \alpha_t^3 TAF_{it} + x_{it}\delta + c_i + q_{pc}\zeta + \epsilon_{it}. \qquad (15) \end{aligned}$$

Table 5 displays summary statistics for the variables in the specification. The dependent variable $ffhighdev_{it}$ is the average of daily deviations of the highest observed rate for funds bought from the effective rate for bank i during week t. Because the discount window generally served as a marginal source of funds in the sample period, the highest rate paid is a close proxy to an actual reservation price. Furthermore, comparing the highest rate paid with the effective rate gives an idea of rates paid relative to the market average.

The first set of independent variables reflect borrowings and account activity at the Federal Reserve. $primary_{it}$ is the sum of primary credit borrowing by the bank over the week. The coefficient on this factor is permitted to vary over primary credit regimes. According to the model presented above, rates paid by banks that borrowed primary credit should be lower than those for banks that did not. The second set of variables are daylight overdrafts and TAF

Table 5. Bank-Level Summary Statistics

	Mean	Std. Dev.	Min.	Max.
Regime Probability				
100 bp Regime	0.50	0.50	0	1
50 bp Regime	0.28	0.45	0	1
25 bp Regime	0.23	0.42	0	1
Primary Credit ($Billions)	0.01	0.34	0	17.5
100 bp Regime	0.00	0.01	0	1.16
50 bp Regime	0.00	0.09	0	15
25 bp Regime	0.01	0.33	0	17.5
Daily Total Balances ($Billions)	0.01	0.08	0	7.18
Days in Market	0.62	1.48	0	5
Funds Bought ($Millions)	0.31	2.82	0	92.92
Assets ($Billions)	1.42	23.42	0	1,392.27
Required Reserves ($Billions)	0.03	0.25	0	7.89
N	120,464			

Sources: Federal Reserve Board; Bloomberg; Federal Reserve Bank of New York.
Notes: This table reports descriptive statistics for the variables used in the panel regressions. The sample includes bank-week observations from October 2006 to September 2008.

borrowings. As discussed above, these will be used as instruments in the selection and correction specifications. Both vary at the bank-week level.

The next set of factors control for bank characteristics, denoted x_{it}. One control variable is the number of days bank i participated in the federal funds market during week t. More frequent participation can indicate that the bank is generally a "market maker" in the federal funds market, while infrequent participation suggests transacting to satisfy short-term liquidity needs. The vector also contains bank-specific information, including the assets of the bank, weekly average amount borrowed in the federal funds market, reserve requirements, and average reserve balance holdings over a week.

A Hausman test rejects the hypothesis that a random effects model is sufficient to control for individual-level effects. As a result, fixed effects are assumed in the estimation, denoted by c_i. We include time period controls as well, indicated by q_{pc}, which correspond to the primary credit regime. The error term ϵ_{it} captures all other unobserved factors.

The first column of table 6 presents the results. Primary credit borrowing does not appear to be significantly correlated with rates paid on federal funds. By contrast, daylight overdraft activity is correlated with higher federal funds rates, suggesting that banks with elevated funding demands are forced to pay more to obtain funds. Very roughly, the estimated coefficient on the daylight overdraft term suggests that for each one standard deviation of peak overdrafts, the spread on the high-rate trading above the effective rate increases about 5 to 6 basis points. TAF borrowing in the 50 basis point spread primary credit regime is not statistically significantly associated with higher rates paid in the federal funds market, but in the 25 basis point regime, the effect is positive and significant. This result suggests that the TAF may have offset the need to buy high-rate funds in the market early in the program, but later in the program, TAF borrowers paid significantly higher rates in the federal funds market than other borrowers.

Turning next to bank-level factors, neither the number of days in the market nor the amount of funds borrowed appear to be significantly correlated with high rates paid in the federal funds market. Asset size does not significantly predict high rates. Also, required reserves and total reserve balances are not associated with higher rates paid.

Some of the broad financial market variables are significantly correlated with rates paid in the federal funds market. Although the CDS index is somewhat surprisingly negatively correlated with trading in the federal funds market, it could be a result of selection in the federal funds market. At the same time, the spread of the repo rate to the target rate has an intuitive sign, with a more negative spread associated with higher-rate trading.

The intercept terms, reported in the last three lines of the table, suggest that rates rose as time wore on. On average, high rates were about 12 basis points higher in the 25 basis point regime than in the 100 basis point regime, after controlling for the factors listed above.

5.4 Controlling for Selection and Endogeneity

With the baseline panel results in mind, the next columns present results that test and control for selection and endogeneity.

Table 6. Controlling for Selection Bias in and Endogeneity of Discount Window Borrowing

	Panel (1)	Selection (2)	Corrected (3)	Corrected (FHLB) (4)	Corrected (90^{th}) (5)
Primary Credit	2.053	−30.20**	−2.262*	−1.388	−0.00977
	(4.987)	(10.67)	(0.998)	(0.897)	(0.00608)
50 bp Regime	−1.240				
	(5.161)				
25 bp Regime	−1.607				
	(4.921)				
Peak Daylight Overdrafts	0.742***				
	(0.205)				
TAF Borrowing	−0.0438				
	(0.0525)				
25 bp Regime	0.276*				
	(0.138)				
Number of Days in Market	−0.183	3.787	0.119	0.105	0.00657
	(0.106)	(2.912)	(0.588)	(0.586)	(0.00551)
Amount Borrowed	0.202	−2.767*	0.296	0.861	−0.00172
	(0.166)	(1.381)	(0.481)	(0.552)	(0.00264)
Total Assets	−0.00111	−0.0972	−0.00674	−0.00134	−0.0000762
	(0.00124)	(0.0665)	(0.0230)	(0.0240)	(0.000176)
Required Reserves	1.423	34.95	14.17	13.04	0.106
	(1.933)	(21.75)	(9.873)	(10.40)	(0.0725)
Total Reserve Balances	−1.839	−33.67	−12.22	−9.890	−0.00477
	(2.007)	(19.86)	(14.40)	(13.11)	(0.0609)
FHLB Borrowings				−489.0	
				(357.6)	
CDS Index	−3.203***	−15.56	−4.723	−3.610	−0.0371
	(0.401)	(10.45)	(2.831)	(2.804)	(0.0262)
Repo-Target Spread	−0.0602***	0.114	−0.0451	−0.0401	−0.000873***
	(0.00405)	(0.0907)	(0.0277)	(0.0247)	(0.000232)
Selection					
100 bp Regime		31.82*	1.459	1.141	−0.0125
		(12.55)	(1.474)	(1.760)	(0.0145)
50 bp Regime		1.086	−4.594	−2.422	−0.0754**
		(6.914)	(3.389)	(3.081)	(0.0258)
25 bp Regime		−26.49*	−9.434***	−8.331***	−0.0607**
		(10.48)	(2.748)	(2.144)	(0.0199)
Constant	9.317***	−33.35	4.838	5.278	0.132***
	(0.602)	(22.21)	(3.863)	(4.247)	(0.0357)
50 bp Regime	9.232***	92.35**	25.87***	21.03***	0.254***
	(0.411)	(30.32)	(6.318)	(5.793)	(0.0643)
25 bp Regime	12.53***	149.6**	36.83***	32.61***	0.260***
	(0.638)	(47.48)	(8.742)	(8.017)	(0.0727)
N	21,390	642	642	642	643
Number of Banks	547	129	129	129	129
Adj. R-sq.	0.1574	0.0052	0.269	0.324	0.176

Notes: Dependent variable is the deviation of the average observed high rate paid for federal funds from the effective rate. Specifications include Mundlak-Chamberlain fixed effects.
Column 1 presents estimated coefficients for the baseline panel regression:
$f\,fhighdev_{it} = c_i + \alpha_t^1 primary_{it} + \alpha^2 daylight_{it} + \alpha_t^3 TAF_{it} + \gamma days_{it} + x_{it}\delta + c_i + q_{pc}\zeta + \epsilon_{it}$.
Column 2 presents estimated coefficients for the selection test:
$f\,fhighdev_{it} = c_i + \alpha primary_{it} + x_{it}\beta_t + \gamma pcspread_t + \rho_{spread}\hat{\lambda}_{it} + \epsilon_{it1}$.
Columns 3–6 present estimated coefficients that control for selection and endogeneity:
$f\,highdev_{it} = c_i + \alpha\widehat{primary}_{it} + x_{it}\beta_t + \gamma pcspread_t + \bar{z}_i\eta + \rho_{spread}\hat{\lambda}_{it} + \epsilon_{it1}$.
Cluster-robust standard errors are in parentheses on panel and selection estimates. Bootstrapped standard errors on corrected estimates. *, **, and *** denote significance at the 5 percent, 1 percent, and 0.1 percent level, respectively.

As shown in column 2 of table 6, selection in the federal funds market intensified as the spread between the primary credit rate and the target rate narrowed. The coefficient on the ρ_{25} term implies that an unobserved factor suggesting a higher propensity to borrow from the discount window is correlated with lower rates paid in the federal funds market. If this factor is "stigma," then lower stigma leads to lower federal funds rates paid. Consistent with the model's predictions, then, banks that were willing to borrow from the discount window did not pay as high rates for funds. The coefficient suggests that borrowing from the discount window was associated with about a 25 basis point decrease in the average high rate paid. In addition, there appears to be positive selection in the federal funds market when the spread is 100 basis points. Because this was the spread in a period of relative calm, it may be the case that borrowing from the window occurred on days with specific pressures in the funds market, such as quarter-end reporting dates. The magnitude of the coefficient suggests that banks paid an average of about 30 basis points higher for high-rate funds if going to the discount window during normal times. During the 50 basis point regime, the correlation between borrowing from the discount window on high rates paid was not significant.

However, examining the selection terms by themselves does not give a complete picture. As indicated by the coefficient on primary credit, borrowing \$1 billion in primary credit is associated with a 30 basis point lower peak federal funds rate. Taken with the selection terms, the results suggest that borrowing from the discount window substantially reduced funding costs during the 25 basis point regime, somewhat damped them during the 50 basis point regime, and probably had a minimal net effect in the 100 basis point regime.

The final step of the estimation procedure corrects for both the endogeneity of primary credit and the selection for federal funds rates. These results are presented in the third through fifth columns of table 6. The coefficient on the primary credit term is negative and significant, suggesting that banks that borrowed primary credit paid lower federal funds rates. The point estimate suggests that for each \$1 billion borrowed, peak rates fell by about 2 basis points. Taken with the highly statistically significant and negative coefficients on the selection terms, rates appear to be substantially lower for banks willing to go to the discount window. Indeed, during the 25 basis point regime, the net effect of borrowing $1 billion from the discount window was a peak funds rate that was about 10 basis points lower. Overall, the results are consistent with the model and suggest that banks that borrowed from the discount window paid lower rates in the federal funds market, and this phenomenon became stronger as the primary credit rate spread narrowed and the crisis intensified.

One caveat is that banks may have been using another source of funding, other than the discount window, and that our results might mask the effect of this other source. In particular, as discussed in Ashcraft, Bech, and Frame (2010), many institutions substituted Federal Home Loan Bank loans for discount window loans, as the FHLB loans generally had lower interest rates. As a robustness check, column 4 of table 6 tests if there was any influence on rates paid in the federal funds market that depended on FHLB borrowings; we use the level of FHLB borrowings as reported quarterly on the Call Report as a control variable in our specification. Interestingly, the level of FHLB borrowings is not correlated with paying lower rates in the federal funds market. Controlling for FHLB borrowings shows that primary credit borrowings are still weakly correlated with lower rates paid in the federal funds market; each $1 billion borrowed is associated with a 1 basis point lower rate paid in the market; selection-term coefficients are roughly the same as in column 3. More generally, even if rates only weakly fall for each dollar borrowed (the intensive margin), rates do fall with a willingness to go to the discount window (the extensive margin).

Finally, some may question the choice of dependent variable. Although the high rate paid on the day is a metric that is consistent with the model presented above, there may be some biases due to data limitations and also to using an extreme value of a distribution. Other plausible suggestions include the 90th percentile of trades, expressed as a deviation from the effective rate and measured over a week. Results with this dependent variable are shown in the final column of table 6. Specifically, while the coefficients on the selection terms remain significant, the magnitude is far less. The coefficients suggest that borrowing from the discount window depressed a wide range of rates, but the most dramatic effect was on the highest rates paid.

Across all specifications, the adjusted R-squared statistics suggest a reasonable amount of variation is explained by these variables. In particular, the specifications in columns 3–5 center around explaining roughly 25 percent of overall variation, consistent with meaningful impact of these factors on trading in the federal funds market.

6. Robustness and Diagnostics

Two-stage estimation procedures with selection can be plagued by a number of weaknesses. This section discusses potential weaknesses of our results related to weak instruments, failure to satisfy overidentifying restrictions, or overly restrictive selection parameters. The section also presents some results from the later crisis period to gauge how federal funds rates were related to discount window borrowing as the crisis wore on.

6.1 First-Stage Results

The selection and endogeneity first-stage results are shown in columns 1 and 2 of table 7. Column 1 displays the results of estimating the first-stage equation described in equation (10), which is the probability of borrowing at the discount window. The two instrumental variables are daylight overdrafts and TAF borrowing; there are also fixed-effect terms for each of these variables. In addition, all exogenous variables from the second stage are also included in the first-stage specification. Looking at the individual coefficients, daylight overdrafts do not appear to be statistically significantly correlated with discount window borrowing. At the same time, TAF borrowing is positively correlated with the probability of borrowing from the discount window. This latter result suggests some complementarity between funding sources during the early stages of the crisis. The result is also consistent with our interpretation of TAF borrowing as an instrument for the probability of borrowing from the discount window, but perhaps uncorrelated with unexpected account shortfalls. For the exogenous factors, the number of days in the market is positively correlated with the probability of borrowing at the discount window, while total assets, required reserve balances, and holdings of reserve balances are negatively correlated. Column 2 presents results from the first-stage specification used to

Table 7. First-Stage Results and Robustness Checks

	First Stage		Cubic Spline		
	Pr(Borrow) (1)	Amount Borrowed (2)	Probit (3)	Logit (4)	Hazard Ratio (5)
Peak Daylight Overdrafts	0.00264	−0.0338			
	(0.00929)	(0.0335)			
TAF Borrowing	0.00584*	0.159***			
	(0.00291)	(0.0215)			
Primary Credit			−2.432**	−2.381**	−2.303**
			(0.861)	(0.845)	(0.841)
Number of Days in Market	0.0286*	0.0435	0.219	0.166	0.191
	(0.0113)	(0.0523)	(0.589)	(0.588)	(0.580)
Amount Borrowed	0.0169*	−0.0524	0.231	0.248	0.268
	(0.00824)	(0.0345)	(0.441)	(0.444)	(0.436)
Assets	−0.000721*	−0.00763***	−0.0102	−0.0119	−0.00731
	(0.000280)	(0.00197)	(0.0188)	(0.0197)	(0.0180)
Required Reserves	−0.634***	0.589	15.28	15.34	15.39
	(0.162)	(0.721)	(8.175)	(8.245)	(8.126)
Total Reserve Balances	−0.746***	0.0579	−12.57	−12.21	−12.27
	(0.199)	(0.630)	(11.96)	(11.96)	(12.28)
CDS Index		−0.446	−3.360	−3.607	−4.891
		(0.308)	(2.662)	(2.679)	(2.774)
Repo-Target Spread		0.00419*	−0.0404	−0.0415	−0.0506*
		(0.00209)	(0.0234)	(0.0233)	(0.0251)
Selection					
100 bp Regime		0.434*	3.847	4.806	79.12*
		(0.171)	(3.505)	(4.440)	(34.65)
50 bp Regime		−0.397	−13.75	−16.86	−241.8
		(0.229)	(7.440)	(9.789)	(205.2)
25 bp Regime		−0.908**	−24.40***	−32.20***	−9.640***
		(0.324)	(7.048)	(8.885)	(2.775)
Hazard Spline					
100 bp Regime					−76.55*
					(33.56)
50 bp Regime					234.9
					(204.5)
25 bp Regime					−11.66
					(71.85)
N	124,403	642	642	642	642
Number of Banks	1,238	129	129	129	129
R^2	0.100	0.292	0.289	0.290	0.298
F-statistic		29.81			
Hansen J Statistic		3.617			
P-value		0.06			
Endogeneity χ^2 Statistic		5.67			
P-value		0.02			

Notes: Specifications include Mundlak-Chamberlain fixed effects and indicators for discount window regime.
Column 1 presents estimated coefficients for the selection equation $Pr(s_{it} = 1|z_i) = \Phi(z_{it}\delta_t^a + \bar{z}_i\xi_t^a)$.
Column 2 presents estimated coefficients for the first-stage equation $primary_{it} = c_i + z_{it1}\alpha + x_{it}\beta_t + \gamma pcspread_t + \bar{z}_i\eta + \rho_{spread}\hat{\lambda}_{it} + \epsilon_{it1}$.
Columns 3–6 present estimated coefficients that control for selection and endogeneity: $f\ highdev_{it} = c_i + \alpha\widehat{primary}_{it} + x_{it}\beta_t + \gamma pcspread_t + \bar{z}_i\eta + \rho_{spread}\hat{\lambda}_{it} + \epsilon_{it1}$.
Robust standard errors are in parentheses. *, **, and *** denote significance at the 5 percent, 1 percent, and 0.1 percent level, respectively.

instrument the amount of primary credit. Coefficients for both the probability of borrowing and the amount borrowed are similar.

Despite the individual insignificance of the daylight overdraft result, a battery of first-stage diagnostics suggest that the TAF and daylight overdraft instruments are both sufficiently strong and satisfy overidentifying restrictions. The statistics for these are presented in the bottom lines of the table in column 2. The F-statistics are significantly higher than the critical values suggested by Stock and Yogo (2005). In addition, the Hansen J-statistic does not reject the hypothesis that overidentifying restrictions are satisfied.

6.2 Other Functional Forms

An issue that arises with selection estimators is the degree to which results depend on the form of the control function. This issue was raised by Newey (2009) and addressed by Semykina and Wooldridge (2010) within the context of their model. Columns 3–6 report results from specifications using a range of control functions. These include three cubic spline estimators: probit, logit, and a hazard function, as in Semykina and Wooldridge (2010).

The results in the rows marked “Selection” and “Hazard Spline” imply similar selection and endogeneity results to the baseline. For the probit and logit spline specifications, banks that opt to use the discount window pay a few basis points less for their highest-rate trades than banks that remain in the market. In addition, all other coefficients on included variables in the specification are qualitatively and quantitatively similar to those presented in the baseline. The hazard rate model does suggest slightly different magnitudes of selection than the probit and logit models spline models. However, the results are still qualitatively similar and there may be some functional form effects that should be accounted for more generally.

6.3 Later Crisis Period

The focus of this paper is on the early stages of the crisis. For robustness, it is important to explore how stigma and sample selection shifted after the collapse of Lehman Brothers in the fall of 2008. To do so, the baseline specification is evaluated on a sample with data from September 2008 to April 2010, which marked the conclusion of the TAF program.

There are some caveats with this exercise. Not only did the collapse of Lehman Brothers signal the acute stage of the financial crisis, but it was also met with a substantial change in the Federal Reserve's monetary policy operating framework. Specifically, as a result of a wide range of lending programs and the first rounds of quantitative easing, reserve balances ballooned, from an average level of roughly $25 billion before September 2008 to 100 times that level afterwards. In turn, the level of daylight overdrafts cratered, as banks generally had substantial funds in their reserve accounts to cover payments without incurring overdrafts.

Against that backdrop, table 8 reports the results of estimating equation (13). Comparing the baseline panel specifications shown in column 1 of table 8 with the baseline specification in column 1 of table 6, there are some key differences in the interplay between borrowing and rates paid. Specifically, the coefficient on primary credit borrowing is positive, not insignificant or negative. Still, TAF borrowing is also negatively correlated with higher rates paid, and the coefficient on peak daylight overdrafts is positive and significant. Taken together, these coefficients suggest that banks that borrowed from the TAF paid lower rates and banks with overdrafts paid higher rates than those that did not, similar to what our early crisis hypothesis would suggest.

That said, evidence in columns 2 and 3 implies that high-rate trading may have become somewhat decoupled from primary credit as the crisis continued and the Federal Reserve's balance sheet ballooned. Column 2 suggests that there was no selection evident in high rates paid in the federal funds market. As such, the results in columns 2 and 3 are not remarkably different. Moreover, as shown in the bottom lines of column 3, first-stage tests suggest that primary credit was no longer endogenous to rates paid in the federal funds market. Importantly, federal funds market participation declined dramatically as reserve balances climbed and counterparty credit risk intensified. Moreover, only the best credits remained in the market. While lenders were wiling to extend credit at higher rates to borrowers during the early stages of the crisis, after Lehman, overnight unsecured credit became scarce and rates became less dispersed, with many banks turning to the TAF for funding.

Table 8. Later Crisis Period

	Panel (1)	Selection (2)	Corrected (3)
Primary Credit	−0.000562	−0.334	0.193
	(0.00200)	(0.566)	(0.618)
Target = 2 Percent	0.0294**		
	(0.00977)		
Target = 1.5 Percent	0.0116***		
	(0.00310)		
Target = 1 Percent	0.00676*		
	(0.00319)		
Peak Daylight Overdrafts	0.0142***		
	(0.00170)		
TAF Borrowing	−0.000938***		
	(0.000221)		
Target = 2 Percent	0.0110**		
	(0.00344)		
Target = 1.5 Percent	0.00174		
	(0.00212)		
Target = 1 Percent	−0.000137		
	(0.000609)		
Number of Days in Market	0.00754*	0.178	−0.0541
	(0.00305)	(0.326)	(0.150)
Amount Borrowed	0.000788	0.00151	0.133
	(0.00102)	(0.192)	(0.147)
Assets	0.000224	−0.206	−0.0697
	(0.000134)	(0.377)	(0.270)
Required Reserves	−0.0421	−39.13	13.04
	(0.0236)	(64.19)	(43.63)
Reserve Balances	−0.00346***	−0.0515	−0.111
	(0.000974)	(0.0587)	(0.197)
CDS Index	0.0123**	−0.328	0.126
	(0.00444)	(0.565)	(0.389)
Repo-Target Spread	−0.244***	−0.653	−0.138
	(0.0258)	(0.722)	(0.235)
Selection			
Target = 2 Percent		0.288	0.0713
		(0.728)	(0.463)
Target = 1.5 Percent		0.365	0.174
		(0.695)	(0.908)
Target = 1 Percent		0.0473	0.0218
		(0.362)	(0.364)
Target = 25 bp		0.149	−0.0214
		(0.273)	(0.301)
N	9,301	503	503
Number of Banks	395	121	121
Adj. R-sq.	0.277	0.046	
Endogeneity χ^2 Statistic			3.233
P-value			0.07

Notes: Dependent variable is the deviation of the average observed high rate paid for federal funds from the effective rate. Specifications include Mundlak-Chamberlain fixed effects.
Column 1 presents estimated coefficients for the baseline panel regression:
$f\,fhighdev_{it} = c_i + \alpha_t^1 primary_{it} + \alpha^2 daylight_{it} + \alpha_t^3 TAF_{it} + \gamma days_{it} + x_{it}\delta + c_i + q_{pc}\zeta + \epsilon_{it}$.
Column 2 presents estimated coefficients for the selection test:
$f\,fhighdev_{it} = c_i + \alpha primary_{it} + x_{it}\beta_t + \gamma pcspread_t + \rho_{spread}\hat{\lambda}_{it} + \epsilon_{it1}$.
Columns 3–6 present estimated coefficients that control for selection and endogeneity:
$f\,highdev_{it} = c_i + \alpha\widehat{primary}_{it} + x_{it}\beta_t + \gamma pcspread_t + \bar{z}_i\eta + \rho_{spread}\hat{\lambda}_{it} + \epsilon_{it1}$.
Cluster-robust standard errors are in parentheses. *, **, and *** denote significance at the 5 percent, 1 percent, and 0.1 percent level, respectively.

7. Conclusion

This paper presents a theoretical framework and empirical results that can explain the interaction of the Fed's liquidity provision and the federal funds market in the first stages of the recent financial crisis. If aversion to obtaining funds from the discount window differs across banks, a lower spread of the primary credit rate over the target rate can help lenders price discriminate in a way that is impossible with a wider spread. And, although this price discrimination may lead to higher market rates, overall funding costs may still be lower as a result of the narrowing of the spread between the primary credit rate and the target rate.

Furthermore, the lowering of the primary credit rate may have supported trading in the federal funds market to continue despite the financial crisis, as lenders were more able to price discriminate. Three salient empirical facts help to show this point: (i) as the spread between the primary credit rate and the target rate narrowed, the number of primary credit borrowers and the level of primary credit increased, while the number of participants in the federal funds market decreased; (ii) on an aggregate level, trading above the primary credit rate is correlated with various measures of banking industry stress; and (iii) on an institution level, there is evidence of selection in the federal funds market—as the spread between the primary credit rate and the target rate narrowed, banks that did not go to the discount window paid significantly higher rates in the federal funds market.

By and large, most of the time federal funds were brokered below the primary credit rate and occasions where funds were brokered above the primary credit rate were infrequent. But it is still instructive to study these episodes of above-rate trading to understand the interaction between unsecured interbank markets and central bank liquidity provision in the early days of a financial crisis.

Appendix. Comparative Statics

This appendix reviews some basic comparative statics from the model.

- If q decreases, the equilibrium rate rises.

Taking the total differential of (2) shows

$$\frac{dr^*}{dq} = r^{alt} - \left(r^{dw} + \sum_k p_k \theta_k \right). \tag{A.1}$$

Since $r^{alt} < r^{dw}$ and $sum_k p_k \theta_k \geq 0$ by assumption, this statement is necessarily true. This result intuitively makes sense: as the bargaining power of the buyer falls, the equilibrium rate necessarily rises. During the beginning stages of the financial crisis, as banks were increasingly under scrutiny for their safety and soundness, one might suspect that their bargaining power might fall a bit.

- If stigma increases, the equilibrium rate rises.

There are two ways stigma can increase: either the stigma parameter θ_k or the share of banks with a high stigma p_k can increase. For the first case, taking the total differential of (2) shows

$$\frac{dr^*}{d\theta_k} = p_k, \tag{A.2}$$

which is necessarily positive.

For the second, we assume that an increase in p_k causes the shares p_j to decrease equally, $j \neq k$. Thus, if dp_k is the change in type $k's$ share, we have $dp_j = -\frac{dp_k}{n-1}$ for all $j \neq k$. As a result, we see that

$$\frac{dr^*}{dp_k} = \theta_k - \frac{1}{n-1} \sum_{j \neq k} \theta_j \Longrightarrow \tag{A.3}$$

$$\frac{dr^*}{dp_k} = \theta_k - \bar{\theta}_{-k}, \tag{A.4}$$

where n is the number of banks in the federal funds market and $\bar{\theta}_{-k}$ is the average stigma of institutions not of type k. This simply implies that the effective rate goes up if banks with above-average stigma increase in share.

- If the discount window rate decreases, the equilibrium rate can rise or fall.

This is a result of the direct effect of the discount window rate on pricing in the federal funds market, as well as the indirect effect on participation. Taking the total differential of (2) shows

$$\frac{dr^*}{dr^{dw}} = (1-q) + (1-q)\left(\frac{d}{dr^{dw}}\left(\sum_k p_k\theta_k\right)\right). \tag{A.5}$$

The direct effect of a decrease in the discount window rate on the equilibrium rate is a fall in the rate, as shown by the first term, $(1 - q)$. However, note that $\frac{d}{dr^{dw}}\left(\sum_k p_k\theta_k\right)$ increases with a decrease in the discount window rate—that is, while there are fewer buyers if the primary credit rate falls, those with lower stigma drop out, because of selection. Consequently, the average level of stigma increases. If this effect is sufficiently positive, then overall, the equilibrium rate will rise. Empirically, we will show that there are instances where this selection effect dominates and the lowering of the primary credit rate resulted in higher federal funds rates.

Of course, the same factors that could lead policymakers to narrow the spread between the primary credit rate and the target rate could also cause some of the parameters of this equilibrium condition to shift. For example, these factors could lead to an increase in the average level of stigma. As a result, the term $\frac{d}{dr^{dw}}\left(\sum_k p_k\theta_k\right)$ could be boosted, and the equilibrium rate could rise coincident with the primary credit rate.

Another item to note is how this affect adjusts with a change in the bargaining power parameter, q. If conditions are such that the discount window rate would be lowered, this could also be reflected as a fall in the bargaining parameter, q. This could serve to boost the direct effect of lowering the discount window rate. However, the same factors that could lead to a reduction in the discount rate could also lead to an increase in the average level of stigma, and therefore, boost $\frac{d}{dr^{dw}}\left(\sum_k p_k\theta_k\right)$, and cause the equilibrium rate to increase more with the discount rate.

References

Afonso, G., A. Kovner, and A. Schoar. 2011. "Stressed, Not Frozen: The Federal Funds Market in the Financial Crisis." *Journal of Finance* 66 (4): 1109–39.

Afonso, G., and R. Lagos. 2015. “Trade Dynamics in the Market for Federal Funds.” *Econometrica* 83 (1): 263–313.

Armantier, O., and A. Copeland. 2015. “Challenges in Identifying Interbank Loans.” *Economic Policy Review* (Federal Reserve Bank of New York) 21 (1, October): 1–17.

Armantier, O., E. Ghysels, A. Sarkar, and J. Shrader. 2015. “Discount Window Stigma during the 2007-2008 Financial Crisis.” *Journal of Financial Economics* 118 (2): 317–35.

Artuc, E., and S. Demiralp. 2010. “Discount Window Borrowing after 2003: The Explicit Reduction in Implicit Costs.” *Journal of Banking & Finance* 34 (4): 825–33.

Ashcraft, A. B., M. L. Bech, and W. S. Frame. 2010. “The Federal Home Loan Bank System: The Lender of Next-to-Last Resort?” *Journal of Money, Credit and Banking* 42 (4): 551–83.

Bartolini, L., S. Hilton, and J. J. McAndrews. 2010. “Settlement Delays in the Money Market.” *Journal of Banking and Finance* 34 (5): 934–45.

Bech, M. L., and E. Klee. 2011. “The Mechanics of a Graceful Exit: Interest on Reserves and Segmentation in the Federal Funds Market.” *Journal of Monetary Economics* 58 (5): 415–31.

Bernanke, B. S. 2008. “Liquidity Provision by the Federal Reserve.” Speech at the Federal Reserve Bank of Atlanta Financial Markets Conference (via videoconference), Sea Island, Georgia, May 29.

Calomiris, C. 1994. “Is the Discount Window Necessary? A Penn-Central Perspective.” *Review* (Federal Reserve Bank of St. Louis) 76 (3): 31–55.

Clouse, J., and J. P. Dow. 1999. “Fixed Costs and the Behavior of the Federal Funds Rate.” *Journal of Banking and Finance* 23 (7): 1015–29.

Demiralp, S., B. Preslopsky, and W. Whitesell. 2006. “Overnight Interbank Loan Markets.” *Journal of Economics and Business* 58 (1): 67–83.

Ennis, H., and J. A. Weinberg. 2013. “Over-the-Counter Loans, Adverse Selection, and Stigma in the Interbank Market.” *Review of Economic Dynamics* 16 (4): 601–16.

Friedman, M., and A. J. Schwartz. 1963. *A Monetary History of the United States*. Princeton, NJ: Princeton University Press.

Furfine, C. 1999. "The Microstructure of the Federal Funds Market." *Financial Markets, Institutions, and Instruments* 8 (5): 24–44.

———. 2003. "Standing Facilities and Interbank Borrowing: Evidence from the Federal Reserve's New Discount Window." *International Finance* 6 (3): 329–47.

Goodfriend, M. 1983. "Discount Window Borrowing, Monetary Policy, and the Post-October 6, 1979 Federal Reserve Operating Procedure." *Journal of Monetary Economics* 12 (3, September): 343–56.

Livne, Z. A. 1988. "The Bargaining Problem with an Uncertain Conflict Outcome." *Mathematical Social Sciences* 15 (3): 287–302.

Madigan, B., and W. Nelson. 2002. "Proposed Revision to the Federal Reserve's Discount Window Lending Programs." *Federal Reserve Bulletin* (July): 312–19.

Newey, W. K. 2009. "Two-step Series Estimation of Sample Selection Models." *Econometrics Journal* 12 (s1): S217–S229.

Peristiani, S. 1998. "The Growing Reluctance to Borrow at the Discount Window: An Empirical Investigation." *Review of Economics and Statistics* 80 (4): 611–20.

Semykina, A., and J. M. Wooldridge. 2010. "Estimating Panel Data Models in the Presence of Endogeneity and Selection." *Journal of Econometrics* 157 (2): 375–80.

Sidel, R., G. Ip, and V. Bauerlein. 2007. "Banks Step Up to Fed's Window." *The Wall Street Journal*, August 23, p. A3.

Stigum, M., and A. Credenzi. 2007. *Stigum's Money Market.* 4th Edition. New York: The McGraw-Hill Companies.

Stock, J., and M. Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models,* ed. D. W. K. Andrews, 80–108. New York: Cambridge University Press.

International Inflation Spillovers: The Role of Different Shocks*

Gregor Bäurle,[a] Matthias Gubler,[a] and Diego R. Känzig[b]
[a]Swiss National Bank
[b]London Business School

How do international price fluctuations spill over to country-specific inflation? We show that accounting for the drivers of international inflation and their effects on overall economic conditions is crucial to getting a more thorough view of spillover effects. We find substantial heterogeneity in the magnitude of spillovers, depending on the shocks driving inflation abroad. While all identified shocks are inflationary, their effects on activity, interest rates, and exchange rates differ. Disaggregated price responses suggest that these general equilibrium effects are important. We show this by looking at spillovers to Switzerland using a structural dynamic factor model relating disaggregated prices to key macroeconomic factors.

JEL Codes: C11, C32, E31, E52, F62.

*We thank Juan Antolin-Diaz, Carlos Lenz, Ana Mitreska, Tsvetelina Nenova, Juan Rubio-Ramirez, Paolo Surico, the editor, Pierpaolo Benigno, and two anonymous referees as well as participants at the SNB Brown Bag Seminar, the CFE conference 2015 in London, the SSES meeting 2016 in Lugano, the EEA conference 2016 in Geneva, and the NBP-SNB joint seminar 2017 in Zurich for useful comments and suggestions. We would also like to thank Santiago Ernesto Alvarez-Blaser for excellent research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect those of the SNB. A previous version of this paper circulated under the title "Does Independent Monetary Policy Shield against International Spillovers to Inflation? An Empirical Study with Swiss Data." Author contact: Bäurle: Swiss National Bank, Regional Economic Relations, P.O. Box, CH-8022 Zurich, Switzerland, e-mail: gregor.baeurle@snb.ch. Gubler: Swiss National Bank, Economic Data Science, P.O. Box, CH-8022 Zurich, Switzerland, e-mail: matthias.gubler@snb.ch. Känzig: London Business School, London NW1 4SA, United Kingdom, e-mail: dkaenzig@london.edu.

1. Introduction

In recent decades, the world has moved closer together along many dimensions. Increased economic integration goes hand-in-hand with potentially major, international spillover effects. A deep understanding of spillovers is thus key to ensuring optimal policy decisions, particularly in open economies with strong international ties. From a monetary policy perspective, international spillovers to consumer prices are of particular interest. However, the empirical evidence on the impact of spillovers to country-specific inflation is ambiguous. Some authors find that "inflation is largely a global phenomenon" (Ciccarelli and Mojon 2010), while others document that the importance of domestic factors for country-specific inflation has not diminished (Rieth 2015).

Recent, influential studies have proposed a new focus on pass-through effects: accounting for the underlying shocks and the overall changes in economic conditions, i.e., general equilibrium effects. Comunale and Kunovac (2017) and Forbes, Hjortsoe, and Nenova (2018) examine the exchange rate pass-through, while Bobeica, Ciccarelli, and Vansteenkiste (2019) look at the pass-through of labor costs to prices. In this paper, we take up the idea of shock-dependent pass-through by examining how various foreign shocks, all of which have in common that they push up international inflation, affect country-specific inflation and its subcomponents. To do so, we use a structural dynamic factor model (SDFM) for Switzerland—a small, open economy that is particularly suited for studying international spillover effects. This model relates a large set of disaggregated Swiss consumer prices to key international and domestic macroeconomic factors, and allows for multiple transmission channels.

Accounting for the underlying shocks of international inflationary pressures may help to reconcile the ambiguous empirical evidence. Although the identified shocks all push up international inflation, their effects on other foreign and domestic macrovariables, such as real activity, interest rates, or exchange rates, may differ. As a result, this could lead to different spillover effects. To the best of our knowledge, we are the first to examine whether there is empirical evidence for shock dependence of international inflation spillovers. Moreover, by comparing the relative price changes of goods and services with different degrees of tradability, we can get a sense of the importance of indirect, general equilibrium effects relative to the direct, mechanical pass-through of international to country-specific inflation and of how this may vary with the shocks.

Our framework also allows us to study the joint responses of the exchange rate and the interest rate differential, shedding some light on how monetary policy can shape the response to the respective shocks. Traditionally, flexible exchange rates in tandem with independent monetary policy are thought to be effective in cushioning the effect of international shocks on the domestic economy (see, e.g., Woodford 2007). However, if country-specific inflation is mainly driven by global factors, this could indicate that "domestic inflation rates may (at least partly) escape the control of the national central bank" (Monacelli and Sala 2009).

1.1 Preview of Results

We find that foreign shocks explain up to 50 percent of Swiss price variations, while common domestic shocks account for approximately 20 percent (the remaining part being due to item-specific shocks). To a substantial degree, domestic inflation is thus driven by foreign factors. However, this does not necessarily imply that Swiss monetary policy has not been able to have an impact on international spillover effects to domestic inflation. In fact, we show that spillover effects on Swiss prices depend on the nature of the underlying shocks, because their transmission varies, among other factors, with the distinct foreign and domestic monetary policy responses.

Following an increase in inflation abroad due to a positive demand shock, foreign monetary policy counteracts the business cycle upturn strongly, while the Swiss monetary policy reaction turns out to be less restrictive. Consistent with the change in the relative monetary policy stance, the Swiss franc depreciates and inflation picks up, even somewhat more than abroad. By contrast, in response to an increase in foreign inflation due to an expansionary monetary policy shock, monetary policy becomes relatively tighter in Switzerland and the exchange rate appreciates—mitigating spillover effects to Swiss inflation. Finally, a cost-push shock driving up inflation (and decreasing real activity) abroad has no significant effect on the relative monetary policy stance. The effects on the exchange rate turn out to be negligible, and the increase in Swiss inflation is comparable to that of inflation abroad. These results indicate that spillover effects need to be analyzed in a framework allowing for different transmission channels: an increase in inflation abroad may affect inflation in an open economy differently, depending on the source of the foreign shock and thus on movements in other factors such as interest and exchange rates.

Our analysis of the different items of the Swiss consumer price index (CPI) points to substantial heterogeneity in the transmission of foreign inflationary shocks. It turns out that energy prices play a crucial role. The impact of foreign inflationary shocks on the Swiss CPI is markedly lower, and the transmission appears to be slower when energy prices are excluded. Furthermore, there is some heterogeneity in the transmission to the prices of imported goods, domestic goods, and services, which are likely related to differences in tradability and exchange rate sensitivity. This suggests that while a certain part of spillovers is likely to be mechanical, general equilibrium effects are important as well. While we find short- to medium-run changes in relative prices in response to the foreign shocks, we do not find significant effects on relative prices in the long run, in line with previous findings in the literature (Boivin, Giannoni, and Mihov 2009; Mumtaz and Surico 2009). This further underlines the importance of the relative stance of monetary policy through its effect on the exchange rate.

Our results turn out to be robust along a number of dimensions, including the model specification, the choice of the prior, and the sample period. In the baseline model, we use the euro area—Switzerland's largest trading partner by far—as the foreign block. However, the results based on a global foreign block consisting of export-weighted indicators of Switzerland's major trading partners are very similar, suggesting that our findings do not uniquely pertain to spillovers from the euro area, but to international spillovers to Switzerland more generally.

1.2 Related Literature

This paper is related to at least three different strands of the literature. First, it is related to a large body of literature studying the co-movement of international inflation rates (Ciccarelli and Mojon 2010, Neely and Rapach 2011, Mumtaz and Surico 2012). This literature finds that country-specific inflation rates are largely a global phenomenon, i.e., individual countries tend to inherit global inflationary pressures. In contrast, various other recent studies focusing on the impact of specific global factors, such as commodity prices or global business cycles, on domestic inflation dynamics find ambiguous empirical results (see Rieth 2015 for an overview).

Second, our analysis is also related to the extensive literature on the international transmission of external shocks (see Eichenbaum and Evans 1993; Kim 2001; Canova 2005; Maćkowiak 2007; Aastveit, Bjørnland, and Thorsrud 2016; Georgiadis 2016; Dedola, Rivolta, and Stracca 2017; Potjagailo 2017, among others). Our paper connects these two strands of the literature by analyzing the role of international factors for domestic inflation from a more structural perspective, building on an approach that has recently gained a lot of attention in the literature on the exchange rate pass-through (Shambaugh 2008; Comunale and Kunovac 2017; Forbes, Hjortsoe, and Nenova 2017, 2018). In contrast to these studies, our focus is not on the exchange rate pass-through per se, but on international inflation spillovers. In our analysis, the exchange rate is one of several *endogenous* variables, although an important one, through which the effects of the foreign inflationary shocks transmit to Switzerland. As such, our results are not intended to be directly comparable to studies that quantify how exogenous exchange rate fluctuations affect prices (see, e.g., Stulz 2007 and Fleer, Rudolf, and Zurlinden 2016 for evidence on Switzerland).

Third, our approach of incorporating a set of disaggregated prices allows us to compare our results with a relatively new and growing literature that looks at how global factors affect specific subcomponents of inflation, and how important they are for headline inflation (see, e.g., Mumtaz and Surico 2009, Halka and Szafranek 2016, Parker 2016, Altansukh et al. 2017).

Moreover, we are aware of two papers close to the topic of our article. Mumtaz and Surico (2009) use a similar framework to study how global shocks transmit to the U.K. economy. However, while they look at spillover effects more generally, our focus is on inflation dynamics. Furthermore, we provide new empirical evidence on how the strength of the spillovers varies with the relative monetary policy stance and the exchange rate response. Halka and Kotlowski (2017) also study the effects of global shocks on inflation of three small open economies (Czech Republic, Poland, and Sweden) using aggregate and disaggregated price data. However, they use a different modeling approach and focus on contemporaneous effects, while we examine the dynamic effects of shocks driving international inflation. This can be important because the transmission of these shocks might take some time, as can be seen from our results. Moreover, we provide new evidence on the quantitative importance of foreign inflationary pressures using variance decompositions, and we show that the impact on the distribution of prices depends on the nature of the underlying inflationary shock.

1.3 Structure of the Paper

The remainder of this paper is organized as follows. Section 2 covers our econometric approach, including details about the modeling framework, the model specification, the data, and the estimation and identification strategy. In section 3, we present our results and discuss their implications. Section 4 concludes the paper.

2. Econometric Approach

To study the potential spillover effects of foreign inflationary pressures on the Swiss economy and in particular on Swiss prices, we set up a structural dynamic factor model for the Swiss economy. The model relates a large set of disaggregated price data to the key domestic and foreign macroeconomic factors. Building on the framework proposed by Bäurle and Steiner (2015), it takes into account the characteristics of a small open economy. The structural shocks are identified using two different types of restrictions. First, we exploit that economic conditions in Switzerland, a small open economy, are unlikely to affect global economic conditions. This allows us to separate foreign from domestic shocks. Second, we use sign restrictions motivated by economic theory to disentangle different types of foreign shocks.

2.1 Modeling Framework

A dynamic factor model is a framework relating a large panel of economic indicators to a number of observed and unobserved common factors. The premise behind this type of model is that the economy can be characterized by a limited number of factors that drive the co-movements of the indicators in the panel. Formally, the model consists of two different equations: an observation equation and a state equation. The observation equation relates the panel of economic indicators X_t^S to the common factors f_t that drive the economy:

$$X_t^S = \lambda(L) f_t + v_t, \tag{1}$$

where $\lambda(L) = \lambda_0 + \lambda_1 L + \lambda_2 L^2 + \cdots + \lambda_q L^q$ are the factor loadings, L is the lag operator, and v_t is a vector of item-specific components. Thus, the indicators X_t^S are allowed to load on the factors both contemporaneously and on their lags.[1] Following Boivin and Giannoni (2006), we allow v_t to be autocorrelated of order one by specifying $v_t = \psi v_{t-1} + \xi_t$ with $\xi_t \sim N(0, R)$. For our specific application, X_t^S comprises a large number of disaggregated data on Swiss consumer prices. To make our model suitable for a small open economy, we partition the common factors f_t into two blocks: a foreign and a domestic block. The domestic block is further partitioned into a block of unobserved factors and a block of observed factors. Hence, the common factors can be written as $f_t = (f_t^{S'}, \; X_t^{M'}, \; X_t^{M*'})'$, where f_t^S are the domestic unobserved common factors, X_t^M are the domestic observed common factors, and X_t^{M*} are foreign observed common factors. The joint dynamics of these factors are described by the following state equation:

$$\phi(L) f_t = Q\varepsilon_t, \tag{2}$$

where $\phi(L) = I - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ is a lag polynomial, ε_t is a vector of common structural shocks with the same dimension as f_t, and Q is the structural impact matrix mapping the shocks to the

[1] Note that it is possible to rewrite the model in "static form," by including lagged factors into the state vector. This reveals that there is a close connection between the dimension of f_t and the number of lags in the observation equation q. However, increasing q may be a more parsimonious way than increasing the dimension of f_t to relate observed variables to lagged factors as the former implies restrictions on the state equation in the static form. Ultimately, whether these restrictions are supported by the data is an empirical question. We investigate the robustness of our results to the choice of the number of factors and q in section 3.4.

common factors. This is essentially a vector autoregression (VAR) in the factors. The shocks ε_t are assumed to be Gaussian white noise, i.e., $\varepsilon_t \sim N(0, I)$. Moreover, the common shocks ε_t and the idiosyncratic shocks ξ_t, which we call item-specific shocks, are postulated to be uncorrelated. The vector of common shocks ε_t can be partitioned into vectors of foreign shocks ε_t^{M*} and domestic shocks ε_t^{M}, whose dimensions correspond to X_t^{M*} and X_t^{M}, respectively. The small open economy assumption is then implemented by modeling the foreign block of the model as exogenous to the Swiss economy. To this end, we assume that foreign variables do not react to domestic shocks at all lags by restricting $\phi(L)$ and the covariance matrix Q appropriately. More precisely, we restrict the block of $\phi(L)$ that relates X_t^{M*} to the lags of X_t^{M} and the elements of Q that relate X_t^{M*} to the domestic shocks ε_t^{M} to zero.

2.2 Discussion of Modeling Choice

The dynamic factor model described in the previous section allows us to model a large set of time series jointly. As compared with a VAR in all variables, the factor structure reduces the number of parameters substantially. Indeed, the number of parameters grows linearly with the number of observed series (for a given number of factors), while the number of parameters in a VAR increases quadratically with the number of series. As an alternative to the factor structure, Bayesian versions of the VAR have been proposed to deal with this curse of dimensionality. We do not pursue this route mainly for two related reasons. First, our setting allows for a different treatment of our macro series and individual price series. The fact that aggregate series do not react to individual price series, but only to factors containing aggregate price information, is a reasonable assumption in our view. In a VAR, each individual price series reacts to any other series including the macro series and vice versa. Second, and related to this, the distinction between macro series (including the factors) and price series helps to identify macroeconomic shocks. In addition to the "standard" identifying assumption needed within a VAR (see also discussion in section 2.5), we impose that the idiosyncratic components are orthogonal to the macroeconomic shocks, making identification of aggregate shocks even possible.

A note on the terminology is due at this point. Our model may also be described as a factor-augmented VAR (FAVAR). A FAVAR is, formally, a special case of a SDFM with some factors perfectly observed (section 5.2 in Stock and Watson 2016 and Bernanke, Boivin, and Eliasz 2005). Indeed, as we assume that some factors, X_t^M and X_t^{M*}, are observed, our implementation of the dynamic factor model fits into this category. We prefer to use the general term "dynamic factor model," however, because our motivation is not primarily to "augment" the VAR with information from price series, but to model the joint dynamics of macrovariables and the price series.

2.3 Specification and Data

Our baseline specification includes six observed common factors. As discussed above, these factors are grouped into two blocks: a foreign and a domestic one. The foreign block contains measures for output, the short-term interest rate, and consumer prices. The domestic block consists of the same type of measures except consumer prices because they are implicitly contained in the disaggregated price data. To link the domestic to the foreign economy, we also include the nominal exchange rate. In this modeling framework it is possible, on the one hand, to study spillovers to disaggregated Swiss consumer prices and, on the other hand, to keep the model as parsimonious as possible. Furthermore, the inclusion of the abovementioned observed common factors allows us to identify standard macroeconomic shocks.

As discussed in Bäurle and Steiner (2015), the selection of the remaining model dimensions is not trivial. Parsimony in mind, we start with dimensions in the lower range of what is chosen in the literature in our baseline specification, setting the number of unobserved factors to one and the lag order in the state and the observation equation to $p = 2$ and $q = 1$, respectively.[2] Later, we will check the robustness of our results with respect to this choice.

[2]As an example, Mumtaz and Surico (2009) set $p = 4$ and (implicitly) $q = 0$. They set the number of unobserved factors in the domestic economy to four. However, as we add three observed factors, the total number of domestic factors in our model is four as well.

Because the euro area is Switzerland's most important trading partner, we choose it as the foreign block of the model. For the measures of output, short-term interest rates, and consumer prices, we use euro-area real gross domestic product (GDP), the three-month euro-area interbank offered rated (3M EURIBOR), and euro-area CPI, respectively. For Switzerland, we use Swiss real GDP and the three-month London interbank offered rate (3M LIBOR). Finally, the EURCHF is selected as the relevant nominal exchange rate. The exchange rate is quoted inversely; hence, a positive exchange rate change implies an appreciation of the Swiss franc. For the disaggregated price data, we rely on a panel of 148 Swiss CPI items.[3] The frequency of the data is quarterly, and the sample spans the period from 1992:Q1 to 2011:Q2. We choose this particular sample period because it was characterized by a relatively stable monetary policy regime and flexible exchange rates, which is important, as our model does not allow for time variation in the model parameters. All variables enter the model as quarter-on-quarter (qoq) growth rates except for the interest rates, which enter in levels. Following the literature, the series are standardized such that they have zero mean and a variance equal to one. After estimation, the quantitative results are transformed back into the original scale.

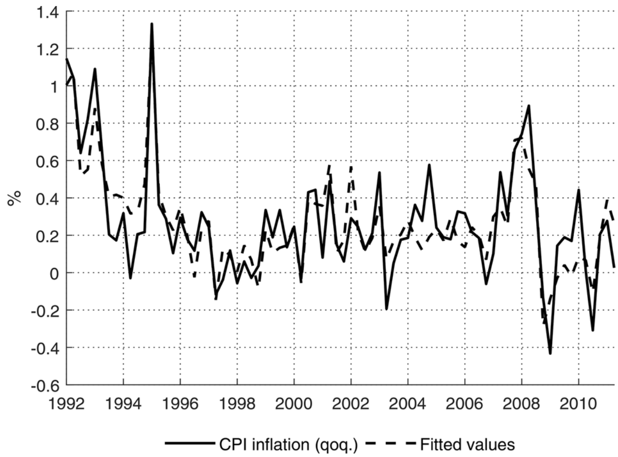
It is important to note that despite not including Swiss CPI inflation explicitly as an observed factor in the state equation, the remaining factors contain almost all consumer price information. Figure 1 shows Swiss CPI inflation (qoq) together with the fitted values of the following ordinary least squares (OLS) regression:

$$\pi_t = \beta f_t + u_t. \tag{3}$$

One sees that the fit based on the seven factors in the baseline model (the six observed factors and the unobserved factor) is excellent, matching basically all peaks and troughs, with an R^2 of slightly above 75 percent. Thus, misspecification as a consequence of excluding inflation explicitly should be minor.

[3]The panel is constructed from item-level price data collected by the Swiss Federal Statistical Office (SFSO). A more detailed description of the disaggregated price data can be found in online appendix A, available with the online version of this paper at http://www.ijcb.org.

Figure 1. Swiss CPI Inflation (qoq) versus Fitted Values



Notes: The figure illustrates Swiss CPI inflation (qoq) and the fitted values of an OLS regression of Swiss CPI inflation on one unobserved factor and six observed factors. The observed factors are euro-area real GDP, the 3M EURIBOR, euro-area CPI, the EURCHF, Swiss real GDP, and the 3M LIBOR. All observed variables enter the model as quarter-on-quarter growth rates except for the interest rates, which enter in levels.

2.4 Estimation

The model is estimated using Bayesian methods. Because it is not possible to derive analytical results for high-dimensional estimation problems such as the one at hand, we have to rely on numerical techniques to approximate the posterior. In particular, we use a Gibbs sampler, iterating over the following two steps (see, e.g., Kim and Nelson 1999). First, for a given (initial) set of model parameters, a realization of the distribution of the factors conditional on this set of parameters is drawn. Given this draw, a new set of parameters can be drawn from the distribution of parameters conditional on the draw of the factors.

The two steps are repeated $J = 100{,}000$ times. From these draws, we discard the first 20,000 to assure that the chain has converged to its ergodic distribution. Geweke's spectral-based measure of relative numerical efficiency (RNE; see, e.g., Geweke 2005) suggests that efficiency loss of the algorithm due to the remaining autocorrelation in these evaluated draws is minimal.[4] The efficiency loss is less than 50 percent for almost all of the parameters, i.e., vis-à-vis a hypothetical independence chain, and we need no more than 50 percent additional draws to achieve the same numerical precision. Moreover, the maximum inverse RNE is 4.6, which is well below the value of 20 that is mentioned in the literature as a critical threshold (see, e.g., Primiceri 2005, Baumeister and Benati 2013, or Carriero, Clark, and Marcellino 2014). Additionally, we use Geweke (1992)'s test to assess the convergence of the algorithm, confirming that posterior means for partitions of the chain do not differ.[5] We also investigate convergence visually by looking at the posterior means based on an expanding number of draws, finding no evidence of changes after less than half of the draws.

Our choices for the prior distributions are the following. The prior for the coefficients in the observation equation is proper. This mitigates the problem that the likelihood is invariant to an invertible rotation of the factors. The problem of rotational indeterminacy in this Bayesian context is discussed in detail in Bäurle (2013).[6] The determination of the coefficients describing the factor dynamics reduces to the estimation of a standard VAR. We implement the restrictions reflecting the exogeneity assumption on foreign factors following Bauwens, Lubrano, and Richard (1999) and Karlsson (2013). Furthermore, we impose stationarity by rejecting the draws that do not satisfy the stationarity condition. It is important to note that the likelihood is only informative about $\Sigma = QQ'$, but not about Q directly. Therefore, we first derive the posterior distribution of Σ and impose certain restrictions based on economic considerations to pin down the distribution of Q in a second step. The strategy for identifying Q depends on the specific application

[4]The spectrum at frequency zero is calculated using a quadratic spectral kernel as described in Neusser (2009).

[5]We follow Geweke (1992) and test whether the parameter means based on the first one-tenth of the draws (after discarding the burn-in sample) are significantly different from the second half of the draws.

[6]Bayesian analysis is always possible in the context of nonidentified models as long as a proper prior on all coefficients is specified; see, e.g., Poirier (1998). Note that rotating the factors does not have an impact on the impulse response functions as long as no restrictions on the responses of the factors to shocks are set.

and is described in the subsequent subsection. As compared to the procedure in Bäurle and Steiner (2015), we implement two changes to the prior distribution. Both changes help us to make the estimation procedure more robust, especially in short samples. First, we assume that a priori, the variances of the parameters in $\lambda(L)$ are decreasing with the squared lag number. Second, we assume a Minnesota-type prior for the parameters in the state equation as described in Karlsson (2013). We set the hyperparameters as follows: in Karlsson (2013)'s notation, we use $\pi_1 = \pi_2 = \pi_3 = 1$ to implement a very loose prior and set the prior mean of the first own lag to zero as we model stationary series. Further details on the estimation method and the implementation can be found in online appendix B.

2.5 Identification

To analyze how foreign inflationary pressures affect the Swiss economy, we identify three different foreign inflationary shocks: a demand shock, a monetary policy (MP) shock, and a cost-push shock, all originating in the euro area. The shocks are identified using two different types of restrictions. First, we exploit that economic conditions in Switzerland are unlikely to have an impact on economic conditions abroad. Thus, domestic shocks are restricted to have no effect on foreign variables as implemented by short-run zero restrictions on Q. In this way, domestic shocks are separated from foreign ones. Note that in combination with the restrictions on $\phi(L)$ described in subsection 2.1, domestic shocks do not influence foreign variables at all lags. Second, we use sign restrictions to disentangle the different types of foreign shocks. Following Uhlig (2005), we restrict the sign of the response of selected elements of X^M_{t+h}, but do not directly impose restrictions on the reaction of X^S_{t+h}. Specifically, we assume that a positive shock to foreign demand leads to an increase in output, prices, and the real interest rate (nominal interest rate minus CPI inflation) in the euro area. In contrast, an expansionary foreign monetary policy shock is assumed to decrease the policy rate and to increase output and prices in the euro area. Finally, we assume that a cost-push shock in the euro area causes output to fall and foreign prices to rise. An overview of the sign restrictions used can be found in table 1. It is important to note that we

Table 1. The Identification Scheme

Variable/Shock	Demand	Monetary Policy	Cost-Push
Real GDP Growth Euro Area	+	+	−
Policy Rate Euro Area	(+)	−	*
CPI Inflation Euro Area	+	+	+
Real Interest Rate Euro Area	+	(−)	*

Note: The signs in parentheses are implicitly fulfilled, given the explicit sign restrictions imposed to identify the shock. An asterisk (*) indicates no sign restriction imposed.

place restrictions only on the responses of foreign factors and remain agnostic about the reaction of the domestic economy as well as the exchange rate.[7] As a baseline, we impose these restrictions for $h \leq 1$ periods.[8]

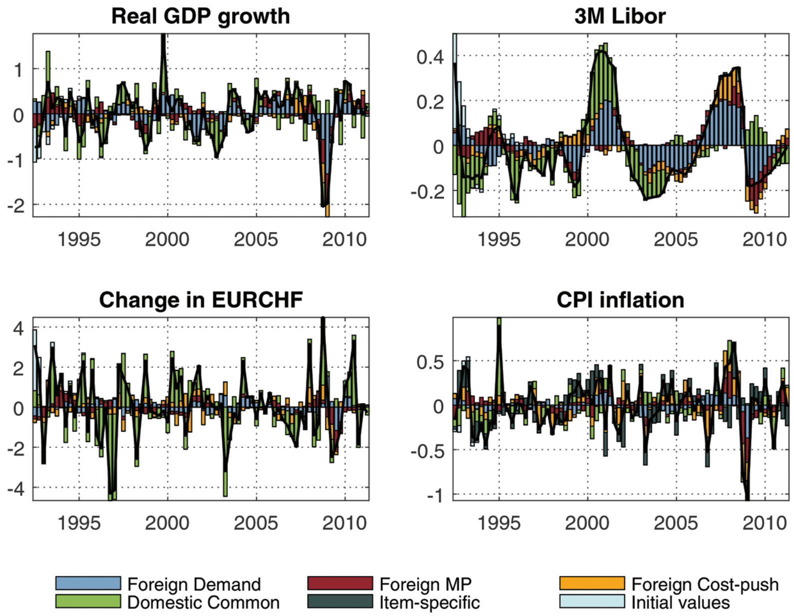
We chose this identification scheme with short-run and sign restrictions because it is well established in the literature and theoretically founded. To check whether the identification scheme makes sense, we also compute the contributions of the structural foreign shocks and the domestic shocks to quarterly changes in the detrended levels of the Swiss real GDP, the Swiss CPI, and the EURCHF as well as the detrended level of the 3M LIBOR.[9] The results point to reasonably identified shocks as shown in figure 2. For example, positive foreign demand shocks contributed strongly

[7]As the foreign shocks are block identified, additionally identifying domestic shocks is irrelevant for the identification of the foreign shocks for a given set of reduced-form parameters. However, if sign restrictions are used, they may have an influence on the posterior distribution of reduced-form parameters. This is because different draws of the reduced-form parameters may have different probabilities of satisfying the sign restrictions. We do not think that it is sensible to use sign restrictions solely to inform us on the probability of the reduced-form parameters, such that we do not follow this route.

[8]By restricting two quarters (the current and one future quarter), this horizon is consistent with the horizon chosen by Uhlig (2005), who uses five periods with monthly data.

[9]All variables are detrended with the use of the two-sided Hodrick-Prescott filter. Note that the variables are not detrended in the baseline model. However, as shown in the robustness analysis, detrending the variables does not alter our conclusions.

Figure 2. Historical Decomposition of Detrended Swiss Macroeconomic Variables



Notes: The figure shows the historical contributions of the structural foreign shocks and the domestic shocks to quarterly changes in the detrended levels of the Swiss real GDP, the Swiss CPI, and the EURCHF as well as the detrended level of the 3M LIBOR. All variables are detrended with the use of the two-sided Hodrick-Prescott filter.

to real GDP growth in Switzerland and also supported Swiss CPI inflation in the period from around 2006 to 2008. During that period, real GDP growth in the euro area was particularly strong. The same is true for the early 2000s. At the end of 2007 and the beginning of 2008, Swiss inflation picked up due to cost-push shocks. In this period, the oil price increased strongly, before collapsing right after the onset of the financial crisis. This is reflected in negative cost-push shocks in late 2008. At that time, negative foreign demand shocks and restrictive monetary policy shocks also weighed on Swiss inflation.

To implement the sign restrictions conditional on the zero restrictions, we use the method proposed by Arias, Rubio-Ramírez, and

Waggoner (2018).[10] Based on the draws that satisfy the identification scheme, we compute statistics that facilitate the interpretation of the results. In particular, we look at impulse response functions (IRFs) and the fraction of forecast error variance decomposition (FEVD). Highest probability density (HPD) intervals on these statistics are calculated "pointwise," i.e., for each horizon separately.

3. Results

In this section, we present the results of our empirical analysis. We start by discussing the transmission of inflationary shocks in the euro area to the Swiss economy. Subsequently, we analyze the quantitative importance of foreign and domestic shocks on a set of Swiss macroeconomic variables—with a specific focus on consumer prices. After studying the effects at the aggregate level, we investigate whether disaggregated Swiss consumer prices are affected differently by international inflation spillovers. Finally, we check the robustness of our results and discuss the implications for monetary policy.

3.1 International Spillovers to the Swiss Economy

How do foreign inflationary pressures originating from different shocks in the euro area transmit to the Swiss economy and in particular to consumer prices? We analyze this question by looking at the

[10] This is a difference to a previous version of this paper (Bäurle, Gubler, and Känzig 2017), in which we rely on the algorithm proposed in Arias, Rubio-Ramírez, and Waggoner (2014). As Arias, Rubio-Ramírez, and Waggoner (2018) show, in order to correctly draw from the structural parameterization, one needs to take into account the effects of the change of variable induced by the mapping from the orthogonal reduced-form parameterization to the structural parameterization by introducing an additional importance sampling step. As we implement zero restrictions on the reduced form, we cannot use the conjugate prior suggested in Arias, Rubio-Ramírez, and Waggoner (2018). However, as we use the same structural parameterization, applying their importance sampling step to our "proposal distribution" (i.e., the distribution derived with the algorithm from Arias, Rubio-Ramírez, and Waggoner 2014) correctly adjusts for the change of variable. Note that as robustness tests, we first estimate the model with our prior but refraining from implementing zero restrictions on the reduced form. It turns out that the results hardly differ when the small open economy assumption is dropped. We then proceed by using the conjugate prior of Arias, Rubio-Ramírez, and Waggoner (2018). The results are again robust to this change in the specification. These results are available from the authors on request.

impulse responses to the identified shocks. The impulse responses to the three identified foreign shocks—demand, monetary policy, and cost-push shocks—are presented in figure 3 and figure D.1 (in online appendix D). In addition, figure 4 shows the responses of the relative consumer price indexes as well as the nominal and real interest rate spreads between the euro area and Switzerland to these shocks. The median response is depicted by the solid black line. The light-gray shaded areas represent 68 percent HPD intervals. Cumulative responses are shown for all variables except the interest rates. The response of the Swiss CPI is calculated based on the disaggregated price responses and the corresponding CPI weights.[11] Similarly, we can compute the responses of different categories of the CPI.

3.1.1 Response to Foreign Demand Shocks

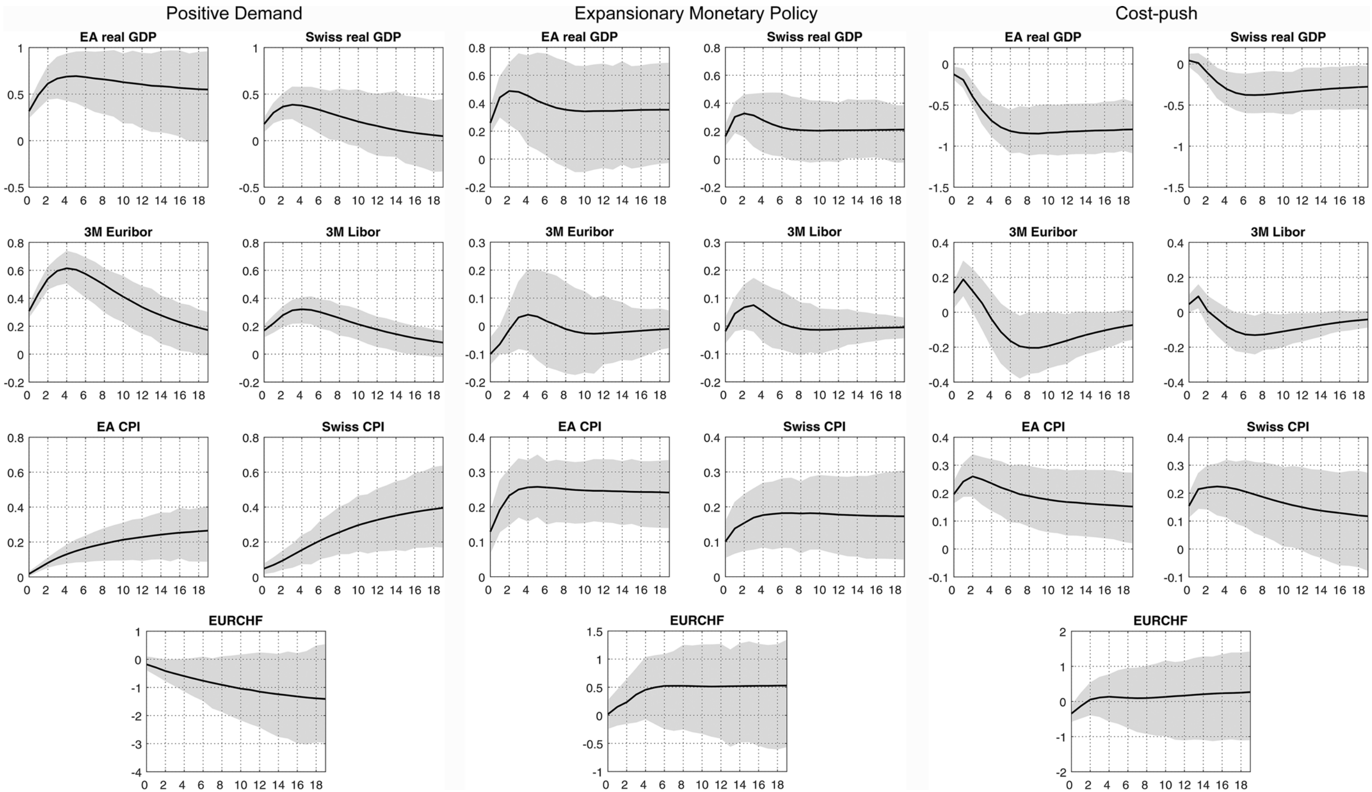
A positive shock to demand in the euro area leads to a persistent rise in foreign output, consumer prices, and the real interest rate—consistent with our identifying restrictions. The demand-driven boom in the euro area has substantial spillover effects on the Swiss economy. Both Swiss output and prices rise strongly, consistent with the fact that Switzerland is an open economy and thus heavily dependent on the foreign economic development.[12] However, while the demand-driven upturn is counteracted by substantial hikes in policy rates in the euro area, the Swiss monetary policy reaction turns out to be less restrictive (as reflected by a weaker response of the real interest rate shown in figure 4). Consistent with these changes in the relative monetary policy stance, the Swiss franc depreciates against the euro.[13] Spillover effects to Swiss consumer prices

[11]To be more precise, the CPI is computed as a weighted average of the prices of the different CPI items, $\log(\mathrm{CPI}_t) = \sum_{i=1}^{148} \omega_{i,t} \log(\mathrm{price}_{i,t})$, where $\omega_{i,t}$ is the weight of item i in the CPI from the SFSO.

[12]In contrast, Mumtaz and Surico (2009) find no significant change in real activity in the United Kingdom in response to an unanticipated increase in foreign real activity.

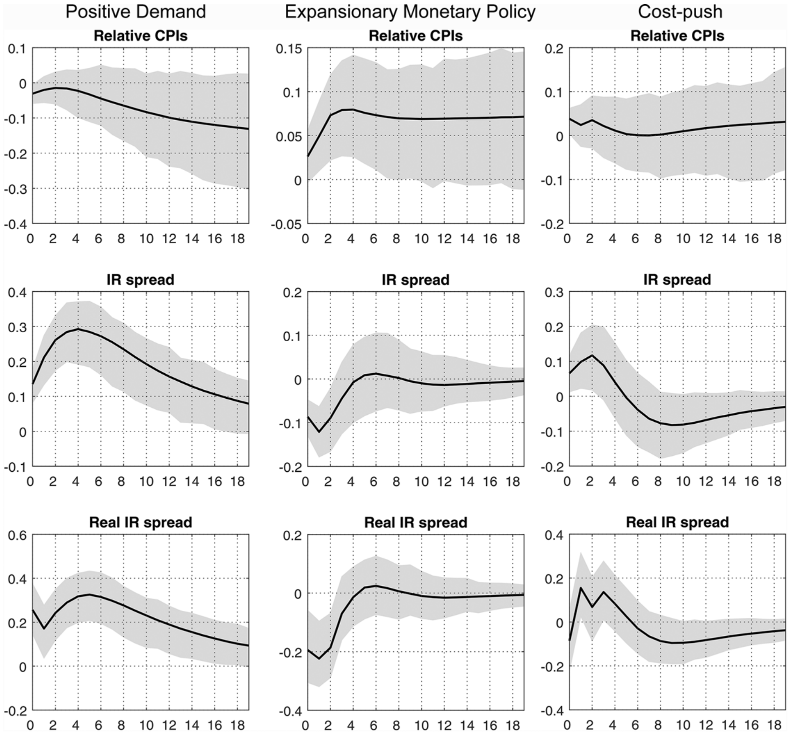
[13]The exchange rate response is broadly consistent with the predictions of the uncovered interest rate parity (UIP). While the focus of this paper is not on the UIP (puzzle) in particular, we nevertheless compute the responses of the forward discount premium, defined as in Mumtaz and Surico (2009), to all three shocks. After the demand shock, it is, although hardly statistically significant, negative over a prolonged period. This reflects the longer-lasting depreciation of

Figure 3. Impact of Foreign Inflationary Shocks on Common Factors and Aggregate Swiss CPI



Notes: The figure illustrates the impulse responses to one-standard-deviation structural shocks at horizons up to 20 quarters (along the x-axis). The median response is depicted by the bold black line. The light-gray shaded area represents the 68 percent HPD interval. For all variables, the cumulative responses are shown except for the interest rates. The responses of the interest rates along the y-axis can be interpreted as the annualized quarter-on-quarter change in percentage points. All other responses along the y-axis denote percentage changes.

Figure 4. Impact of Foreign Inflationary Shocks on Spreads between Swiss and Euro-Area CPI and Interest Rates



Notes: The figure illustrates the impulse responses to one-standard-deviation structural shocks at horizons up to 20 quarters (along the x-axis). The median response is depicted by the bold black line. The light-gray shaded area represents the 68 percent HPD interval. For the relative CPIs, the cumulative responses are shown. The responses of the interest rate spreads along the y-axis can be interpreted as the annualized quarter-on-quarter change in percentage points. The responses of the relative CPIs along the y-axis denote percentage changes. Relative CPIs: log euro-area CPI – log Swiss CPI. IR spread: 3M EURIBOR – 3M LIBOR. Real IR spread: Real 3M EURIBOR (3M EURIBOR – euro-area CPI inflation) – real 3M LIBOR (3M LIBOR – Swiss CPI inflation).

---

the Swiss franc depicted in figure 3. After the monetary policy and the cost-push shock, the forward discount premium is slightly positive in the short run, but not statistically significant. The response of the forward discount premium to the monetary policy shock is in line with the evidence found by Mumtaz and Surico (2009) for domestic monetary policy shocks in the United Kingdom. The results are available from the authors upon request.

turn out to be substantial. Compared with the euro area, Swiss consumer prices initially increase more sluggishly but eventually attain a higher level in the longer run (see figure 4).

3.1.2 Response to Foreign Monetary Policy Shocks

By construction, an expansionary monetary policy shock in the euro area leads to a fall in the 3M EURIBOR stimulating consumption and investment, which in turn causes output and consumer prices to increase. The economic upturn initiated by the expansionary monetary policy shock also has substantive effects on Switzerland: both output and prices increase significantly. In contrast to what we observe in response to foreign demand shocks, however, the monetary policy stance becomes relatively more restrictive in Switzerland for approximately one year. Furthermore, the Swiss franc now appreciates against the euro. This has likely cushioning effects on the magnitude of spillovers to Swiss prices. Indeed, it turns out that Swiss consumer prices rise by less than in the euro area, as shown in figure 4.

In light of the substantial positive impact of an expansionary monetary policy shock in the euro area on Swiss output, our results do not confirm that a beggar-thy-neighbor mechanism is at work. Liu, Mumtaz, and Theophilopoulou (2014) find similar results for the United Kingdom since the 1990s. This result suggests that the expenditure switching effect—where Swiss consumers increasingly buy imported instead of locally produced products as imports become relatively cheaper because of the Swiss franc appreciation—does not dominate. Indeed, import prices fall in contrast to the increase following a positive foreign demand shock, but the response remains limited (see figure D.1 in online appendix D). In particular, there is no clear evidence of relative price changes between domestic and imported goods in the long run.

3.1.3 Response to Foreign Cost-Push Shocks

A cost-push shock in the euro area is associated with a rise in consumer prices together with a fall in output. This shock introduces a tradeoff for most central banks. Even if price stability is the primary concern, central banks often also consider developments in the real economy for their decisionmaking.[14] We find that, on impact, the price response dominates and monetary policy in the euro area becomes more restrictive to counteract the inflationary pressures. As time evolves, the adverse effects on output become more pronounced and euro-area monetary policy becomes more expansive again. The economic downturn in the euro area also has non-negligible effects on the Swiss economy. After a slight delay, Swiss output starts to fall significantly, but the response turns out to be less pronounced than in the euro area. Despite the 3M EURIBOR rising slightly more strongly than the 3M LIBOR in the short term, the relative monetary policy stance remains fairly unchanged given that the real interest rate in Switzerland moves almost in step with that of the euro area. This may prevent the Swiss franc from depreciating more strongly. Swiss consumer prices also rise significantly but the magnitude of the response is comparable to the price response in the euro area (see figure 4).

3.1.4 Comparison of Responses

To summarize, all three foreign shocks result by construction in temporary higher inflation in the euro area. Likewise, the shocks are associated with temporary higher Swiss inflation. This is not surprising given the strong trade linkages between the euro area and Switzerland. Interestingly, however, we find that the magnitude of the spillover effects on Swiss prices depends crucially on the underlying forces driving foreign inflationary pressures and the associated general equilibrium effects.

While the inflation differential between the euro area and Switzerland narrows in response to positive demand shocks, it widens in response to expansionary monetary policy shocks and does

[14]The Swiss National Bank's mandate is described in the National Bank Act (Article 5, Paragraph 1): "It shall ensure price stability. In so doing, it shall take due account of economic developments." The Lisbon Treaty (Article 127, Paragraph 1) states that "The primary objective of the European System of Central Banks [...] shall be to maintain price stability. Without prejudice to the objective of price stability, [the European System of Central Banks] shall support the general economic policies in the Union with a view to contributing to the achievement of the objectives of the Union." These objectives include "full employment" and "balanced economic growth."

not change significantly in response to cost-push shocks. This illustrates the importance of taking the underlying source of the international inflationary pressures as well as the corresponding general equilibrium effects into account when analyzing spillovers to domestic inflation. In particular, the foreign shocks have very different implications on the relative monetary policy stance and exchange rates. While monetary policy becomes relatively more restrictive in the euro area and the exchange rate depreciates after demand shocks, the relative monetary policy becomes less restrictive and the exchange rate appreciates after monetary policy shocks and remains broadly unchanged after cost-push shocks. These differences are consistent with the varying degree of spillovers to consumer prices. In addition, also note that the responses of the exchange rate and inflation abroad and in Switzerland to all three shocks are in line with the purchasing power theory, without imposing any restrictions on the joint behavior of these variables.

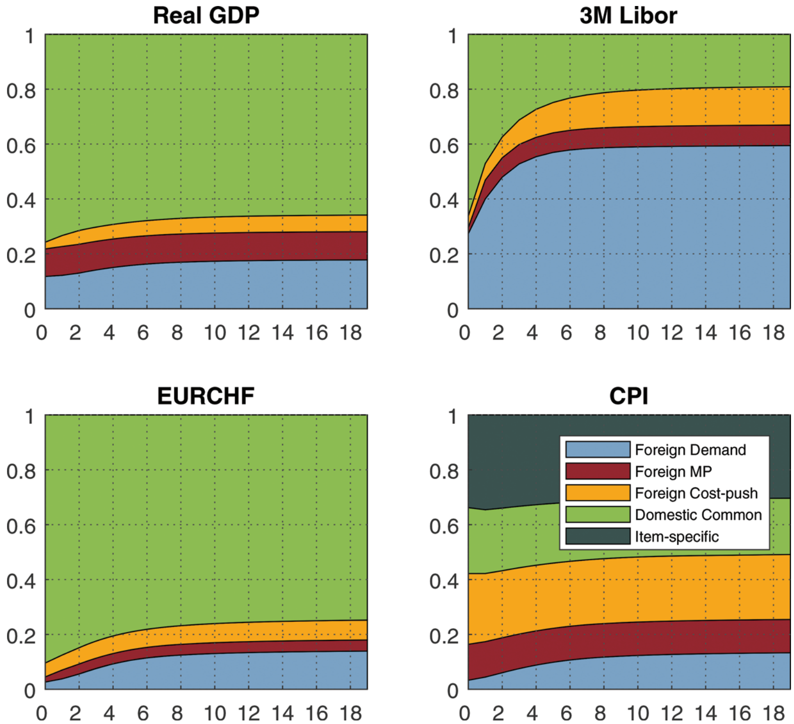
Overall, our findings are in line with recent evidence of Comunale and Kunovac (2017) and Forbes, Hjortsoe, and Nenova (2018) who document that the exchange rate pass-through is dependent on the nature of the shocks, and Bobeica, Ciccarelli, and Vansteenkiste (2019), who find shock-dependent pass-through effects of labor costs to prices. We show that the same is true for international inflation spillovers. This can also help to reconcile the ambiguous empirical evidence on the co-movement of domestic and global inflation.

3.2 Foreign versus Domestic Inflationary Pressures

As shown in the previous subsection, foreign inflationary shocks can have substantial spillover effects on the Swiss economy. An important question in this context is, how important are spillover effects induced by foreign shocks relative to domestic forces? To answer this question and to obtain a better understanding of the relative importance of the different foreign shocks, we conduct a variance decomposition exercise.

Figure 5 shows the variance decomposition for the domestic common factors and the CPI constructed from the disaggregated price data. Depicted is the fraction of forecast error variance that is explained by the three identified foreign shocks as well as the unidentified domestic common and item-specific shocks at different

Figure 5. Variance Decomposition of Swiss Common Factors and Aggregate CPI



Note: The figure illustrates the posterior mean of the forecast error variance decomposition of shocks (along the y-axis) at horizons up to 20 quarters (along the x-axis).

horizons. Note that the set of unidentified domestic common shocks also includes reduced-form shocks to the exchange rate. It turns out that foreign shocks account for a substantial part of the variance of Swiss variables. In the medium run, they explain up to about 25 percent of the variation in the exchange rate, 40 percent of real GDP, 80 percent of the 3M LIBOR, and 50 percent of the CPI. In the case of the CPI, the remaining part is explained by domestic common and item-specific shocks.[15] Approximately 30 percent of the variations

[15]Recall that the CPI does not enter our system as an observable but is constructed from the disaggregated price data and thus features, in contrast to the other aggregates shown, an idiosyncratic part.

in the CPI are explained by item-specific shocks, while domestic common shocks account for approximately 20 percent. The finding that approximately half of the variation of the Swiss CPI is driven by foreign shocks is in line with the findings of Jordan (2015). It is also in line with results for other (small) open economies (Aastveit, Bjørnland, and Thorsrud 2016).

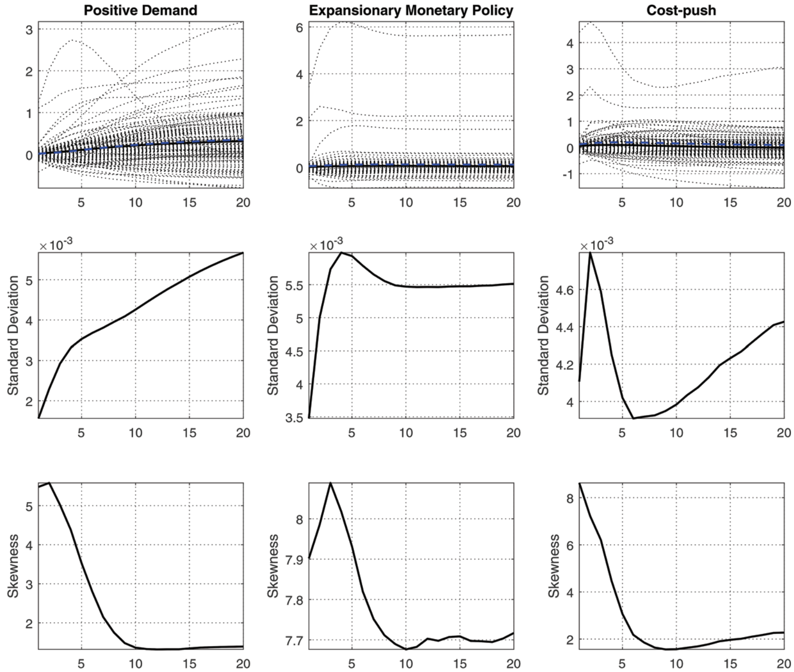
The bulk of the foreign contribution to output, the interest rate, and the exchange rate is due to foreign demand shocks. Monetary policy shocks generally account for a smaller fraction, which is a common finding in the literature and can be reconciled with the fact that this shock is thought to capture unsystematic variations in the policy stance, which should be small. Consumer prices turn out to be heavily driven by foreign cost-push shocks, particularly in the short run. At longer horizons, however, demand shocks become more important, whereas the contribution of foreign cost-push shocks slightly decreases. The variations in the relative importance of the different foreign shocks point to heterogeneous spillover effects to Swiss prices.

By way of summary, our results indicate that international spillovers to the Swiss economy and to Swiss prices in particular are quantitatively important. Foreign demand shocks turn out to be an important driver of Swiss macroeconomic variables in general, and for Swiss prices, foreign cost-push shocks appear to be particularly important as well.

3.3 Heterogeneity in Spillovers to Swiss Consumer Prices

So far, we have focused our analysis on spillover effects at the aggregate level. However, our dynamic factor modeling framework allows us to study these effects at a highly disaggregate level as well, as it includes a vast number of disaggregated data on Swiss consumer prices. This may give valuable insights on item-specific differences. Furthermore, by comparing items with different degrees of tradability, we can shed some light on the importance of indirect, general equilibrium effects relative to the direct, mechanical pass-through of international to country-specific inflation and how this varies with different foreign inflationary shocks. Finally, working with disaggregated prices also allows us to get a sense of how the aggregation level of inflation can influence the estimated degree of spillovers.

Figure 6. Impact of Foreign Inflationary Shocks on Disaggregated Prices



Notes: The top panels of the figure illustrate the posterior median responses of the 148 CPI items for the three identified shocks at horizons up to 20 quarters (along the x-axis). The middle (bottom) panels show the standard deviation (skewness) of the responses across items for the three identified shocks at horizons up to 20 quarters (along the x-axis).

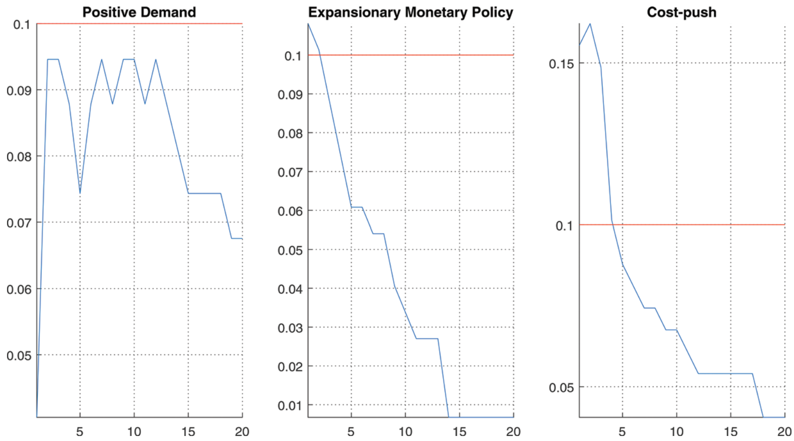
In the top panels of figure 6, we report the posterior median responses of the 148 CPI items for the three identified shocks. There is substantial heterogeneity in the price responses to all three shocks. Most prices tend to increase; however, some prices increase by a substantially smaller amount, while other prices even decrease. Interestingly, the price responses are less dispersed in the short run for foreign demand shocks but then become more dispersed over time. In contrast, the price dispersion tends to be larger in the short run for cost-push shocks and does not change much over the response horizon for monetary policy shocks. This is also confirmed by the middle panels in figure 6, which show the standard deviation of the responses across items over the response horizon. One can see that for the demand shock, the standard deviation increases gradually. For the monetary policy shock, the price dispersion is relatively constant over the response horizon after a strong and quick initial increase. Finally, the price dispersion after cost-push shocks spikes significantly in the first couple of quarters, which appears to be driven by the responses of energy prices, and then fluctuates at a lower level. In the bottom panels of figure 6, we show that the distribution of price responses is positively skewed, particularly in the first 8–10 quarters. The increase in skewness indicates that foreign inflationary shocks transmit to the Swiss economy as shocks to relative prices in Switzerland. In line with Mumtaz and Surico (2009), we find a positive relationship between skewness and the aggregate price response, which is supportive of the fact that shocks to relative prices can be inflationary.

To take into account the uncertainty around these estimates, we follow Mumtaz and Surico (2009) and analyze for each item its distribution of relative price responses. For each item i and draw from the posterior j, we compute the relative price response $\ln p_i^j - \ln \bar{p}^j$, where $\ln \bar{p}^j$ is the average (log) price response over all items for draw j. After having done this for all draws j, we compute for each item the fraction of relative price responses that are positive (across j), which we denote by α_i. By looking at the proportion of items which fall into $\mathcal{S}_\alpha = \{i : \alpha_i < 0.05 \text{ or } \alpha_i > 0.95\}$ (i.e., items for which more than 95 percent of the responses decrease or increase, respectively, compared with the average response over all items), we can then evaluate whether the change in relative prices shows some statistical significance. If the share of items falling into $\mathcal{S}_\alpha$ is larger than 10 percent, we may conclude that the measured change in relative prices is not the result of estimation uncertainty.[16]

Figure 7 shows the fraction of items falling into $\mathcal{S}_\alpha$. Our results for the foreign monetary policy shock and the cost-push shock are in line with Boivin, Giannoni, and Mihov (2009) for domestic monetary policy shocks and Mumtaz and Surico (2009) for international

[16] Mumtaz and Surico (2009) interpret a fraction of responses above this threshold as significant, stating that "one would typically expect 10% of the price responses to be significantly different from the average."

Figure 7. Fraction of Items for which $\alpha_i < 0.05$ or $\alpha_i > 0.95$



Notes: For each item i and draw j, we compute the relative price response $\ln p_i^j - \ln \bar{p}^j$, where $\ln \bar{p}^j$ is the average (log) price response over all items for draw j. After having done this for all draws j, we compute for each item the fraction of relative price responses that are positive (across j), which we denote by α_i. The figure shows for the three identified shocks the fraction of items for which $\alpha_i < 0.05$ or $\alpha_i > 0.95$.

supply shocks: in the short to medium run, there is evidence for significant relative price movements, as can be seen by the fact that the proportion of items falling into $\mathcal{S}_\alpha$ lies above 10 percent at horizons for up to one year. In the longer run, however, the responses converge to the average, as can be seen from the fact that the share of items that differ significantly from the average converges to zero. The results for the demand shock turn out to be quite different. While there seems to be no significant change in relative prices in the very short run, there is some evidence of significant relative price changes in the medium run as the share gets close to 10 percent (even though it never surpasses the threshold) and only slowly diminishes toward the forecast horizon. This is consistent with the persistent aggregate price response after a positive foreign demand shock.

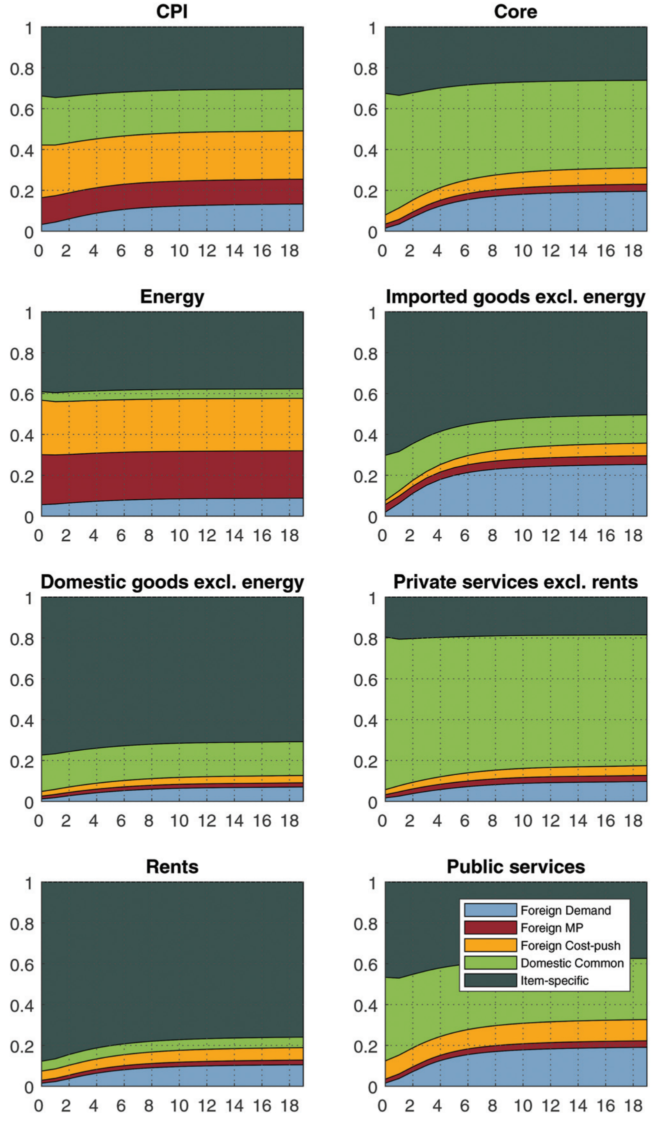
An analysis of a selection of different categories of the CPI reveals insights on the channels leading to the dispersed responses. Specifically, we look at core CPI, energy, imported goods excluding energy, domestic goods excluding energy, private services excluding rents, and public services. In defining the core measure, we follow the Swiss Federal Statistical Office and exclude fresh and seasonal items as well as energy. Moreover, we define all items with an average import share of above 50 percent over the sample period to be imported. Analogous to aggregate CPI, the statistics for these categories are computed based on the weights of the items. We focus here on the variance decomposition of the different categories; however, the corresponding impulse responses can be found in online appendix D.[17] Figure 8 presents the variance decomposition for aggregate CPI and the selected price categories at different horizons.

The main results are twofold: first, the contribution of foreign shocks to core CPI is substantially lower when compared with the headline, particularly in the shorter term. It turns out that these differences are likely driven by energy prices. Energy prices are heavily affected by foreign shocks, and the transmission appears to occur quite fast, as the foreign contribution stands at approximately 60 percent on impact and remains roughly at the same level afterward. A large part of this contribution can be attributed to foreign cost-push shocks, which seems quite intuitive because these shocks likely reflect to a large extent unexpected changes in global energy prices (e.g., supply-driven oil price shocks). The strong and direct impact of foreign cost-push shocks on energy prices appears to be transmitted to headline CPI, for which cost-push shocks also explain a dominant share, particularly in the shorter term. In contrast, the major part of the foreign contribution in core CPI, which does not include energy prices, is due to demand shocks, whereas cost-push shocks explain considerably less.

Second, there are some interesting heterogeneities in the relative importance of foreign shocks, which are likely related to differences in tradability and exchange rate sensitivity of the respective price categories. For categories featuring a high tradability—e.g., energy prices—foreign shocks explain a large share of the price variations. In contrast, foreign shocks account for a smaller fraction of categories

[17]Note that we calculate the decomposition for the particular indexes, which are obtained by aggregating the items included in the categories of interest. Because of this, however, our results are not directly comparable to those of Monacelli and Sala (2009), who use a simple average of the decompositions for the single items in a given category. In online appendix C, we discuss the role of the aggregation in more detail.

Figure 8. Variance Decomposition of Different Categories of Swiss Consumer Prices



Note: The figure illustrates the posterior mean of the forecast error variance decomposition of shocks (along the y-axis) at horizons up to 20 quarters (along the x-axis).

that are hardly tradable, such as domestic private services, while domestic common shocks turn out to be relatively more important. Still, foreign shocks explain a non-negligible part of the variation in domestic goods and service inflation. This is in line with the findings of Halka and Szafranek (2016) for the Czech Republic, Poland, and Sweden.

These results suggest that while a certain part of spillovers to Swiss consumer prices is likely mechanical due to direct effects on import prices, general equilibrium effects are important as well. The relative importance of these effects also seems to vary with the shocks. While monetary policy and cost-push shocks explain a comparably rather low but roughly equal share of the variations in the prices of imported (excluding energy) and domestic goods, demand shocks account for a much larger share of the variations in prices of imported goods, excluding energy. In contrast, cost-push shocks (and to some extent also monetary policy shocks) explain a much higher share of the variance in energy prices. The contribution of demand shocks to energy prices is, however, much smaller and more in the range of those to prices of other domestic goods and services.

To summarize, our empirical findings point to significant differences between different price categories. It turns out that energy prices play an important role. Indeed, foreign shocks explain a markedly lower share of consumer prices when energy prices are excluded. This finding is in line with recent empirical evidence presented in other studies. Parker (2016) documents that global factors are particularly important in explaining energy prices, and Halka and Kotlowski (2017) conclude that commodity-specific shocks are an important source of inflation variability in the Czech Republic, Poland, and Sweden. Altansukh et al. (2017) argue that in a low-inflation environment such as the one analyzed in this paper, the volatility of energy inflation has become relatively more important for explaining short-run changes in headline inflation.

3.4 Robustness

We check the robustness of our results along a number of dimensions, including the specification of the model, the choice of priors, the identifying assumptions, as well as the sample period. All the results are shown in online appendix D.

3.4.1 Model Specification

An important question is whether our results do uniquely pertain to spillovers from the euro area or to international inflation spillovers more generally. To address this issue, we replace the euro-area block of the model with a global block consisting of export-weighted indicators of Switzerland's most important trading partners.[18] The results based on the global factors are in line with the results based on the factors for the euro area (see figures D.2 and D.3). Consequently, our findings not only pertain to spillovers emerging from the euro area but to international inflation spillovers to Switzerland more generally.

The baseline specification also does not include a measure of global financial conditions, which may be another transmission channel for international inflation spillovers. To control for this channel, we augment the global block by a trade-weighted financial conditions index (FCI).[19] The impulse responses of our constructed FCI to the three identified shocks are as expected (see figure D.4). More importantly, the conclusions drawn from the baseline model remain intact (see also figure D.5), even after controlling for financial conditions abroad.

Another potential concern is that in our baseline we include short-term interest rates, which were constrained by the effective lower bound (ELB) in the last part of our sample.[20] To analyze whether our results are effected by this, we reestimate our model

[18]In particular, we use export-weighted real GDP as a measure of foreign demand; an export-weighted CPI as a measure of foreign CPI; an export-weighted policy rate of the United States, euro area, and Japan as a measure of the foreign policy stance; and the nominal effective exchange rate (NEER) as the relevant exchange rate index. A positive exchange rate change implies an appreciation of the Swiss franc as in the baseline analysis with the EURCHF.

[19]We focus here on the model using the global variables because of the lack of a good proxy for financial conditions with sufficient time coverage for the euro area. For the period from 1999 onwards, we construct a trade-weighted financial conditions index using the U.S. excess bond premium provided by Gilchrist and Zakrajšek (2012) and a BBB bond spread for the euro area as used in Jarocinski and Karadi (2018). For the period before 1999, we extrapolate the series with the U.S. excess bond premium only.

[20]Note that in the last part of our sample the 3M LIBOR and 3M EURIBOR were close to or stuck at zero. The European Central Bank and the Swiss National Bank lowered short-term rates to negative levels only after the end of our sample.

using short-term shadow interest rates. The shadow rates are from Krippner (2013) (see figure D.6).[21] The results are shown in figures D.7 and D.8. Overall, the results turn out to be robust to this change.

Some of the (nominal) variables in our model exhibit a slight downward trend in the first part of the sample. Therefore, we check the robustness of our conclusions to estimates with all variables detrended using the two-sided Hodrick-Prescott filter. From the results shown in figures D.9 and D.10, we can conclude that our findings are robust to the treatment of these trends. However, the impulse responses of Swiss GDP and CPI to foreign demand and cost-push shocks are somewhat more volatile than in the baseline estimation.

Finally, in our baseline specification, the disaggregated prices load on seven factors—six observed and one unobserved—both contemporaneously and on one lag. Instead of allowing for one lag in the observation equation, we estimate the model with two unobserved factors. The results support our conclusions. While the variance decomposition in the model with two unobserved factors and $q = 0$ points to slightly stronger spillover effects of foreign demand shocks on Swiss inflation of domestic goods (see figure D.12), the impulse responses of the observed factors and Swiss CPI inflation are very similar to those in our baseline model (see figure D.11).

Figures D.13 and D.14 show the results with fewer lags in the estimation equations. The state equation contains only one lag ($p = 1$) and the observation equation only the contemporaneous impact ($q = 0$). Having fewer lags in the model results in more persistent impulse responses, particularly for the interest rates. Moreover, the variance decomposition suggests smaller explanatory power of foreign shocks for Swiss inflation. Adding lags, however, does not alter the baseline model results considerably, with one exception—the tendency of the Swiss franc to appreciate in response to an expansionary foreign monetary policy shock is weaker. Figures D.15 and D.16 show the impulse responses and variance decomposition with the lag order

[21] We are grateful to Leo Krippner for sharing the data. Since they were only available from a later point in time than the starting quarter of our estimation period, we link them with the observed interest rates in 1999, when the euro was established. This should not be a problem, given that interest rates were not constrained by the ELB in the period from 1992 to 1999.

in the state and the observation equations of $p = 3$ and $q = 2$, respectively. The findings suggest that the number of lags in the baseline model is sufficient to capture the relevant dynamics.

3.4.2 Estimation and Identification

We also checked the sensitivity to the prior tightness. We find that the results are robust to a change in the prior parameters from $\pi_1 = \pi_2 = \pi_3 = 1$ to $\pi_1 = 0.2$, $\pi_2 = 0.7$, $\pi_3 = 2$ (which is in the region used by Karlsson 2013 in his forecasting exercise). Given the increased prior tightness, we allow for four lags in the state equation. Figures D.17 and D.18 show the results. The most noticeable difference is that the exchange rate response to a foreign demand shock becomes somewhat weaker, but the sign remains the same such that our conclusions are still supported.

In our baseline model specification, both demand shocks seem to have permanent effects on the GDP level, although not or only hardly statistically significant. Nevertheless, this indicates that the assumption of monetary policy neutrality in the long term is violated. Therefore, we impose long-run zero restrictions on the effect of the two demand shocks on foreign GDP. The results (see figures D.19 and D.20) remain robust to this modification in the shock identification strategy.

3.4.3 Subsample Analysis

To ensure that the impact of the financial crisis and the Great Recession does not drive our results, we estimate the model with data until 2007:Q4. We find that the main results do not change, with one exception (see figures D.21 and D.22). After a demand shock, the interest rate differential between the EURIBOR and the 3M LIBOR is more stable in nominal terms, and in real terms, there is almost no reaction once the crisis period is removed from the sample.[22]

The fact that the interest differential is less stable when the crisis period is included might reflect that the 3M LIBOR was close to

[22]Note that the size of the demand shock is smaller in the shorter sample, offering an explanation for the more muted reaction of the foreign interest rate level.

zero in 2010 and 2011. As a result, after the onset of the financial crisis, Swiss monetary policy might have been limited to some extent in counteracting negative foreign demand shocks with its interest rate instrument, making a stabilization of the interest rate differential infeasible. Indeed, Bäurle and Kaufmann (2018) find that the response of the trade-weighted Swiss franc exchange rate to risk shocks becomes more pronounced at the effective lower bound for nominal interest rates. Thus, our result that the relative interest rate reaction is influenced by the ELB periods and the exchange rate reaction changes supports their result. Furthermore, it supports our conclusion that monetary policy can affect the spillovers of shocks. Indeed, in line with a stable interest rate differential in the pre-crisis period, the Swiss franc does not depreciate in response to a positive foreign demand shock.[23]

4. Conclusion

In this paper, we analyze how different shocks driving up inflation abroad translate into inflationary pressures in Switzerland, putting particular emphasis on general equilibrium effects such as the relative monetary policy and exchange rate responses. Based on a structural dynamic factor model relating a large set of disaggregated Swiss consumer prices to key foreign and domestic factors, we study how foreign inflationary shocks are transmitted to Swiss prices. We identify three different types of inflationary shocks that are widely discussed in the literature: a positive demand shock, an expansionary monetary policy, and a cost-push shock, all originating abroad.

We find that foreign shocks explain up to approximately 50 percent of Swiss price variations, while common domestic shocks account for only approximately 20 percent (with the remaining part being due to item-specific shocks). Thus, domestic inflation is to

[23]We even see a tendency of a slight appreciation. However, the appreciation disappears when we replace the euro area with a broader measure for the foreign economy as described above. With the other responses remaining very similar (this result is not shown but can be obtained by the authors), we do not want to over-emphasize this result. In line with the exchange rate response, the spillover effects through imported inflation to Swiss CPI inflation are more muted in the shorter sample.

a substantial degree driven by foreign shocks. However, this does not imply that Swiss monetary policy has not been able to affect international spillover effects to domestic inflation. Indeed, spillover effects on Swiss prices are crucially dependent on the nature of the underlying shock and the associated general equilibrium effects. Following an increase in foreign inflation due to a positive demand shock, foreign monetary policy counteracts the business cycle upturn strongly, while the Swiss monetary policy reaction turns out to be less restrictive. Consistent with the change in the relative monetary policy stance, the Swiss franc depreciates and inflation picks up even somewhat more than abroad. In contrast, in response to an increase in inflation abroad due to an expansionary monetary policy shock, monetary policy becomes relatively tighter in Switzerland and the franc appreciates—mitigating spillover effects to Swiss inflation. Finally, a cost-push shock driving up inflation (and decreasing real activity) abroad has no significant effects on the relative monetary policy stance. The effects on the exchange rate turn out to be negligible, and the increase in Swiss inflation is comparable to inflation abroad.

Our analysis of the different items of the Swiss CPI points to substantial heterogeneities in the transmission of foreign inflationary shocks. It turns out that energy prices play a crucial role in the transmission, particularly in the short run. The impact of foreign inflationary shocks on Swiss CPI is lower, and the transmission appears to be slower when energy prices are excluded. Furthermore, by comparing the relative price changes of goods and services with different degrees of tradability, we can get a sense of the importance of the indirect, general equilibrium effects relative to direct, mechanical spillover effects. Our results indicate that while a certain part of spillovers is likely mechanical, general equilibrium effects are important as well. Interestingly, the importance of these effects also seems to vary with the underlying shocks.

These results indicate that spillover effects need to be analyzed in a framework allowing for different transmission channels: an increase in inflation abroad may affect the inflation in an open economy differently, depending on the source of the foreign shock, and thus, movements in other factors such as activity, interest rates, and exchange rates. This may also partly explain the ambiguity of the empirical evidence on the impact of spillovers found so far.

References

Aastveit, K. A., H. C. Bjørnland, and L. A. Thorsrud. 2016. "The World is Not Enough! Small Open Economies and Regional Dependence." *Scandinavian Journal of Economics* 118 (1): 168–95.

Altansukh, G., R. Becker, G. Bratsiotis, and D. R. Osborn. 2017. "What is the Globalisation of Inflation?" *Journal of Economic Dynamics and Control* 74 (January): 1–27.

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner. 2014. "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications." Working Paper No. 2014-1, Federal Reserve Bank of Atlanta.

———. 2018. "Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications." *Econometrica* 86 (2): 685–720.

Baumeister, C., and L. Benati. 2013. "Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound." *International Journal of Central Banking* 9 (2, June): 165–212.

Bäurle, G. 2013. "Structural Dynamic Factor Analysis Using Prior Information from Macroeconomic Theory." *Journal of Business and Economic Statistics* 31 (2): 136–50.

Bäurle, G., M. Gubler, and D. Känzig. 2017. "International Inflation Spillovers — The Role of Different Shocks." Working Paper No. 7/2017, Swiss National Bank.

Bäurle, G., and D. Kaufmann. 2018. "Measuring Exchange Rate, Price, and Output Dynamics at the Effective Lower Bound." *Oxford Bulletin of Economics and Statistics* 80 (6): 1243–66.

Bäurle, G., and E. Steiner. 2015. "How Do Individual Sectors Respond to Macroeconomic Shocks? A Structural Dynamic Factor Approach Applied to Swiss Data." *Swiss Journal of Economics and Statistics* 151 (3): 167–225.

Bauwens, L., M. Lubrano, and J.-F. Richard. 1999. *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press.

Bernanke, B., J. Boivin, and P. S. Eliasz. 2005. "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics* 120 (1): 387–422.

Bobeica, E., M. Ciccarelli, and I. Vansteenkiste. 2019. "The Link between Labor Cost and Price Inflation in the Euro Area." Working Paper No. 2235, European Central Bank.

Boivin, J., and M. Giannoni. 2006. "DSGE Models in a Data-Rich Environment." Working Paper No. 0332, National Bureau of Economic Research.

Boivin, J., M. P. Giannoni, and I. Mihov. 2009. "Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data." *American Economic Review* 99 (1): 350–84.

Canova, F. 2005. "The Transmission of US Shocks to Latin America." *Journal of Applied Econometrics* 20 (2): 229–51.

Carriero, A., T. Clark, and M. Marcellino. 2014. "No Arbitrage Priors, Drifting Volatilities, and the Term Structure of Interest Rates." Discussion Paper No. 9848, Centre for Economic Policy Research.

Ciccarelli, M., and B. Mojon. 2010. "Global Inflation." *Review of Economics and Statistics* 92 (3): 524–35.

Comunale, M., and D. Kunovac. 2017. "Exchange Rate Pass-Through in the Euro Area." Working Paper No. 2003, European Central Bank.

Dedola, L., G. Rivolta, and L. Stracca. 2017. "If the Fed Sneezes, Who Catches a Cold?" *Journal of International Economics* 108 (Supplement 1): S23–S41.

Eichenbaum, M., and C. L. Evans. 1993. "Some Empirical Evidence on the Effects of Monetary Policy Shocks on Exchange Rates." Working Paper No. 4271, National Bureau of Economic Research.

Fleer, R., B. Rudolf, and M. Zurlinden. 2016. "Price Change Dispersion and Time-Varying Pass-through to Consumer Prices." Working Paper No. 2016-17, Swiss National Bank.

Forbes, K. J., I. M. Hjortsoe, and T. Nenova. 2017. "Shocks versus Structure: Explaining Differences in Exchange Rate Pass-through across Countries and Time." Discussion Paper No. 50, External MPC Unit, Bank of England.

———. 2018. "The Shocks Matter: Improving Our Estimates of Exchange Rate Pass-through." *Journal of International Economics* 114 (September): 255–75.

Georgiadis, G. 2016. "Determinants of Global Spillovers from US Monetary Policy." *Journal of International Money and Finance* 67 (October): 41–61.

Geweke, J. 1992. "Bayesian Statistics." In *Bayesian Statistics*, Vol. 4, ed. J. Bernardo, J. Berger, A. Dawid, and A. Smith. Oxford University Press.

———. 2005. *Contemporary Bayesian Econometrics and Statistics.* John Wiley & Sons, Inc.

Gilchrist, S., and E. Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102 (4): 1692–1720.

Halka, A., and J. Kotlowski. 2017. "Global or Domestic? Which Shocks Drive Inflation in European Small Open Economies?" *Emerging Markets Finance and Trade* 53 (8): 1812–35.

Halka, A., and K. Szafranek. 2016. "Whose Inflation Is It Anyway? Inflation Spillovers Between the Euro Area and Small Open Economies." *Eastern European Economics* 54 (2): 109–32.

Jarocinski, M., and P. Karadi. 2018. "Deconstructing Monetary Policy Surprises: The Role of Information Shocks." Working Paper No. 2133, European Central Bank.

Jordan, T. J. 2015. "The Impact of International Spillovers on Inflation Dynamics and Independent Monetary Policy: The Swiss Experience." Paper presented at the 39th Economic Symposium on "Inflation Dynamics and Monetary Policy," Jackson Hole, Wyoming, August 28.

Karlsson, S. 2013. "Forecasting with Bayesian Vector Autoregressions." In *Handbook of Economic Forecasting*, Vol. 2B, ed. G. Elliott and A. Timmermann, 791–897 (chapter 15). Elsevier.

Kim, C.-J., and C. R. Nelson. 1999. *State-Space Models with Regime Switching.* The MIT Press.

Kim, S. 2001. "International Transmission of U.S. Monetary Policy Shocks: Evidence from VAR's." *Journal of Monetary Economics* 48 (2): 339–72.

Krippner, L. 2013. "A Tractable Framework for Zero Lower Bound Gaussian Term Structure Models." Discussion Paper No. 2013/02, Reserve Bank of New Zealand.

Liu, P., H. Mumtaz, and A. Theophilopoulou. 2014. "The Transmission of International Shocks to the UK. Estimates Based on a Time-Varying Factor Augmented VAR." *Journal of International Money and Finance* 46 (September): 1–15.

Maćkowiak, B. 2007. "External Shocks, U.S. Monetary Policy and Macroeconomic Fluctuations in Emerging Markets." *Journal of Monetary Economics* 54 (8): 2512–20.

Monacelli, T., and L. Sala. 2009. "The International Dimension of Inflation: Evidence from Disaggregated Consumer Price Data." *Journal of Money, Credit and Banking* 41 (s1): 101–20.

Mumtaz, H., and P. Surico. 2009. "The Transmission of International Shocks: A Factor-Augmented VAR Approach." *Journal of Money, Credit and Banking* 41 (s1): 71–100.

———. 2012. "Evolving International Inflation Dynamics: World and Country-Specific Factors." *Journal of the European Economic Association* 10 (4): 716–34.

Neely, C. J., and D. E. Rapach. 2011. "International Comovements in Inflation Rates and Country Characteristics." *Journal of International Money and Finance* 30 (7): 1471–90.

Neusser, K. 2009. *Zeitreihenanalyse in den Wirtschaftwissenschaften*, 3rd edition. Vieweg + Teubner Verlag.

Parker, M. 2016. "Global Inflation: The Role of Food, Housing and Energy Prices." Discussion Paper No. 2016/05, Reserve Bank of New Zealand.

Poirier, D. 1998. "Revising Beliefs in Nonidentified Models." *Econometric Theory* 14 (4): 483–509.

Potjagailo, G. 2017. "Spillover Effects from Euro Area Monetary Policy across Europe: A Factor-Augmented VAR Approach." *Journal of International Money and Finance* 72 (April): 127–47.

Primiceri, G. E. 2005. "Time Varying Structural Vector Autoregressions and Monetary Policy." *Review of Economic Studies* 72 (3): 821–52.

Rieth, M. 2015. "Can Central Banks Successfully Lean against Global Headwinds?" DIW Roundup No. 88, Deutsches Institut fur Wirtschaftsforschung.

Shambaugh, J. 2008. "A New Look at Pass-through." *Journal of International Money and Finance* 27 (4): 560–91.

Stock, J. H., and M. W. Watson. 2016. "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics." In *Handbook of*

Macroeconomics, Vol. 2A, ed. J. B. Taylor and H. Uhlig, 415–525 (chapter 8). Elsevier.

Stulz, J. 2007. "Exchange Rate Pass-through in Switzerland: Evidence from Vector Autoregressions." Economic Studies No. 2007-4, Swiss National Bank.

Uhlig, H. 2005. "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure." *Journal of Monetary Economics* 52 (2): 381–419.

Woodford, M. 2007. "Globalization and Monetary Control." Working Paper No. 13329, National Bureau of Economic Research.

Always Look on the Bright Side? Central Counterparties and Interbank Markets during the Financial Crisis*

Massimiliano Affinito and Matteo Piazza
Bank of Italy

This paper joins the literature on the growing use of central counterparties (CCPs) in interbank markets by analyzing a scarcely explored source of risk, namely that CCPs may provide riskier banks that are cut off from the bilateral segment of the market with an alternative channel to access interbank funds, thereby eluding peer monitoring and potentially increasing the risks borne by the financial system. We investigate this issue using monthly granular data on Italian banks from June 2004 to June 2013 and we find that during the global financial crisis riskier banks increased the share of their interbank funding obtained via CCPs, due to both the impact of general market uncertainty and the heightened attention to counterparty risk in the bilateral segment of the market. More tellingly, we show that, *for riskier banks only*, this increase was associated with a decline in the *duration* of bilateral relationships, indicating that longer-standing counterparties, typically those with more information, tended to withdraw from relationships with those banks. This suggests that during our sample period the pool of banks operating via CCPs may have become riskier, confirming, from a novel perspective, the importance of the policy efforts to ensure that CCPs have a proper risk-management framework.

JEL Codes: G01, G21, G23.

*The views expressed herein do not necessarily reflect the views of the Bank of Italy. We would like to thank for their comments two anonymous referees, Giorgio Albareto, Giuseppe Cappelletti, Riccardo De Bonis, Marc Deloof, Giovanni Guazzarotti, Gaetano Marseglia, Francesco Palazzo, Mario Pietrunti, Valerio Vacca, Barry Williams, and participants at the "Paris Financial Management Conference," "IFABS Barcelona Conference," "25th EFMA Conference," and "Infinity Conference." Usual disclaimers apply. Corresponding author (Affinito): massimiliano.affinito@bancaditalia.it.

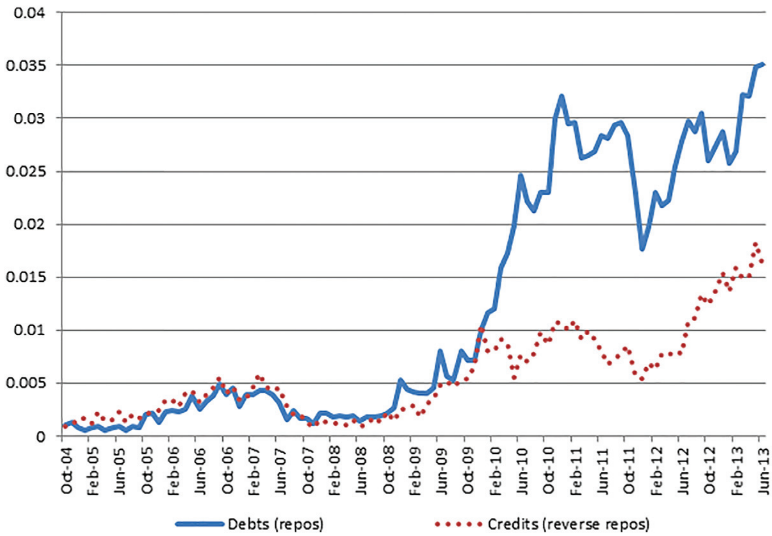
1. Introduction

A well-known feature of the global financial crisis has been its impact on interbank markets and the repercussions on the transmission mechanism of monetary policy and the whole financial system (e.g., Allen, Carletti, and Gale 2009; Brunnermeier 2009; Taylor and Williams 2009; Freixas, Martin, and Skeie 2011; Garcia-de-Andoain et al. 2016). In some countries, interbank activity did not freeze but showed, however, a remarkable change in its characteristics with a significant surge in secured lending, notably via central clearing counterparties (CCPs). While in the traditional interbank market transactions occur between pairs of banks (*bilateral* interbank market), may be secured or unsecured, and are nominative, in interbank transactions via CCPs lending and borrowing banks are no longer direct counterparties to each other, but all of them have the CCP as their counterparty. Moreover, exposures are secured (because they take place as repurchase agreements) and, at least in the European interbank market, anonymous (Mancini, Ranaldo, and Wrampelmeyer 2016).[1] CCPs are therefore third parties that stand between banks for the purpose of mitigating counterparty credit risk: according to some views, this transfer of counterparty risk to CCPs is precisely what makes acceptable the anonymity of (ultimate) counterparties which, in turn, allows for expanding the set of possible trades.[2]

[1]A CCP can be generally defined as an entity that interposes itself between (two or more) counterparties, becoming the buyer to every seller and the seller to every buyer. While in the bilateral transactions of interbank markets there is one contract, in the transactions involving a CCP there are more contracts: one between the buyer and the CCP and another one between the seller and the CCP. The CCP transforms the risk exposure among interbank counterparties into a risk exposure of each counterparty with the CCP. While repo activity via CCPs is in principle not limited to banks, in Europe, during our sample period "practically all counterparties involved in repos via CCPs have been euro area MFIs or non-euro area residents" (European Central Bank 2012). Note also that non-euro-area residents were basically banks, at least in the Italian case. This continued to be the case also in more recent periods.

[2]The reduction of counterparty risk in transactions via CCPs occurs through loss mutualization, high levels of collateralization, and multilateral netting. To manage the risk borne by the CCPs, members post initial margins and make contributions to the CCPs' default fund. CCPs are active in several markets in addition to repo transactions, notably in derivative markets. CCPs' functions for

Figure 1. Interbank Exposures through CCPs as Shares of Total Assets



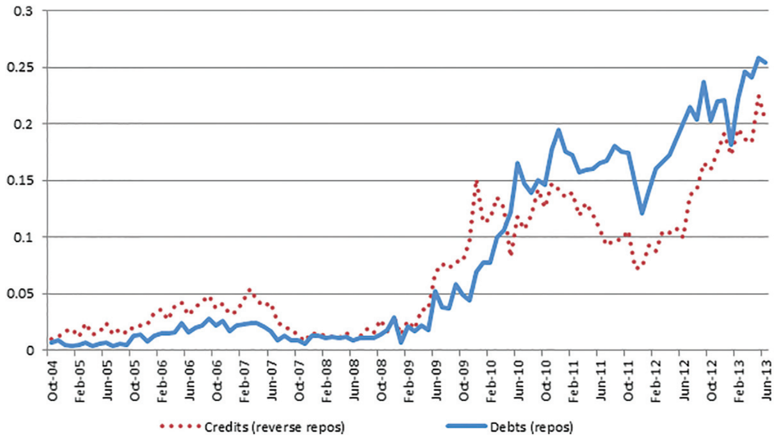
Source: Authors' computations on Bank of Italy prudential supervisory reports.

In Italy domestic banks stepped up their interbank funding via CCPs in a striking way since 2009, just after a key event of the global crisis (the Lehman Brothers default), with a sixfold increase of borrowed funds in less than four years, both as a share of total assets (figure 1) and as a share of total interbank exposures (figure 2). The ratio between the number of banks operating via CCPs and the total number of banks operating in the interbank markets also increased significantly (figure 3). This exponential increase mostly made up for the sharp decline in bilateral interbank funding with foreign banks (figure 4), in turn due to the euro-area financial fragmentation during the crisis (Banca d'Italia 2013a, 2013b; International Monetary Fund 2013, Garcia-de-Andoain et al. 2016).

---

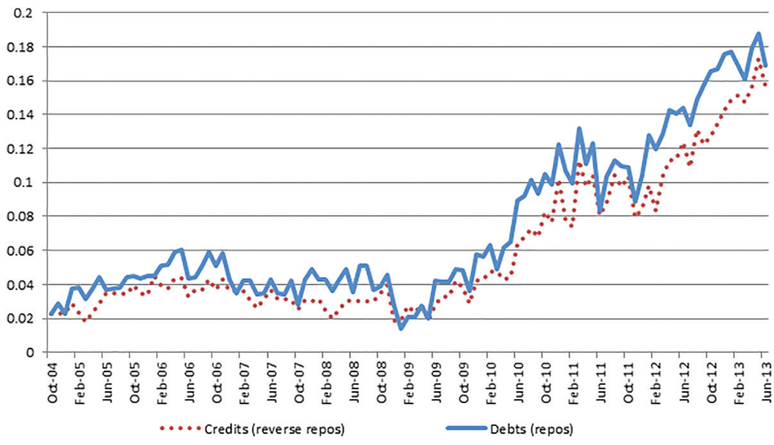
derivatives and for wholesale short-term funding present relevant differences and serve different economic purposes, as the former pursues a goal of insurance and the latter pursues a goal of funding. More institutional details are provided in section 2.

Figure 2. Interbank Exposures through CCPs as Shares of Total Extragroup Interbank Exposures



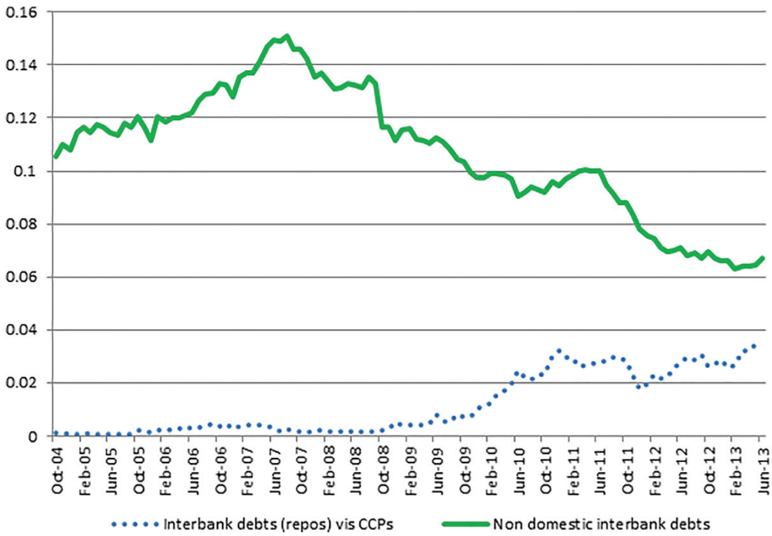
Source: Authors' computations on Bank of Italy prudential supervisory reports.

Figure 3. Number of Banks Operating via CCPs as a Share of the Total Number of Banks Operating in the (extragroup) Interbank Markets



Source: Authors' computations on Bank of Italy prudential supervisory reports.

Figure 4. Interbank Exposures through CCPs and Abroad as Shares of Total Assets



Source: Authors' computations on Bank of Italy prudential supervisory reports.

The Italian experience seems to lend support to the thesis that "jurisdictions that had CCPs for their repo markets in place before the crisis were relatively less affected than those that did not" (Chatterjee, Embree, and Youngman 2012). A number of papers (e.g., Cappelletti et al. 2011; Heider, Hoerova, and Holthausen 2015; Mancini, Ranaldo, and Wrampelmeyer 2016; Cappelletti and Guazzarotti 2017) refer to the benefits that a CCP may bring to the functioning of interbank transactions in periods of turmoil. A key aspect is that the increasing role of centrally cleared transactions addressed the general increase in uncertainty and risk aversion of lending banks during the financial crisis, thereby allowing interbank activity to keep playing its crucial role for monetary policy transmission and financial system functioning.

This larger role, however, may conceal a possible drawback in terms of financial stability, which has been scarcely explored so far. In fact, the increased use of CCPs could be concentrated among a pool of borrowers that would have been otherwise cut off from the bilateral segment of the interbank market due to their riskiness. In this case, the discipline exercised by peer monitoring in the bilateral interbank market could be lost, with a potential impact on financial stability.[3]

Investigating the possibility that the CCPs may be taking risks that would not be accepted on the bilateral segment of the market is relevant for three different reasons. First, as mentioned, it is important to analyze if the risk borne by the financial system may increase unintendedly, ceteris paribus, due to weakened peer monitoring. Second, an increase in the risk taken by CCPs may be potentially dangerous in light of their growing importance. Indeed, while the increased role of CCPs facilitates interbank activity and the related benefits, it may also increase the overall risk borne by the financial system, contributing to a general trend toward concentration of risks in CCPs that may turn them into institutions of systemic importance. In the words of policymakers, "CCP's criticality to the overall safety and soundness of the financial system means that authorities must take steps to ensure that CCPs do not themselves become a source of systemic risk" (Basel Committee on Banking Supervision et al. 2017). Third, the risk faced by a member of a CCP can increase, due to the mutualization of the losses, even if its own exposure does not change (Arnsdorf 2012). Therefore, a riskier pool of borrowers may reduce the incentives of sounder participants to centrally clear and potentially encourage a return to bilateral trading, losing the benefits of centrally cleared transactions.

For our analysis, we rely on a granular data set containing monthly data on all banks operating in Italy since 2004, when the Italian CCP started operating on the repo market, up until 2013. In addition to bank balance sheet variables, our data contain information on the identity of the parties and the duration of each interbank bilateral relationship, as customer relationships are quite relevant in the Italian interbank market (Affinito 2012). These data allow us (i) to identify banks that use CCPs, as well as

[3]Similar potential drawbacks of the use of CCPs may be found, for example, in Thompson (2010); Pirrong (2011); Stephens and Thompson (2011); Koeppl (2012); and Biais, Heider, and Hoerova (2013, 2016). Benefits and drawbacks of CCPs according to the literature are reviewed in sections 2 and 3.

when they started to do so; (ii) to connect choices in terms of participation in CCPs and intensity of their use to a large number of bank-specific characteristics and to bilateral interbank relationships; and (iii) to verify how the bilateral relationships were affected by the risk of borrowing banks and how this may have affected the use of CCPs.

Our empirical analysis runs in two steps. In the first step, we study the determinants of the share of interbank transactions conducted via CCPs, and we show that both general uncertainty and individual risk were relevant in determining the recourse to CCPs.[4] Taken alone, however, the fact that individual bank risk was positively influencing the recourse to CCPs is not sufficient to conclude that CCPs were taking up risks that were shunned by bilateral counterparties. This leads us to our second step, where we take advantage of the granular nature of our data to infer, from the actual behavior of bilateral interbank counterparts, whether the use of centrally cleared transactions was associated with a loss of their usual interbank bilateral counterparties. In more detail, we examine the relation between variations in the use of CCPs and the weighted average duration of all bilateral interbank relationships of each borrowing bank. The hypothesis we test is whether, *for riskier banks*, an increase in the share of CCP transactions is significantly associated with a decrease in the duration of bilateral relationships, while for less risky banks the relationship is positive or nil. The underlying idea is that—due to the informational advantages of long-term relationships compared with short-term ones, a well-established result in the literature (reviewed in section 3)—long-standing counterparts should be more able to discriminate between banks and to preserve bilateral relationships with the less risky ones. This implies that older interbank relationships are affected relatively more than newer ones by bank-specific characteristics and risks.

In other words, for riskier banks (those in the upper deciles of the distribution of our risk indicators), increases in the share of CCP transactions and decreases in the duration of bilateral relationship

[4]The *participation* of riskier banks in CCPs became instead less likely during the crisis, possibly due to the increased costs to use CCPs as a consequence of the stricter risk control frameworks gradually adopted. The increased use of CCPs in our sample period is mostly explained, however, by the intensive margin.

would be a sign of the drying-up of interbank funding from longer-standing (i.e., more informed) counterparts in the bilateral segment of the interbank market and of its replacement with transactions via CCPs. Instead, less risky banks may have no need at all to recur to CCPs, as they can keep existing relationships with long-standing counterparts: if any, they could use CCPs to replace newer counterparts that may be less able to recognize the low risk of these banks. This means that a null or positive relationship between increases in CCP use and duration could be expected for less risky banks. Such finding would suggest that the discipline exercised by interbank peer monitoring was in fact relaxed by the availability of anonymous CCP transactions.

Our empirical approach also allows us to disentangle our hypothesis that riskier banks may prefer anonymous trades to elude peer monitoring, with a possible detrimental effect on financial stability, from the alternative hypothesis that the shift to transactions via CCPs is simply driven by the desire to avoid a stigma.[5] In our framework, the latter hypothesis would imply no differential impact on *existing, long-standing* relationships while, to the contrary, in our hypothesis long-standing counterparts would be those better placed to first exercise peer monitoring and refrain from transactions with the riskiest counterparties.

Our results show that different banks may have indeed different motivations behind their recourse to CCPs. We show that, *for riskier banks only*, the increase in the use of CCPs was associated with a decline in the *duration* of bilateral relationships, indicating that longer-standing counterparties, typically those with more information, tended to withdraw from relationships with those riskier banks. This is not the case for less risky banks. The policy implication of our results supports, from a novel perspective, the ongoing effort to ensure that CCPs put in place adequate risk control frameworks, an essential corollary to the growing importance of CCPs promoted by financial reforms in the aftermath of the global financial crisis, with the aim of improving market transparency, mitigating systemic risk, and preventing market abuse (Committee on Payment

[5]This stigma would be related to the fact that, in a period of uncertainty, interbank market participants could identify additional borrowing in that market as a sign of financial difficulties.

and Settlement Systems (CPSS) and International Organization of Securities Commissions (IOSCO) 2012; Committee on Payments and Market Infrastructures (CPMI) and IOSCO 2016; Basel Committee on Banking Supervision et al. 2017).[6]

The rest of the paper illustrates in detail the features of our analysis, starting in section 2 with a description of some institutional background on the development of CCPs. Section 3 summarizes the literature on benefits and risks of CCPs. Sections 4–7 describe respectively the data, our empirical strategy, the main results, and the robustness checks. Section 8 concludes.

2. Institutional Background

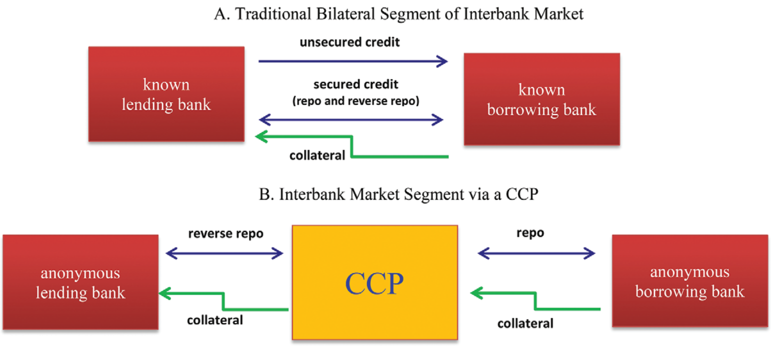
The use of CCPs to clear interbank repurchase agreements has strongly increased since the financial crisis. Repurchase agreements with CCPs quickly became a sizable alternative to bilateral transactions, reaching an outstanding amount of almost 300 billion in the euro area already in July 2012 "as repo operations through CCPs provide better protection against counterparty risk than bilateral repo transactions" (ECB 2012). In addition to reducing counterparty risk, recourse to CCPs may bring several other benefits, including saving collateral, through greater netting efficiency, and promoting transparency.[7]

The typical structure of interbank transactions via CCPs in the euro area can be broadly described as follows (figure 5): (i) the borrowing bank enters into a repurchase agreement with the CCP, borrowing the required amount and providing collateral; (ii) the lending bank enters into a reverse repo with the CCP; and (iii) the CCP acts

[6]Recourse to central clearing has been strongly promoted, in the aftermath of the global financial crisis, for over-the-counter (OTC) derivatives, starting with the work of the Financial Stability Board (FSB, formerly Financial Stability Forum, FSF) in 2008 and the ensuing G-20 commitments in Pittsburgh in 2009 (FSF 2008; FSB 2013). As of mid-2017, 17 of 24 FSB member jurisdictions have a legislative framework in force for mandatory central clearing requirements (FSB 2017).

[7]See, for example, ECB (2007), FSF (2008), Cecchetti, Gyntelberg, and Hollanders (2009), Leitner (2012), Biais, Heider, and Hoerova (2012, 2016), FSB (2013, 2017), Acharya and Bisin (2014), Loon and Zhong (2014), Duffie, Scheicher, and Vuillemey (2015); Baklanova, Dalton, and Tompaidis (2017).

Figure 5. Structure of Two Segments of the Interbank Market: Bilateral and via a CCP



Notes: The figure shows schematically the structure of the interbank market: Panel A shows the typical structure of the bilateral segment, and panel B shows the typical structure of the segment via a CCP. The traditional interbank bilateral transactions occur between pairs of banks, are nominative, and may be secured or unsecured. Interbank transactions via CCPs occur usually through repos (and are thus secured), and in Europe they are typically anonymous. The structure of the segment via a CCP typically works as follows: (i) the borrowing bank enters into a repurchase agreement with the CCP, borrowing the required amount and providing collateral; (ii) the lending bank enters into a reverse repo with the CCP; (iii) the CCP acts as the direct counterparty to the seller and the buyer, thus assuming the risk of borrower default, and manages the transaction and the collateral. In addition, collateral management is highly standardized in terms of profiling and margining, which enhances transparency, and the administrative burden for borrower and lender is significantly lower than in a bilateral repo.

as the direct counterparty to the seller and the buyer, thus assuming the risk of borrower default, and manages the transaction and the collateral.[8] In addition, collateral management is highly standardized in terms of profiling and margining, which enhances transparency, and the administrative burden for borrower and lender is significantly lower than in a bilateral repo.

[8]If lending and borrowing banks or one of them are not clearing members of the CCPs, we have the so-called client-clearing models, where a counterparty is not itself a clearing member but accesses a CCP via a third party who is a clearing member. It results in the creation of a distinct legal contract between the clearing member and its client (a back-to-back contract) in addition to the legal contract between the CCP and the clearing member. For more details, see European Securities and Markets Authority (2017).

In Italy only one central counterparty is authorized: Cassa di compensazione e garanzia S.p.A. (CC&G).[9] Italian intermediaries can however decide to (also) adhere to foreign CCPs, and symmetrically CC&G accepts foreign intermediaries as clearing members. Moreover, thanks to interoperability arrangements, intermediaries can belong either to CC&G or to the French central counterparty LCH.Clearnet SA, as if the two partner institutions formed a single virtual central counterparty.[10] In the Italian case, participants in this market were basically all banks, and this was broadly the case in other countries in the euro area.[11]

The use of CCPs may bring a number of benefits (e.g., Hardouvelis and Peristiani 1992; Borio 2004; ECB 2007; FSF 2008; Cecchetti, Gyntelberg, and Hollanders 2009; and FSB 2015, and the literature reviewed in the next section). First, CCPs are supposed to reduce counterparty risk, making the entire financial system safer, by means of mutualization of credit risk (sharing it among all participants and insuring idiosyncratic risks) and the reduction of information asymmetries (allowing participants to trade with only one counterparty). Second, as counterparties of all trades, CCPs can net multilaterally, and, thanks to the multilateral netting, CCPs can increase the amount of available collateral. Third, by facilitating data collection, CCPs may improve market transparency and help a correct assessment of outstanding risks.

On the other hand, the rising importance of CCPs may be associated with a number of side effects, such as a concentration of

[9]At its outset CC&G dealt only with financial derivatives, but over time its activities expanded to include shares (on a compulsory basis), Italian government securities (on an optional basis), and a broad range of trading platforms and financial instruments, including the collateralized interbank deposit market.

[10]As mentioned, in the European interbank repo market the majority of repos are traded anonymously via CCPs. Furthermore, the interoperability agreements between the Italian and French CCP imply that parties in the repo transaction may carry out their side of the transaction with a different CCP, adding a further reason why (ultimate) parties in the repos may be unaware of the identity of their counterparties in the transaction and accordingly not exercise any monitoring on them.

[11]For this reason, the ECB decided in 2012 to exclude, retroactively from June 2010, repos with CCPs from the reference monetary aggregate M3, considering de facto this activity as part of the interbank activity.

risks that may assume systemic importance and potential contagion effects (in terms of losses and liquidity shortfalls). Typically, CCPs adopt a multi-level system of safeguards to protect themselves and their members from losses. First, clearing members have to post an "initial margin," which is a form of collateral initially collected by the CCP and retained in the event of default. The initial margin is commensurate with the value and risk of contracts, and it is typically delivered either in cash or in the form of securities that have high credit quality and can easily be sold. Second, a "variation margin" is charged or credited daily to clearing members to cover any mark-to-market changes in their portfolio. This means that CCPs control daily the revaluation of open positions at current market prices and calculate any gains or losses that have to be paid or received each day. In periods with high volatility, positions may even be marked to market intradaily. CCP risk control usually entails stricter rules on the posting of collateral than those used in bilateral markets.[12] Third, CCPs have an equity buffer provided by shareholders as well as their own assets. Fourth, every member contributes to the clearing house "default fund," which acts as a mutualized insurance for uncollateralized losses. Fifth, each clearing member is usually committed to providing further funds if necessary (recovery procedure). The so-called default waterfall refers to the order in which these resources are used. Typically, the waterfall envisages first the use of the available resources of the defaulting member (initial margins and then its default fund contribution). Next, the CCPs' capital is used and then the default fund contributions of surviving members. Further down, other rules may be envisaged to face the situation, either as part of the waterfall or as a part of so-called end-of-the-waterfall situations, following the exhaustion of all the safeguards contemplated in the default waterfall (for further details, see CPSS-IOSCO 2012; CPMI-IOSCO 2014, 2016).

Significant efforts have been deployed to ensure an improved resilience of CCPs and, according to some views, they now employ "risk management methods that do not exist to the same extent in

[12]Rules establish what assets are allowed as collateral, how much of a haircut should be given to specific assets in determining their value as collateral, and how often margin calls should take place.

the bilateral world" (Cœure 2014). However, there were also dissenting views, at least in the initial phase of CCPs' activity.[13]

Whatever the judgment about the CCPs' risk control frameworks, as long as the resources provided by the *defaulting member* (either margins or contributions to the default fund) are enough to compensate the lender, centrally cleared transactions are not different in substance from secured bilateral transactions. However, if and once these specific resources are no longer sufficient, the quality of the pool of borrowers starts to matter, and this is what motivates our paper.

3. Related Literature

Our work is related to a wide literature that explores benefits and risks of CCPs, usually in comparison with a situation where only the bilateral market exists. On benefits, Bernanke (1990) highlighted two positive roles of a clearinghouse: reducing transaction costs of consummating agreed-upon trades (analogous to a bank that clears checks) and standardizing contracts by setting terms and format and guaranteeing performance to both sides of trade (analogous to an insurance company). Koeppl and Monnet (2010) show that the benefit of centralized clearing is in the mutualization of counterparty default risk. Biais, Heider, and Hoerova (2012) find that an appropriately designed centralized clearing mechanism enables trading parties to benefit from the mutualization of (the idiosyncratic component of) risk. Loon and Zhong (2014) use data on voluntarily cleared CDS contracts to document a reduction of both counterparty and systemic risk. Another benefit pointed by the literature is the saving of collateral: a number of empirical works have assessed

[13]For example, Pirrong (2011) claimed that "CCP margins typically depend on product risk characteristics, rather than the creditworthiness of the clearing member" and that "margins that do not vary meaningfully [. . .] underprice the risks of less creditworthy firms and overprice the risks of more creditworthy firms, which tends to lead the former to trade too much, and the latter too little." Furthermore, he also adds that CCPs "monitor the creditworthiness of their members, but this monitoring is largely based on standards and information (e.g., accounting statements) that do not reflect variations in creditworthiness among members in a discriminating way" and that "the CCP typically does not impose differential capital or margin requirements on members that meet a certain creditworthiness threshold."

changes in collateral demand due to mandatory central clearing (Heller and Vause 2012; Sidanius and Zikes 2012; Duffie, Scheicher, and Vuillemey 2015) and conclude that mandatory central clearing substantially lowers systemwide collateral demand, unless there is significant proliferation of CCPs. According to Cappelletti and Guazzarotti (2017), the benefit of CCPs is that the perception of a substantial stigma effect may lead borrowers to prefer anonymous to transparent markets for interbank transactions: having in place anonymous trades via CCPs could be therefore welfare increasing, as it reduces some harmful effect of imperfect information.

The literature more closely related to our paper, however, is the sizable work focusing on moral hazard issues. The central clearing mechanism may generate two types of moral hazards. The first one is the moral hazard of participants, which derives from the mutualization of losses, that weakens or cancel participants' incentives to find and monitor solid counterparties, in comparison with what happens in the bilateral market. The second type of moral hazard is due to the CCPs themselves, which counting on their systemic relevance (i.e., on being too big or too interconnected to fail) could fail to properly monitor counterparts (Stephens and Thompson 2011; Jones and Perignon 2013; Biais, Heider, and Hoerova 2016). Pirrong (2011) and Koeppl (2012) both conclude that use of CCPs is not welfare improving relative to bilateral transactions because it can lead to an inefficient increase in the risk of contracting with a bad protection seller and it can weaken market discipline. Jones and Perignon (2013) show that, in order to cope with the moral hazard problems in the clearing mechanism, an incentive-compatible system must be put in place. Biais, Heider, and Hoerova (2013, 2016) point out that, in order to overcome both moral hazard issues, the CCP has to limit the amount of insurance it provides to clearing members so as to give them incentives to seek out sound counterparties that enhance the risk-bearing capacity of the CCP. Hansen and Moore (2016) show that mandatory central clearing is welfare improving thanks to the mutualization of counterparty credit risk, but only if initial margin requirements are set optimally, due to the tradeoff between the default insurance that a CCP provides and the incentive for market participants to trade too much when default losses are mutualized through the CCPs' default fund.

Finally, our work relates to the literature on peer monitoring among banks, which points out that interbank borrowing may serve, through peer monitoring, to monitor and discipline borrowing banks. This discipline effect may work through three channels. First, banks are better informed on the standing of their peers than retail depositors and they have more incentives to monitor them, as exemplified by the absence of deposit insurance for interbank deposits. Second, this literature applies to the relationships among banks the same underlying concepts developed in the literature on the relationships between banks and firms. In particular, it shows that a closer relationship among banks allows lending banks to obtain more information about the borrowing bank because it increases lenders' incentives to gather information and monitor borrowers. Third, the interbank funding assumes a disciplining role because, while retail depositors tend to show a high degree of inertia in their behavior, interbank exposures are typically at very short maturities and lending banks may promptly decide not to roll them over. This literature includes both theoretical and empirical works (e.g., Calomiris and Kahn 1991; Rochet and Tirole 1996; Furfine 2001; Huang and Ratnovski 2008; King 2008; Cocco, Gomes, and Martins 2009; Angelini, Nobili, and Picillo 2011; Affinito 2012; Distinguin, Roulet, and Tarazi 2013).[14]

4. Data

Our sample period extends from June 2004, when centrally cleared repo transactions started in Italy, to June 2013. With the exception of the measures of uncertainty and the rating scores, all our data are drawn from the Bank of Italy prudential supervisory reports. These data include granular information on interbank transactions with both domestic and foreign banks. Since liquidity management

[14]Some doubts have been raised (for example, by Duffie 2019) on the effectiveness of market discipline as opposed to the use of a stricter regulation and supervision. We share the view that interbank peer monitoring, like other forms of market discipline, cannot be considered a substitute for effective supervision. Our focus is rather in stressing that peer monitoring may be a (timely) complement to supervision and that eluding it may contribute to create additional financial stability risks. We thank an anonymous referee for helping us to clarify the point.

is typically centralized at the group level, data of intermediaries that are part of a banking group are consolidated at each point in time (considering the group as a single entity) and we do not consider intragroup transactions.[15] This is done for all variables in our data set, and in the paper we refer to both banking groups and stand-alone banks in our sample as "banks."[16] While data were available for each resident bank, we excluded from our analysis cooperative banks because they are typically very small and tend to manage their liquidity needs and surpluses through a dedicated intermediary which acts as a liquidity hub. Our final sample is an unbalanced panel including about 200 banks on average in each of our 109 monthly periods. The banks in our sample represent on average about 90 percent of the total assets of the Italian banking system along our sample period. Tables 1–3 describe our explanatory variables and provide summary statistics.

We use end-of-month outstanding amounts for all types of interbank exposures. Common to other contributions in the literature (e.g., Furfine 2004, 2009; King 2008; Cocco, Gomes, and Martins 2009; Dinger and von Hagen 2009; Affinito 2013), we do not have data on prices for over-the-counter transactions, which are very relevant in the interbank market. While this is clearly a limit, it is important to remark that, according to the majority of the accounts of developments during the financial crisis, prices were basically moving in response to changes in quantities.[17] The use of end-of-month outstanding amounts is likewise explained by data availability. In

[15]Intragroup transactions tend to fit into a group-specific scheme and are likely to be decided by the parent bank (e.g., Houston, James, and Marcus 1997; de Haas and van Lelyveld 2010). In order to eliminate the intragroup exposures, we used information on the identity of each counterpart and its group. For the banks that changed group during our sample period, we traced the current group of affiliation in each period and analyzed their effective extragroup relationships in each period.

[16]We consider all *extragroup* secured and unsecured transactions executed both on regulated and over-the-counter markets.

[17]The extreme example were transactions on the e-MID, the electronic platform for unsecured interbank activity in Italy, where exchanges dramatically dropped, making the quoted prices basically non-informative. Also note that even with data on interest rates, it would not be easy to assess all the different aspects directly or indirectly involved in the relative cost comparison between CCPs and bilateral transactions: haircuts, cost of collateral, contributions to CCPs default funds, etc.

Table 1. Definition of Variables and Summary Statistics

Variables' Notation	Variables' Content	Definition	N	Mean	SD	Min.	p50	Max.
SH_{jt}	CCP Debts	Interbank debts through CCPs / Total interbank debts	15,279	0.02	0.10	0.00	0.00	1.00
UNC_t	Uncertainty	Ratio between density estimated using historical data from the benchmark index for the Italian stock exchange and the risk-neutral density derived from the options on the index	15,279	0.84	0.30	0.21	0.86	1.48
$Risk_{jt}$	Bad Loans	Bad loans / Total loans	15,279	0.03	0.03	0.00	0.02	0.12
	Rating	Rating agency scores	15,279	9.87	2.47	2.00	11.00	11.00
	Banks without Rating (0–1)	Banks without rating (0–1)	15,279	0.82	0.38	0.00	1.00	1.00
Bilateral Relationships$_{jt}$	Interbank Counterparties Concentration ICC_{jt}	Log (degree of concentration of interbank debts)	15,279	0.44	0.36	0.00	0.34	1.00
	Interbank Relationship Duration IRD_{jt}	Weighted average length of all interbank borrowing relationships	15,279	2.80	1.50	0.00	3.28	5.14
	Foreign Interbank Debts	Interbank debts from abroad / Total interbank debts	15,279	0.20	0.32	0.00	0.02	1.00
Bilateral Network Centrality$_{jt}$	Interbank Network Degree	The number of interbank connections of each bank	15,279	2.60	1.04	0.69	2.40	6.57
	Interbank Network Betweeness	An index of interbank centrality of each bank that seizes the banks that each bank has to go through in order to reach another bank in the minimum number of hops	15,279	3.25	3.12	0.00	2.82	12.41
	Interbank Network Closeness	An index of interbank centrality of each bank that captures the length of shortest path to all others	15,279	0.36	0.04	0.25	0.35	0.60
Control Variables KR_{jt}	Retail Fundraising	Total retail deposits and bonds / Total assets	15,279	0.47	0.30	0.00	0.57	1.00
	Central Bank Loans	Total loans from central bank / Total assets	15,279	0.02	0.05	0.00	0.00	0.36
	Tier 1 Capital	Tier 1 / Risk-weighted assets	11,606	0.17	0.13	0.02	0.13	1.00
	ROE	Net profits / Capital	15,279	0.06	0.17	−0.89	0.05	0.90
	Size	Log (Total Assets)	15,279	7.79	1.96	1.95	7.72	13.67
	Loans to Private Sector	Loans to private sector / Total assets	15,279	0.57	0.24	0.00	0.63	0.99
	Portfolio of Government Bonds	Portfolio of government bonds / Total assets	15,279	0.06	0.09	0.00	0.03	0.86

Table 2. Intensive and Extensive Margins of Interbank Exposures through CCPs (millions of euros and as a share of total assets)

	Total		Intensive		Extensive	
		%		%		%
2009–2008	10.955	*0.31*	10.923	*0.31*	32	*0.00*
2010–2009	52.841	*1.53*	46.741	*1.36*	6.100	*0.18*
2011–2010	20.602	*0.59*	20.209	*0.58*	393	*0.01*
2012–2011	−885	*−0.02*	*−4.033*	*−0.11*	3.148	*0.09*
2013–2012	17.246	*0.45*	13.726	*0.36*	3.521	*0.09*
2013–2008	**100.759**	***2.64***	**87.564**	***2.29***	**13.194**	***0.35***

Notes: The extensive margin is computed as the sum of the current-year average interbank exposure through CCPs of each bank whose previous-year average interbank exposure through CCPs is equal to zero. The intensive margin is computed as the sum of differences of the current- and previous-year average interbank exposures of each bank whose previous-year average interbank exposure through CCPs is greater than zero.

fact, micro bank-by-bank data with the details of our data set do not exist with a higher frequency. However, it is worth noticing that, although interbank activity is usually at very short maturities, the persistence of exposures and positions is very high, even toward specific counterparties (Affinito 2012, 2013; Affinito and Pozzolo 2017).

5. Outline of the Empirical Analysis

Our analysis focuses on borrowing banks as a possible source of risk for CCPs. In Italy banks have typically been net borrowers on centrally cleared repo transactions (figures 1 and 2), since the ultimate lenders are mostly foreign intermediaries.[18]

[18]Based on available evidence, Italian borrowers—and foreign lenders—operating via CCP were both almost exclusively banks, as discussed in section 2.

Table 3. Correlations among Variables

	CCP Debts	Foreign Interbank Debts	Borrowing IRD	Borrowing ICC	Interbank Network Betweeness Centrality	Interbank Network Degree	Interbank Network Closeness Centrality	Size	Loans to Private Sector
CCP Debts	1								
Foreign Interbank Debts	–0.1***	1							
Borrowing IRD	0.1***	–0.3***	1						
Borrowing ICC	–0.09***	–0.069***	0.09***	1					
Interbank Network Betweenness Centrality	0.2***	–0.3***	0.4***	–0.3***	1				
Interbank Network Degree	0.2***	–0.3***	0.4***	–0.3***	0.8885***	1			
Interbank Network Closeness Centrality	0.2***	–0.1***	0.3***	–0.2***	0.7***	0.82***	1		
Size	0.2***	–0.02*	0.4***	–0.2***	0.62***	0.7***	0.61***	1	
Loans to Private Sector	–0.1***	0.1***	0.03***	0.02*	–0.2***	–0.2***	–0.1***	–0.09***	1
Portfolio of Government Bonds	0.1***	–0.3***	0.07***	0.085***	0.07***	0.03***	–0.0066	–0.0064***	–0.3***
Retail Fundraising	0.04***	–0.7***	0.2***	0.0602***	0.1***	0.1***	–0.01	0.03***	0.009
Central Bank Loans	0.1***	–0.1***	0.1***	–0.02	0.2***	0.1***	0.2***	0.1***	–0.1***
Bad Loans	0.1***	–0.3***	0.2***	0.02	0.1***	0.2***	0.1***	0.1***	0.2***
Tier 1 Capital	0.067***	–0.082***	–0.2***	0.1***	–0.1***	–0.2***	–0.1***	–0.2***	–0.3***
ROE	–0.082***	0.2***	–0.002	–0.081	–0.009**	0.01	0.02***	0.089***	0.1***
Rating	–0.1***	0.2***	–0.2***	0.2***	–0.5***	–0.602***	–0.5***	–0.5***	0.069***
Banks without Rating (0–1)	–0.1***	0.2***	–0.2***	0.2***	–0.5***	–0.602***	–0.5***	–0.5***	0.07***
Uncertainty	0.05***	0.01	–0.07*	0.01*	0.0082***	–0.061***	0.04***	–0.02	0.03***

(*continued*)

Table 3. (Continued)

	Portfolio of Government Bonds	Retail Fundraising	Central Bank Loans	Bad Loans	Tier 1 Capital	ROE	Rating	Banks without Rating (0–1)	Uncertainty
Portfolio of Government Bonds	1								
Retail Fundraising	0.3	1							
Central Bank Loans	0.3***	–0.005	1						
Bad Loans	0.1***	0.3***	0.1***	1					
Tier 1 Capital	0.1***	–0.2***	0.04***	–0.1***	1				
ROE	–0.0861***	–0.1***	–0.07***	–0.0882***	–0.2***	1			
Rating	–0.02*	–0.1***	–0.09***	–0.2***	0.1***	–0.005	1		
Bans without Rating (0–1)	–0.01	–0.01***	–0.09***	–0.2***	0.1***	0.0003	0.09***	1	
Uncertainty	0.02	–0.02	0.1***	0.062***	0.05***	–0.09***	0.0669***	0.05***	1

Note: ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

5.1 First Step: Determinants of the Use of CCPs

We start by exploring the determinants of the use of centrally cleared transactions through the following equation:

$$\begin{aligned} SH_{jt} = {} & \alpha_0 UNC_t + \beta_0 Risk_{jt} + \gamma_0 Bilateral_{jt} + \alpha_1 UNC_t * CR1_t \\ & + \alpha_2 UNC_t * CR2_t + \beta_1 Risk_{jt} * CR1_t + \beta_2 Risk_{jt} * CR2_t \\ & + \gamma_1 Bilateral_{jt} * CR1_t + \gamma_2 Bilateral_{jt} * CR2_t + \delta' KR_{jt} \\ & + \zeta' b_j + \eta' p_t + \epsilon_{jt}, \end{aligned} \tag{1}$$

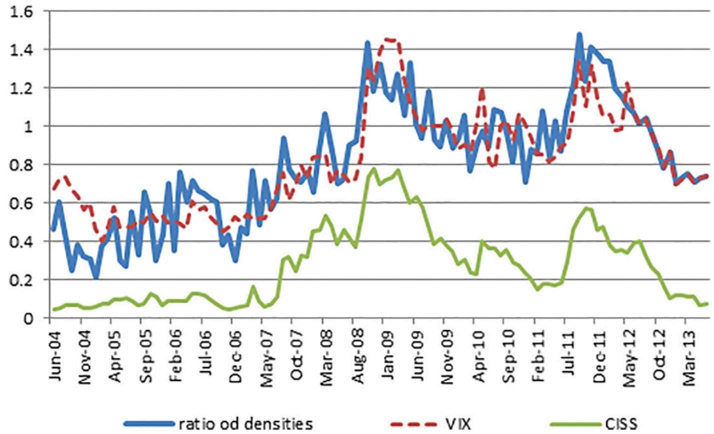
where SH_{jt} is the share of bank borrowing via CCPs over total interbank borrowing (including bilateral transactions, secured and unsecured, domestic and abroad) of bank j at time t, in each month from June 2004 to June 2013.

Explanatory variables are grouped in four categories (table 1), described in more detail below: (i) general market uncertainty and risk aversion (UNC_t); (ii) individual risk of borrowing banks ($Risk_{jt}$); (iii) banks' relationships in the bilateral segment of interbank market ($Bilateral_{jt}$); and (iv) control variables (KR_{jt}). Bank-specific dummies b_j are also included to account for unobservable structural bank characteristics. Time fixed effects p_t and dummies for the crisis periods (CR) are also included.

UNC_t accounts for the role of general market uncertainty and risk aversion, and it is proxied by three different measures, used alternatively for robustness purposes. Our main measure is the ratio between the density estimated using historical data from the benchmark index for the Italian stock exchange and the risk-neutral density derived from the options on the index.[19] We also use alternative measures of UNC_t, such as VSTOXX and CISS (figure 6), as described in more detail in the section on robustness checks (section 7).

[19]The methodology underlying this proxy for risk aversion is described in Jackwerth (2000) and implemented by Tarashev, Tsatsaronis, and Karampatos (2003). As we had this variable available only up to May 2012, we forecast it for the last months in our sample period by using the VSTOXX, the index based on Euro Stoxx 50 options prices according to VIX methodology, which is closely correlated with the first indicator for the overlapping periods. Results do not change with respect to those obtained using data only until May 2012.

Figure 6. Alternative Measures of General Market Uncertainty and Risk Aversion



Sources: For the ratio of densities: Jackwerth (2000) and Tarashev, Tsatsaronis, and Karampatos (2003); for VIX: VSTOXX, the index based on Euro Stoxx 50 options prices according to VIX methodology; for CISS: Holló, Kremer, and Lo Duca (2012).

$Risk_{jt}$ represents our proxies for the individual risk of the borrowing banks. Our default measure is the *Bad Loans* ratio, which is a standard measure of banks' risk, available for each bank.[20] This variable, while available in the supervisory returns used in this analysis, is not known by counterparties on a continuous-time basis (as it is

[20] According to Italian regulation in force during our sample period, nonperforming loans were classified according to four categories: (i) bad loans: exposures to an insolvent counterparty (even if insolvency is not legally ascertained) or in equivalent situations, regardless of any loss estimate made by the bank and irrespective of any possible collateral or guarantee; (ii) substandard loans: exposures to counterparty facing temporary difficulties—defined on the basis of objective factors—expected to be overcome within a reasonable period of time; (iii) restructured loans: exposures in which a pool of banks or an individual bank, as a result of the deterioration of the borrowers' financial situation, agree to change the original conditions (rescheduling deadlines; reduction of interest rate), giving rise to a loss; (iv) past-due loans: exposures other than those classified as bad loans, substandard, or restructured exposure that are past due for more than 90 days on a continuous basis. Our variable, therefore, focuses on the most impaired part of the loan portfolio of a bank, and it is computed as the ratio of bad loans over total loans.

usually published only in the financial statements), and it may be influenced by classification policies. However, it generally provides a fair approximation of the actual risk of each bank also considering that for Italian banks, credit risk typically represents by far the largest source of risk. As an alternative, we also use a pair of variables that capture the point of view of rating agencies and are described in the section on robustness checks.

The third set of regressors, $Bilateral_{jt}$, looks at how the situation and the role of each bank in the bilateral segment of the interbank market affects the choice of recurring to CCPs. We include here two subsets of variables. The first subset, $Bilateral\ Relationships_{jt}$, estimates the effect of interbank bilateral customer relationships on the use of CCPs with two alternative variables which take advantage of our granular information on the identity of each counterpart (domestic and foreign) and the related gross bilateral positions and measure respectively the *strength* and *length* of relationships of each bank in the bilateral interbank market.

The first variable, *Interbank Counterparties Concentration*, ICC_{jt}, measures the degree of concentration of bilateral interbank borrowing of a bank j in period t. The second variable, *Interbank Relationship Duration*, IRD_{jt}, measures in each period the weighted average time length of all interbank relationships of each bank and is a weighted average to take into account the size of each exposure in addition to its duration.

The rationale for the two variables is in the vast literature that documents the advantages of relationship lending. According to this literature, a close relationship allows lenders to obtain more information about the borrower because it increases lenders' incentives to gather information and monitor borrowers. Similar arguments may be applied also to the relations between two banks (see, for example, Cocco, Gomes, and Martins 2009; Affinito 2012). Both our measures of the intensity of $Bilateral\ Relationships_{jt}$ are inspired by that literature, which measures the strength of the customer relationships either through the concentration of loans or through their duration.[21]

[21] For example, as far as the variable ICC_{jt} is concerned, Petersen and Rajan (1994) and Angelini, Di Salvo, and Ferri (1998) find that firms with more concentrated borrowing have better access to credit. Gobbi and Sette (2014) show that

ICC_{jt} is computed as a standard Herfindahl index: $ICC_{jt} = \sum_{i=1}^{N} s_{ijt}^2$, where s_{ijt} is the share of counterpart bank i as lending counterpart of bank j in time t, and N is the total number of banks lending to bank j in time t. This variable, which ranges between 0 and 1, provides a measure of the strength of interbank relationships of each bank j: higher values indicate that a bank tends to hold more exclusive relationships with few counterparts.

IRD_{jt} is computed as follows: $IRD_{jt} = \sum_{i=1}^{N} s_{ijt} * d_{ijt}$, where j, i, t, N, and s_{ijt} are defined as before and d_{ijt} counts in each period t the integer number of consecutive months elapsed since the start of an interbank relationship between bank j and each counterpart bank i. In order to minimize censoring, we collect data for this variable back to June 1998 (i.e., 72 monthly periods before the start of our sample period). The maximum value for the integer number d_{ijt} is accordingly equal to 181 in the last period of our sample if the pair *(j,i)* had a interbank relationship in any period, allowing for one month of interruption as a maximum.[22]

We also include foreign extragroup interbank funding (as a ratio to total interbank funding) as an explanatory variable, as the financial crisis triggered a significant retrenchment of foreign interbank bilateral funding (figure 4).

A second subset of variables, *Bilateral Network Centrality*$_{jt}$, measures the centrality of each bank in the network of bilateral links of the interbank market. We use three standard measures of centrality in the network literature which have been already widely used in

firms with more concentrated borrowing after Lehman's default suffer on average a smaller contraction in bank credit and have a lower probability of being credit rationed. Regarding the variable IRD_{jt}, Bodenhorn (2003) shows that borrowers with longer relations are more likely to have loan terms renegotiated during a credit crunch. Elsas (2005) shows that firms that preserve their relation for a relatively long period face lower financial constraints and experience better credit terms and conditions. Bonaccorsi di Patti and Gobbi (2007) show that longer relationships imply fewer costs and easier sources of finance. Gobbi and Sette (2015) show that the credit growth has been higher after Lehman's default for longer lending relations.

[22] The average IRD amounts to 39 consecutive months on the lending side and 27 months on the borrowing side (the one considered in the paper). As a robustness check, we allowed alternatively for zero, two, and three months of interruption in order to consider a relationship as continuous: results are robust to these different specifications. Section 7 provides more details on this point.

the analysis of interbank markets, although mainly to analyze financial contagion. In this literature, banks are the units (or nodes) and the amounts of interbank exposures constitute the weighted links. The three centrality measures we use are degree (i.e., the number of interbank connections of each bank); betweeness centrality (i.e., an index of interbank centrality of each bank that indicates the banks that each bank has to go through in order to reach another bank in the minimum number of hops); and closeness centrality (i.e., an index of interbank centrality of each bank that captures the length of shortest path to all others).

The subset $Bilateral\ Network\ Centrality_{jt}$ complements $Bilateral\ Network\ Relationships_{jt}$ as it captures the role of each bank in the web of the bilateral market, which could be a central one even if the bank does not have concentrated and/or stable bilateral relationships. A bank could, for example, try to establish a ramified interbank network (e.g., by having multiple, albeit occasional, counterparties) precisely because it lacks strong bilateral relationships: the outcome of such a strategy would be precisely a high centrality measure and low ICC_{jt} and IRD_{jt}.

Other important bank-specific covariates are included as control variables in the matrix KR_{jt}. *Retail Fundraising* and *Central Bank Loans* describe funding sources alternative to the CCPs. *Tier1* and *RoE* describe, respectively, bank capitalization and profitability, while *Size*, *Loans to Private Sector,* and *Portfolio of Government Bonds* approximate important aspects of a bank's business model. The last variable also provides a rough proxy for collateral availability. All variables are described in table 1.

In order to distinguish different phases of the financial crisis and to take into account that in some euro-area countries, including Italy, access to funding was more difficult during the sovereign debt crisis than in the previous phase of the financial crisis, we consider two crisis-related dummies. The dummy *CR1* covers the period from the Lehman Brothers bankruptcy in September 2008 to June 2011, when the sovereign crisis hit Italy. The dummy *CR2* covers the sovereign crisis and runs until the end of the sample period in June 2013. Monthly time dummies p_t are also included, where possible, to take into account the impact of particular events, such as the impact of a change in CCPs' haircuts in November 2011 or the launch of the Long-Term Refinancing

Operations by the ECB, as well as other unobservable time-varying variables.[23]

While our analysis explores the demand (bank) side determinants of CCPs' use, supply factors such as changes in the risk-management policies of the Italian CCP or in its standards and conditions (e.g., fees, margins, collateral requirements) may be very relevant as well. As we have only one CCP operating in Italy, supply-side factors apply to all banks, and therefore either they have the same effect on all banks—and then they may explain a generalized increasing recourse to CCPs, but not a differential use across banks—or they have a different impact on banks but this impact would depend on (heterogeneous) bank characteristics (e.g., a change in CCPs' risk-management policy or collateral requirement can have differential effects on banks' participation due to specific bank riskiness or collateral endowment). In the first case (i.e., in the unlikely case that the effect had the same effects on all banks), supply factors are seized, from an econometric point of view, by the time fixed effects, which capture aggregate fluctuations of the dependent variable over time. In the second case (i.e., when the effect is bank specific), our analysis focusing on the determinants at bank level of the growing use of CCPs should be perfectly able to identify the effect.

We add, however, in some specifications a supply-side variable, $Margins_t$, computed as a monthly average of the margins applied by the CCP to several kinds of securities used as collateral in each month. An increasing value of the variable corresponds to a tightening of supply conditions. In addition to the covariate $Margins_t$, we interact it with each variable measuring banks' characteristics. Should supply-side factors be relevant, these interaction terms would result statistically significant, indicating that banks react heterogeneously to supply changes depending on their characteristics.

To estimate equation (1) we run a zero-inflated beta regression model. The model is made of two steps: in the first step (which explores the determinants of *participation* in CCPs) the dependent

[23]Instead of using time dummies, we also used continuous variables accounting for major developments that could affect our variables, such as the total liquidity injected by the Eurosystem, GDP growth, and inflation rates, with no significant impact on our results.

variable is a dummy 0,1; in the second step (which investigates the *intensity* of the recourse to CCPs conditional on participation) the dependent variable is a ratio. This model has two specific advantages with respect to alternative specifications, as it allows to take into account that (i) most banks do not use the CCPs for their funding (especially during the first part of the sample period); and (ii) our dependent variable is a ratio (the share of CCP exposures over the total interbank exposures).

The zero-inflated beta regression model aims to address the specification errors arising from (i) modeling a ratio variable as a linear function of the explanatory variables; and (ii) ignoring that the conditional variance must be a function of the conditional mean since the former must change as the conditional mean approaches either 0 or 1 (e.g., Papke and Wooldridge 1996; Cook, Kieschnick, and McCullough 2008). In addition, the zero-inflated approach allows us to take into account that determinants of zero and positive observations (once an intermediary decides to use CCPs) may be different, avoiding the related selection bias. While most of the increase in the use of CCPs is driven in each year by the intensive margin, as expected, the data show that between 2009 and 2010 and again between 2011 and 2013 also the contribution of the extensive margin (i.e., the funding obtained by banks which were not operating via CCPs the year before) is not irrelevant (table 2). It is therefore important to have the possibility to look at both aspects as carefully as possible.[24]

5.2 Second Step: Use of CCPs by Riskier Borrowers

Our second step aims to investigate whether recourse to CCPs allowed riskier banks to elude peer monitoring, potentially increasing the risk borne by the financial system as a whole. For such a conclusion, it is not enough to show that individual bank risk is positively associated with CCPs' share in the overall interbank transactions: a measure is needed to link the risk associated with each bank, as

[24] As a robustness check, we also carried out a standard panel regression for equation (1), obtaining fully consistent results, once the limitations of the panel approach in this specific setting were taken into account.

assessed by its bilateral interbank counterparties, to its recourse to CCPs. To evaluate if CCPs run the risk to fund a pool of borrowers that are shunned by (the most informed among) their bilateral interbank counterparties, we need an indicator able to capture the assessment of these bilateral interbank counterparties.

The measure we propose to summarize the assessment made by interbank peers, as revealed by their behavior, is the *change* in the weighted average duration of each intermediary's interbank relationships IRD_{jt}, where IRD_{jt} is the *Interbank Relationship Duration* for bank j at time t, defined above.

In formal terms, we estimate equation (2) with a fixed-effect panel estimation model:

$$\Delta SH_{jt} = \alpha_0 UNC_t + \beta_0 Risk_{jt} + \gamma_0 \Delta IRD_{jt} + \gamma_1 \Delta IRD_{jt} * Risk_{jt} \\ + \delta' KR_{jt} + \zeta' b_j + \eta' p_t + \epsilon_{jt}, \tag{2}$$

where variables are defined as above and changes are over the previous month.

As the literature on relationship lending shows that long-lasting partnerships are characterized by better information (see section 3), a positive ΔIRD_{jt} would signal that on average better-informed counterparts keep their relationship with the bank j while a negative change would signal a drying-up of interbank funding by longer-standing counterparts. The relationship between changes in the share of funding via CCP and changes in the weighted average duration of bilateral interbank relationships should then have, ceteris paribus, a negative sign for riskier banks if CCP transactions replace older bilateral relationships (as the loss of these relationships shortens the weighted average duration of bilateral transactions). Using this measure addresses possible concerns about the precision and/or the observability by counterparties of the measures of risks used in our first step's regressions and it allows to tackle the issue of whether the CCPs are taking risks that are dodged by bilateral counterparts. Moreover, our measure of "duration" refers to the continuity of the relationship between two interbank counterparties, not to the maturity of the contract: this means that the fact that during the crisis long-term deals became increasingly unlikely makes our measure more able to timely record any change in the assessment of the standing of a counterparty as shortened maturities implied a more frequent renegotiation of deals.

6. Results

6.1 First Step: Determinants

The results of our first step are reported in tables 4 and 5.

Table 4 shows the results on the determinants of *participation* in CCPs (the dependent variable is a dummy 0,1), while table 5 shows those related to the *intensity* of the recourse to CCPs, conditional on participation (the dependent variable is a ratio). It is important to note that in the estimation of participation reported in table 4 (first stage of the zero-inflated beta regression model), a positive sign indicates a lower participation (more zeros) and a negative sign a higher participation (fewer zeros).

Starting from the interbank bilateral factors underlying the participation in CCPs transactions, we find that stronger interbank bilateral relationships (the variable ICC) are associated with a lower participation, supporting the idea that the two channels tend to be alternative in normal conditions (table 4). During both phases of the crisis, however, this association tended to fade away, as also banks with established bilateral relationships had to tap all the available sources of funding, including the CCPs. Similar results hold when looking at the intensity of use (share of funding via CCPs), conditional on the participation in the market (table 5): we find that strong bilateral relationships reduce the intensity of CCP use in normal conditions, but that this association disappeared during the crises.

As for foreign extragroup interbank funding (as a ratio to total interbank funding), we find that it has a negative impact on participation (i.e., banks with higher bilateral funding from abroad were less likely to resort to CCPs; table 4). We also use the change in funding from abroad as an explanatory variable and find that, as expected, a negative change in foreign funding is associated with a higher use of CCPs.

Results on network indicators show that before the onset of the crisis, a higher centrality in interbank bilateral market favored both participation and intensity of use of CCPs, while during the crisis

Table 4. Determinants of Interbank Exposures through CCPs: Determinants of *Participation* in CCPs (the dependent variable is a dummy 0,1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
												Only Regressor	Interacted with Margins
Foreign Interbank Debts	29.372***	29.296***	31.281***		26.693***	28.257***	28.619***	28.885***			26.343***		
	9.117	*8.995*	*11.089*		*9.301*	*8.268*	*8.065*	*9.427*			*8.815*		
Delta (Foreign Interbank Debts)				6.925**					7.577**	14.173***		14.685*	−0.003
				3.241					*3.543*	*4.488*		*7.520*	*0.009*
UNC	3.604**	2.604	4.704*	4.713**	4.324***	3.513	2.983	5.279*	3.954	2.769	4.276	1.246**	
	1.543	*1.826*	*2.614*	*1.831*	*1.390*	*2.877*	*2.371*	*3.180*	*3.034*	*2.065*	*3.221*	*0.580*	
UNC × Crisis 1						−3.474*	−6.093***	−3.701***	−3.733	−2.362***	−10.006***		
						2.087	1.730	1.429	2.389	0.732	3.156		
UNC × Crisis 2						−5.715**	−8.240***	−5.781**	−5.968**	−4.539***	−11.956***		
						2.390	*2.004*	*2.292*	*2.617*	*1.143*	*1.982*		
ICC	3.510***	3.940***		4.089*		8.051***	8.345***		8.250***	7.285**		2.331**	0.001
	1.330	*1.410*		*1.701*		*2.525*	*2.714*		*2.531*	*3.170*		*1.189*	*0.001*
ICC × Crisis 1						−5.577**	−5.208**		−5.527**	−5.204*			
						2.402	*2.417*		*2.257*	*3.248*			
ICC × Crisis 2						−5.540**	−5.421*		−5.310**	−4.524			
						2.721	*2.825*		*2.470*	*3.255*			
IRD	0.072		0.121	0.180		0.554		0.316	0.737			−0.796	0.002
	0.177		*0.161*	*0.178*		*0.556*		*0.355*	*0.613*			*1.268*	*0.002*
IRD × Crisis 1						−0.432		−0.259	−0.505				
						0.789		*0.681*	*0.799*				
IRD × Crisis 2						−0.478		−0.157	−0.582				
						0.599		*0.437*	*0.638*				
Betweeness Centrality					*23.707*						*−0.856****		
					19.636						*0.311*		
Betweeness Centrality × Crisis 1											0.707***		
											0.207		
Betweeness Centrality × Crisis 2											0.775***		
											0.295		
Margins												0.598	
												1.351	

(*continued*)

Table 4. (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12) Only Regressor	(12) Interacted with Margins
Bad Loans	0.121	0.555	1.455	−5.391	−0.116	−74.133**	−55.878*	−40.811**	−88.955**	−83.777**	−40.329*	−37.648	0.037
	16.099	*14.094*	*12.543*	*15.185*	*0.194*	*30.873*	*30.244*	*19.594*	*37.926*	*35.374*	*22.620*	*59.017*	*0.057*
Bad Loans × Crisis 1						76.181***	58.478**	60.182***	83.550**	85.862**	44.331		
						27.454	*26.642*	*22.615*	*35.108*	*33.891*	*31.459*		
Bad Loans × Crisis 2						87.453***	71.118**	65.933***	99.470**	91.362**	67.692**		
						31.551	*30.413*	*24.623*	*38.703*	*37.325*	*29.272*		
Size	−2.772**	−3.103**	−2.960**	−3.015**	−3.479**	−2.355***	−2.551**	−2.780**	−2.297**	−0.998	−2.634	1.465	−0.002**
	1.177	*1.358*	*1.347*	*1.490*	*1.698*	*0.920*	*0.998*	*1.324*	*0.893*	*1.325*	*2.213*	*3.863*	*0.001*
Retail Fundraising	−8.753***	−9.549***	−9.505***	−8.542***	−20.609***	−6.666**	−7.713**	−6.755**	−6.511*	−4.651	−18.342***	−23.194***	−0.009
	3.072	*3.179*	*3.434*	*3.584*	*4.917*	*3.363*	*3.274*	*3.416*	*3.827*	*3.644*	*4.547*	*8.149*	*0.008*
Loans to Private Sector	8.517*	7.962*	7.652*	8.217*	12.504**	11.691***	11.498**	8.336*	11.919***	15.654***	15.936**	20.191***	−0.006
	4.859	*4.929*	*4.716*	*4.850*	*5.327*	*4.464*	*4.812*	*4.635*	*4.168*	*4.430*	*6.649*	*6.999*	*0.009*
Central Bank Loans	7.848	8.490	5.753	5.495	−2.623	5.764	6.947	4.591	3.423	4.219	−4.684	32.424	−0.055**
	9.806	*9.819*	*9.464*	*9.864*	*9.082*	*9.804*	*9.848*	*9.409*	*9.627*	*9.460*	*9.674*	*24.619*	*0.027*
Portfolio of Government Bonds	−0.863	−0.762	−0.202	−1.775	1.805	−0.880	−0.680	−1.868	−2.088	−2.896	4.263	44.872***	−0.052***
	5.479	*5.494*	*4.928*	*6.178*	*5.130*	*5.632*	*5.708*	*6.245*	*6.122*	*5.800*	*6.531*	*15.534*	*0.018*
ROE	0.725	0.705	1.489	0.859	1.449	0.284	0.523	1.357	0.471	−0.202	1.070	−12.528	0.012
	1.250	*1.181*	*1.311*	*1.153*	*0.985*	*1.271*	*1.273*	*1.332*	*1.247*	1.449	*1.309*	*13.232*	*0.019*
Constant	28.862**	38.277**	32.536**	33.971**	46.047**	24.817**	32.157	30.773*	26.889**	7.738	44.672	2.803	
	14.173	*16.841*	*15.783*	*17.651*	*22.900*	*12.240*	*13.367*	*16.178*	*12.519*	*16.414*	*28.556*	*47.040*	
No. of Obs.	15,279	15,279	15,279	13,766	15,279	15,279	15,279	15,279	13,766	13,766	15,279	13,766	

Notes: The table reports zero-inflated beta regression model results of the first part of the estimation: factors driving the choice to participate or not in the funding via CCPs. The results of the second part of the estimation (the intensity of the recourse to CCPs) are reported in table 5. Observations are clustered at banking group level (and at bank level for independent banks), thus obtaining heteroskedasticity-robust standard errors and controlling for possible autocorrelations across the same banking group. In the estimation of participation, a positive sign indicates a lower participation (more zeros) and a negative sign a higher participation (fewer zeros). The table reports regression coefficients and associated standard errors in italics. ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively. Results from specification (12) are reported in two columns even if obtained in a single regression: the first column of specification (12) reports the results of the variable $Margins_t$ in addition to the standalone variables of specification (4); the second column of specification (12) reports the interaction terms between the variable $Margins_t$ and each bank characteristic.

Table 5. Determinants of Interbank Exposures through CCPs: Determinants of the *Intensity* of the Recourse to CCPs (the dependent variable is a ratio)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)		Marginal Effects
												Only Regressor	Interacted with Margins	
Foreign Interbank Debts	−1.240	−0.879	−1.222		−0.170	0.693	0.814	−1.166			0.463			ns (−2.5)
	1.647	*1.692*	*1.614*		*1.375*	*1.699*	*1.748*	*1.488*			*1.027*			
Delta (Foreign Interbank Debts)				0.377					0.478	1.028		−3.670***	0.006**	
				1.354					*1.362*	*0.758*		*1.348*	*0.002*	
UNC	1.423	1.864	1.468	1.828*	0.455	0.370	0.270	0.667	0.278	−0.782	−0.407	0.126		ns (5.7)
	1.075	*5.974*	*1.074*	*0.933*	*4.981*	*1.220*	*7.388*	*1.174*	*1.143*	*11.510*	*6.851*	*9.513*		
UNC × Crisis 1						0.565	0.434	0.560	0.582	0.320	−0.508			ns (17.9)
						0.486	1.422	0.487	0.472	5.222	1.606			
UNC × Crisis 2						0.487*	0.336	0.534**	0.510*	0.218	0.188			15.3
						0.296	*1.697*	*0.213*	*0.290*	*5.604*	*1.874*			
ICC	0.013	0.017		−0.012		−3.175***	−2.847***		−2.939***	−2.133**		−0.086	0.002	−14.5
	0.303	*0.331*		*0.287*		*0.913*	*0.772*		*0.945*	*0.955*		*0.271*	*0.002*	
ICC × Crisis 1						3.128***	2.683***		2.873***	2.047**				13.9
						0.758	*0.658*		*0.836*	*0.976*				
ICC × Crisis 2						3.423***	3.131***		3.188***	2.441**				15.8
						0.952	*0.835*		*1.031*	*1.005*				
IRD	0.112		0.113	0.104		0.212		−0.226	0.226			0.391	−0.001	ns (1.8)
	0.098		*0.099*	*0.102*		*0.235*		*0.288*	*0.253*			*0.274*	*0.000*	
IRD × Crisis 1						0.024		0.493	0.012					ns (3.0)
						0.271		*0.318*	*0.286*					
IRD × Crisis 2						−0.170		0.279	−0.182					ns (−1.4)
						0.260		*0.313*	*0.275*					
Betweeness Centrality					−0.230						0.195***			
					0.444						*0.075*			
Betweeness Centrality × Crisis 1											*−0.139***			
											0.068			
Betweeness Centrality × Crisis 2											*−0.322****			
											0.076			
Margins												0.047		
												0.325		

(*continued*)

Table 5. (Continued)

												(12)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	Only Regressor	Interacted with Margins	Marginal Effects
Bad Loans	15.394***	16.063***	15.398***	15.444***	−4.830**	−6.258	−2.149	−5.733	−6.315	−16.191	−4.410	−32.077*	0.005	ns (−3.0)
	3.515	*3.361*	*3.598*	*3.457*	*2.158*	*10.723*	*9.875*	*10.684*	*10.674*	*10.377*	*10.878*	*16.594*	*0.018*	
Bad Loans ×						14.540*	12.668*	14.819**	14.561*	21.141**	14.506**			6.7
Crisis 1						*7.940*	*7.477*	*7.293*	*7.939*	*8.589*	*7.373*			
Bad Loans ×						19.697**	17.805**	18.957**	19.767**	28.272***	18.435**			9.0
Crisis 2						*9.758*	*8.896*	*9.501*	*9.681*	*9.692*	*7.457*			
Size	−0.541	−0.582	−0.552	−0.689*	8.143*	−0.232	−0.219	−0.354	−0.207	0.061	−0.028	0.107	0.000	ns (−4.2)
	0.578	*0.564*	*0.565*	*0.411*	*4.613*	*0.654*	*0.591*	*0.676*	*0.581*	*0.620*	*0.515*	*0.760*	*0.000*	
Retail	0.016	0.086	0.020	0.081	5.450***	0.223	0.279	0.193	0.222	−0.106	6.465***	0.104	0.010**	ns (6.3)
Fundraising	*0.920*	*0.832*	*0.930*	*0.947*	*1.133*	*0.941*	*0.877*	*0.945*	*0.930*	*0.927*	*1.195*	*2.332*	*0.004*	
Loans to	−1.252	−1.342	−1.269	−1.680	3.150	0.100	0.101	−0.119	0.217	0.357	−5.034***	−4.046	−0.004	ns (3.5)
Private Sector	*2.215*	*2.173*	*2.210*	*1.821*	*2.525*	*2.271*	*2.151*	*2.446*	*2.090*	*2.166*	*1.616*	*3.566*	*0.003*	
Central Bank	1.114	1.349	1.120	1.053	−0.008	1.342	1.683	1.194	1.382	1.584	4.772**	−2.384	0.007	ns (1.0)
Loans	*1.729*	*1.793*	*1.727*	*1.672*	*0.062*	*1.580*	*1.688*	*1.698*	*1.597*	*1.602*	*2.186*	*6.061*	*0.008*	
Portfolio of Government	5.954***	5.916***	5.968***	6.097***	2.543**	6.325***	6.238***	6.179***	6.284***	6.303***	1.521*	10.711**	−0.015**	5.6
Bonds	*1.349*	*1.403*	*1.271*	*1.339*	*1.055*	*1.333*	*1.405*	*1.247*	*1.303*	*1.505*	*0.828*	*5.459*	*0.007*	
ROE	0.016	0.044	0.018	0.016	−0.060	0.193	0.193	0.236	0.199	0.168	−0.088	−2.526	0.004	ns (1.9)
	0.380	*0.404*	*0.360*	*0.383*	*0.376*	*0.357*	*0.369*	*0.351*	*0.359*	*0.341*	*0.366*	*2.073*	*0.003*	
Constant	1.811	1.084	1.933	3.521	−5.150***	−0.606	−0.565	0.470	−0.925	−1.965	−5.150***	−0.147		
	7.041	*0.754*	*6.910*	*4.926*	*0.755*	*7.688*	*0.754*	*8.049*	*6.853*	*1.754*	*0.755*	*0.160*		
No. of Obs.	15,279	15,279	15,279	13,766	15,279	15,279	15,279	15,279	13,766	13,766	15,279	13,766		

Notes: The table reports zero-inflated beta regression model results of the second part of the estimation: factors influencing the intensity of the recourse to the CCPs, conditional on participation. The results of the first part of the estimation (determinants of *participation* in CCPs) are reported in table 4. Observations are clustered at banking group level (and at bank level for independent banks), thus obtaining heteroskedasticity-robust standard errors and controlling for possible autocorrelations across the same banking group. The table reports regression coefficients, associated standard errors in italics, and marginal effects. ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively. Results from specification (12) are reported in two columns even if obtained in a single regression: the first column of specification (12) reports the results of the variable $Margins_t$ in addition to the standalone variables of specification (4); the second column of specification (12) reports the interaction terms between the variable $Margins_t$ and each bank characteristic. The marginal effect of each determinant is computed for specification (6) measuring the percentage change of the dependent variable (the share of exposures via CCPs on total interbank exposures) moving from the 25th to the 75th percentile of each regressor. Outcomes are very similar in the other specifications.

banks more central in the bilateral interbank market showed less need to turn to CCPs.[25]

Turning to market uncertainty, we find that it was not a significant factor in driving banks to CCPs until the start of the financial crisis. Then, for both the crisis periods, it became significant and associated with both a larger participation and a larger share of CCPs' transactions, reflecting the general move toward secured transactions at times of heightened risk aversion.[26]

The individual risk of a bank, proxied by its bad loans ratio, affects both the participation and the intensive use of CCPs but in opposite directions.[27] Participation of riskier banks in CCPs is more likely before the crisis and becomes instead less likely in both the crisis periods. By contrast, for banks already using CCPs, individual bank risk becomes a significant positive determinant of the proportion of CCPs transactions during the crisis (coefficients are significant in both subperiods, slightly larger during the sovereign debt crisis phase), in line with the hypothesis that a more intense scrutiny took place in other segments of the interbank market.

Table 5 reports the marginal effects of each regressor on the dependent variable, other things being equal.[28] The total net impact of our measures of individual risk and general uncertainty are sizable and very similar. Moving from the 25th percentile to the 75th percentile of the bad loans ratio, the intensity of the use of CCPs increases during the crisis with an impact ranging from 7 to 9 percent

[25]The results related to the pre-crisis period may reflect the fact that in the infancy of interbank activity on CCPs the banks more active on the interbank market were also experimenting with the new channel while, later on, the two channels may have been substitutes.

[26]To support this interpretation, we ran a similar regression for lenders, who are likely the most affected by uncertainty about counterparty risk. We found that the participation in CCPs is indeed higher when our measure of general uncertainty is higher and when the degree of concentration of bilateral lending is lower.

[27]When banks' individual risk is proxied by the pair of variables on banks' credit rating, results are broadly similar to those of table 4 (not reported).

[28]Marginal effects are computed only for the intensity of the recourse to CCPs measuring the percentage change of the dependent variable moving from the 25th to the 75th percentile of each regressor for specification (6). Outcomes are very similar in the other specifications. Marginal effects on the participation in CCPs (first stage of the zero-inflated beta regression model) are not reported because the dependent variable is a dummy 0,1.

in the two phases of the crisis, while the uncertainty increases the share of CCP transactions during the sovereign part of the crisis by around 15 percent.

Turning to the other covariates, we find that larger banks tend to participate more in CCPs. The share of centrally cleared transactions is also higher for banks with a higher share of government bonds over total assets, broadly confirming the relevance of collateral availability for this type of funding.

To test for the possible influence of supply factors on the use of CCPs, tables 4 and 5 include in specification (12) a supply-side variable, $Margins_t$.[29] In both tables, results from specification (12) are reported in two columns: the first column reports the results of the variable $Margins_t$ in addition to the variables of specification (4); the second column reports the interaction terms between $Margins_t$ and each bank characteristic. The coefficient associated with $Margins_t$ is not statistically significant (first column of specification (12)), while some interaction terms are statistically significant (second column), indicating that supply factors may have different impacts on banks according to their characteristics.[30] The interaction term with the individual risk of a bank is, however, not significant, suggesting that the impact of this variable on the use of CCPs is not channeled via supply factors.

6.2 Second Step: CCPs and Riskier Borrowers

Results of the first step provide a broad view of the factors driving participation and recourse to CCPs transactions before and during the financial crisis, confirming that both uncertainty and risk play a significant role. In the second step we focus on the monthly changes of the weighted average duration of the bilateral interbank relationships of each borrowing bank, IRD_{jt}. If the shift to CCPs derives from bank-specific risk, older (i.e., better-informed) counterparts should maintain relationships with safer borrowing banks and

[29] As mentioned in section 5, this variable is computed as a monthly average of the margins applied by the CCP to several kinds of securities used as collateral in each month.

[30] This is the case, in particular, for *Size, Retail Fundraising, Foreign Interbank Debts*, and *Portfolio of Government Bonds*.

shut down those with riskier banks. The latter could then be forced to recur to CCPs: accordingly, the relationship between changes in SH_{jt} and IRD_{jt} should be negative for riskier banks (and positive and/or not significant for less risky intermediaries). To check if this is indeed the case, we separate banks according to their decile in the bad loans ratio distribution and we then check if the coefficients associated with the interaction terms are negative and significant for the banks belonging to the upper deciles of the risk distribution while positive and/or nonsignificant for the lower deciles.

Table 6 summarizes the results of equation (2). It shows, first, that changes in the use of CCPs are negatively related to changes in the weighted average duration but only during the crisis (specifications (1) and (2)). Moreover, in line with our hypothesis, the driver of this result is the level of individual risk, as indicated by the fact that only the interaction term is significant in specifications (3) and (4). Results are supportive of our interpretation of the weighted average duration variable as the relationship between changes in SH_{jt} and IRD_{jt} becomes negative as we move from the lowest to the highest levels of banks' risk. In particular, interacting the changes in the weighted average duration with the deciles of our risk indicator (bad loans ratio), we find that the negative effect is limited to the highest deciles of the distribution by risk (the last two deciles in the first part of the crisis and the last one in the sovereign debt crisis).

For the riskiest borrowers, therefore, the negative and significant sign of the changes in average duration suggests that a relevant determinant of the increased recourse to the CCPs is the loss of more established interbank customer relationships, a signal that there may be a specific issue with the risk associated with that bank.

7. Robustness Checks

This section summarizes the main robustness checks we carried out.[31]

[31] For the sake of brevity, some checks are not reported in additional tables, but they are all available from the authors. In some of our estimations, the sample may vary due to missing values for some variables or due to the use of $\Delta variables$. As a further check, we restricted all estimations to the largest sample consistent across all specifications, and results remain the same.

Table 6. Determinants of Δ(interbank exposures through CCPs)

	(1)	(2)	(3)	(4)	(5)	(6)
Foreign Interbank Debts	0.010	0.012	0.015	0.016	0.017	0.023
	0.043	*0.043*	*0.043*	*0.043*	*0.044*	*0.040*
UNC	−0.004	−0.003	−0.004	−0.004	−0.009	0.024
	0.006	*0.006*	*0.006*	*0.006*	0.008	*0.018*
Size	0.031***	0.030***	0.031***	0.031***	0.029***	0.017**
	0.009	*0.009*	*0.009*	*0.009*	*0.009*	*0.008*
Retail Fundraising	0.056	0.055	0.056	0.057	0.062	0.038
	0.086	*0.085*	*0.086*	*0.085*	*0.086*	*0.081*
Loans to Private Sector	0.003	−0.004	0.003	0.001	−0.004	−0.028
	0.040	*0.040*	*0.040*	*0.040*	*0.040*	*0.043*
Central Bank Loans	0.041	0.041	0.041	0.042	0.022	−0.009
	0.086	*0.084*	*0.085*	*0.085*	*0.083*	*0.084*
Portfolio of Government Bonds	0.352***	0.345***	0.349***	0.349***	0.336***	0.378***
	0.129	*0.127*	*0.129*	*0.128*	*0.127*	*0.135*
ROE	−0.018	−0.017	−0.018	−0.018	−0.013	−0.016
	0.013	*0.013*	*0.013*	*0.013*	*0.013*	*0.012*
Δ(IRD)	−0.004***	0.002	−0.001	−0.001	0.002	0.003
	0.002	*0.002*	*0.003*	*0.003*	*0.003*	*0.003*
Δ(IRD) × Crisis 1		−0.004*			*0.001*	−0.001
		0.002			*0.002*	*0.002*
Δ(IRD) × Crisis 2		*−0.017***			−0.003	−0.007*
		0.007			*0.003*	*0.004*
Bad Loans	0.316	0.302	0.336	0.348	−0.039	Included as deciles, and unreported. See notes at bottom of table.
	0.243	0.242	*0.344*	*0.244*	*0.252*	
Bad Loans × Crisis 1					*0.253*	
					0.185	
Bad Loans × Crisis 2					0.486*	
					0.288	
Δ(IRD) × Bad Loans			−0.162*	*0.029*		
			0.085	*0.085*		
Δ(IRD) × Bad Loans × Crisis 1				−0.162*		
				0.084		
Δ(IRD) × Bad Loans × Crisis 2				*−0.335***		
				0.159		

(*continued*)

Table 6. (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)
Δ(IRD) × Bad Loans (2° Quartile) × Crisis 1					−0.004	
					0.007	
Δ(IRD) × Bad Loans (2° Quartile) × Crisis 2					−0.024	
					0.021	
Δ(IRD) × Bad Loans (3° Quartile) × Crisis 1					−0.002	
					0.005	
Δ(IRD) × Bad Loans (3° Quartile) × Crisis 2					−0.016	
					0.013	
Δ(IRD) × Bad Loans (4° Quartile) × Crisis 1					−0.012*	
					0.006	
Δ(IRD) × Bad Loans (4° Quartile) × Crisis 2					−0.021*	
					0.012	
Δ(IRD) × Bad Loans (7° Decile) × Crisis 1						0.002
						0.005
Δ(IRD) × Bad Loans (7° Decile) × Crisis 2						0.006
						0.007
Δ(IRD) × Bad Loans (8° Decile) × Crisis 1						0.011
						0.007
Δ(IRD) × Bad Loans (8° Decile) × Crisis 2						−0.027
						0.037
Δ(IRD) × Bad Loans (9° Decile) × Crisis 1						*−0.024***
						0.011
Δ(IRD) × Bad Loans (9° Decile) × Crisis 2						−0.004
						0.008
Δ(IRD) × Bad Loans (10° Decile) × Crisis 1						*−0.017***
						0.008
Δ(IRD) × Bad Loans (10° Decile) × Crisis 2						−0.032*
						0.016
Constant	−0.287***	−0.278***	−0.287***	−0.289***	−0.266***	−0.160**
	0.075	*0.072*	*0.074*	*0.074*	*0.075*	*0.069*
Rho	0.37	0.36	0.37	0.38	0.35	0.32
No. of Obs.	11,008	11,008	11,008	11,008	11,008	11,008

Notes: The table reports fixed-effects panel results, where fixed effects are for banks; time fixed effects also are always included. Observations are clustered at banking group level (and at bank level for independent banks), thus obtaining heteroskedasticity-robust standard errors and controlling for possible autocorrelations across the same banking group. Partial interaction terms are always included even if unreported; in specification (6), the other deciles' results are not reported. The table reports regression coefficients and associated standard errors in italics. ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

Uncertainty and Time Fixed Effects. The effect of market uncertainty and risk aversion on the use of CCPs was tested in two ways. First, as mentioned, we ran our regression with different definitions of the variable UNC_t. A first alternative measure to that presented in section 5.1 directly relies on VSTOXX, the index based on Euro Stoxx 50 options prices computed according to VIX methodology. A second alternative measure is the Composite Indicator of Systemic Stress (CISS) index, which summarizes contemporaneous stress in the financial system (Holló, Kremer, and Lo Duca 2012).[32] The three measures used were moving in a very similar way during our sample period (figure 6) and results are equivalent. In table 7 (specifications (1) and (2)) and table 8 (specification (1)), we report results from regression analogous, respectively, to those in tables 4 and 5 (specifications (1) and (6)) and table 6 (specification (1)) using the CISS index instead of the ratio between the densities: results are unchanged.

As a second way to check the robustness of the variable UNC_t, we either dropped or changed time fixed effects. In tables 4–6 we reported results of equations (1) and (2) that included time fixed effects to allow for all macro unobservable time-varying variables. As time dummies could affect the estimation of the variable UNC_t, absorbing some of its effect on the dependent variable, we ran the same regressions dropping time fixed effects, and the coefficient associated with the variable UNC_t remained stable: table 7, specification (3) and (4), for the first step; and table 8, specifications (2), (3), and (4), for the second step.[33]

Regulatory Drivers to Use CCPs. An important reason for using CCPs may be the regulatory benefits they provide, as a consequence of the regulatory reforms promoted after the financial crisis.

[32] CISS is computed by applying basic portfolio theory to the aggregation of five market-specific subindexes created from a total of 15 individual financial stress measures. The aggregation accordingly takes into account the time-varying cross-correlations between the sub-indexes. As a result, the CISS puts relatively more weight on situations in which stress prevails in several market segments at the same time, capturing the idea that financial stress is more systemic and thus more dangerous for the economy as a whole if financial instability spreads more widely across the whole financial system.

[33] Results are also robust to the choice of the time dummy to be dropped to allow for the inclusion of the measure of market uncertainty.

Table 7. Robustness Checks: Determinants of *Participation* in CCPs and Intensity of the Recourse to CCPs Conditional on Participation

	Participation (Dep. Variable: 0,1 Dummy)				Intensity (Dep. Variable: Ratio)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Using CISS instead of the Ratio of Densities as a Measure of Uncertainty		Dropping Time Fixed Effects		Using CISS instead of the Ratio of Densities as a Measure of Uncertainty		Dropping Time Fixed Effects	
Foreign Interbank Debts	26.925***	28.246***	26.111**	20.682***	0.383	0.688	−6.184***	0.458
	9.241	*8.258*	*11.354*	*7.482*	*1.353*	*1.699*	*2.377*	*1.531*
UNC	16.825***	12.989	1.034*	2.674*	5.080***	0.733	0.206	−0.046
	3.449	*8.493*	*0.522*	*1.370*	*1.481*	*0.830*	*0.245*	*0.372*
UNC × Crisis 1		−8.876***		1.019		1.334		0.620*
		2.936		*1.296*		*1.620*		*0.367*
UNC × Crisis 2		−17.161*		−2.565*		1.781*		0.849**
		8.928		*1.376*		*1.003*		*0.360*
ICC	4.090**	8.049***	1.336	8.688***	−0.012	−3.179***	0.286	−4.337**
	1.701	*2.528*	*1.094*	*3.303*	*0.287*	*0.913*	*0.389*	*1.698*
ICC × Crisis 1		−5.572**		−8.310***		3.134***		5.189***
		2.407		*3.112*		*0.758*		*1.617*
ICC × Crisis 2		−5.538**		−7.225**		3.428***		4.622***
		2.725		*3.050*		*0.952*		*1.655*
IRD	0.180	0.552	0.256	0.419	0.104	0.210	0.017	0.161
	0.178	*0.556*	*0.171*	*0.570*	*0.102*	*0.235*	*0.100*	*0.329*
IRD × Crisis 1		−0.430		−0.236		0.026		−0.006
		0.789		*0.729*		*0.271*		*0.348*
IRD × Crisis 2		−0.477		−0.149		−0.168		−0.099
		0.599		*0.594*		*0.260*		*0.367*

(*continued*)

Table 7. (Continued)

	Participation (Dep. Variable: 0,1 Dummy)				Intensity (Dep. Variable: Ratio)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Using CISS instead of the Ratio of Densities as a Measure of Uncertainty		Dropping Time Fixed Effects		Using CISS instead of the Ratio of Densities as a Measure of Uncertainty		Dropping Time Fixed Effects	
Bad Loans	−5.389	−74.071**	−54.600***	−83.642***	15.452***	−6.302	19.071***	−5.593
	15.186	30.875	10.702	27.304	3.460	10.735	5.088	8.889
Bad Loans × Crisis 1		76.260***		11.408		14.578*		20.174**
		27.455		22.088		7.945		7.919
Bad Loans × Crisis 2		87.526***		51.961**		19.745**		21.031***
		31.572		25.883		9.765		7.705
Size	−3.015**	−2.328**	−4.688***	−4.861***	−0.681*	−0.224	−0.443**	0.867
	1.490	0.918	1.086	1.262	0.412	0.654	0.500	0.720
Retail Fundraising	−8.542**	−6.656**	−10.097***	−9.872***	0.081	0.222	−0.120	0.392
	3.584	3.361	3.486	3.535	0.947	0.941	1.523	1.134
Loans to Private Sector	8.219*	11.764***	−3.952	−2.755	−1.665	0.118	1.949	2.841*
	4.850	4.467	3.088	3.229	1.825	2.271	1.330	1.460
Central Bank Loans	5.497	5.766	−7.237	−2.283	1.046	1.338	0.512	0.179
	9.865	9.803	6.292	6.697	1.673	1.581	1.905	1.784
Portfolio of Government Bonds	−1.775	−0.881	−5.702	−5.435	6.091***	6.320***	5.808	4.432***
	6.179	5.633	4.631	4.276	1.340	1.333	2.461	1.556
ROE	0.859	0.283	0.469	0.816	0.015	0.193	0.114	0.160
	1.153	1.272	1.163	0.928	0.384	0.357	0.381	0.357
Constant	33.507*	24.077*	63.927***	63.451***	3.399	−0.681	−9.789	−13.724
	17.809	12.288	13.119	16.436	5.028	7.917	6.622	8.866
No. of Obs.	15,279	15,279	15,279	15,279	15,279	15,279	15,279	15,279

Notes: The table reports some robustness checks on the first step of our analysis. Specifications (1) and (3) replicate with changes specification (1) of tables 4 and 5, while specifications (2) and (4) replicate with changes specification (6) of tables 4 and 5. Estimation results are zero-inflated beta regression model results of both first and second part of the estimation (i.e., factors driving the choice to participate or not in CCP and factors influencing the intensity of the recourse conditional on participation). Observations are clustered at banking group level (and at bank level for independent banks), thus obtaining heteroskedasticity-robust standard errors and controlling for possible autocorrelations across the same banking group. The table reports regression coefficients and associated standard errors in italics. In the estimation of participation, a positive sign indicates a lower participation (more zeros) and a negative sign a higher participation (fewer zeros). ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

Table 8. Robustness Checks: Determinants of Δ(interbank exposures through CCPs)

	(1)	(2)	(3)	(4)	(5)	(6)
	Using CISS instead of the Ratio of Densities as a Measure of Uncertainty	**Dropping Time Fixed Effects**			**IV Estimations**	
Foreign Interbank Debts	0.010	−0.001	0.000	0.000	0.013	0.029
	0.043	*0.042*	*0.044*	*0.042*	*0.042*	*0.054*
UNC	−0.007	−0.003	−0.005	−0.001	0.343	0.330
	0.009	*0.008*	*0.010*	*0.007*	1.116	*1.119*
Size	0.031***	0.032***	0.033***	0.032***	0.030***	0.029***
	0.009	*0.010*	*0.010*	*0.009*	0.009	*0.009*
Retail Fundraising	0.056	0.068	0.067	0.067	0.054	0.061
	0.086	*0.085*	*0.097*	*0.085*	*0.085*	*0.088*
Loans to Private Sector	0.003	0.030	0.024	0.027	0.000	−0.001
	0.040	*0.037*	*0.037*	*0.037*	*0.040*	*0.040*
Central Bank Loans	0.041	0.039	0.008	0.046	0.043	0.047
	0.086	*0.080*	*0.079*	*0.079*	*0.086*	*0.085*
Portfolio of Government Bonds	0.352***	0.362***	0.334***	0.359***	0.353***	0.351***
	0.129	*0.121*	*0.113*	*0.119*	*0.129*	*0.129*
ROE	−0.018	−0.019	−0.017*	−0.017	−0.018	−0.018
	0.013	*0.012*	*0.011*	*0.011*	*0.013*	*0.013*
Δ(IRD)	−0.004**	−0.004**	0.002	0.002	−0.004**	−0.004**
	0.002	*0.002*	*0.002*	*0.003*	*0.002*	*0.002*
Δ(IRD) × Crisis 1			−0.003	0.002		
			0.002	*0.002*		
Δ(IRD) × Crisis 2			−0.017***	−0.008*		
			0.006	*0.004*		
Bad Loans	0.316	0.451**	0.459**	Included as deciles, and unreported	0.270	0.308
	0.243	*0.223*	*0.221*		*0.270*	*0.237*

(*continued*)

Table 8. (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)
	Using CISS instead of the Ratio of Densities as a Measure of Uncertainty	**Dropping Time Fixed Effects**			**IV Estimations**	
Δ(IRD) × Bad Loans (7° Decile) × Crisis 1				0.000		
				0.005		
Δ(IRD) × Bad Loans (7° Decile) × Crisis 2				0.006		
				0.007		
Δ(IRD) × Bad Loans (8° Decile) × Crisis 1				0.006		
				0.005		
Δ(IRD) × Bad Loans (8° Decile) × Crisis 2				−0.034		
				0.036		
Δ(IRD) × Bad Loans (9° Decile) × Crisis 1				−0.025**		
				0.012		
Δ(IRD) × Bad Loans (9° Decile) × Crisis 2				−0.002		
				0.009		
Δ(IRD) × Bad Loans (10° Decile) × Crisis 1				−0.018**		
				0.008		
Δ(IRD) × Bad Loans (10° Decile) × Crisis 2				−0.031**		
				0.017		
Constant	−0.288	−0.323***	−0.294**	−0.322***	−0.512	−0.507
	0.075	*0.083*	*0.102*	*0.081*	*0.804*	*0.803*
Rho	0.37	0.41	0.38	0.42	0.35	0.35
F test (First Stage)					19.48	22.45
No. of Obs.	11,008	11,008	11,008	11,008	11,008	11,008

Notes: The table reports some robustness checks on the second step of our analysis. Specifications (1), (2), (5), and (6) replicate with changes specification (1) of table 6; specification (3) replicates with changes specification (2) of table 6; and specification (4) replicates with changes specification (6) of table 6. In specifications (1)–(4), the table reports fixed-effects panel results where fixed effects are for banks; time fixed effects are included unless it is indicated differently. In specifications (5)–(6), the table reports IV estimation results alternating instrumental variables. Observations are clustered at banking group level (and at bank level for independent banks), thus obtaining heteroskedasticity-robust standard errors and controlling for possible autocorrelations across the same banking group. Partial interaction terms are always included even if unreported; in specification (4), the other deciles' results are not reported. The table reports regression coefficients and associated standard errors in italics. ***, **, and * denote statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

In particular, using CCPs can significantly reduce the size of the balance sheet items taken into account to compute regulatory requirements (Committee on the Global Financial System 2017). Our analysis is, however, not significantly affected by these requirements, for several reasons.

First, our sample period ends in June 2013 and while the Basel Committee on Banking Supervision mentioned the leverage ratio—the reform more relevant for repos—for the first time in December 2010 (as part of the Basel III package), the related proposal was then revised until January 2014. Moreover, even under the original package, new rules would apply only as from January 2014 (i.e., out of our sample period), with full implementation scheduled on January 2019. Bucalossi and Scalia (2016), indeed, confirm that banks started to adapt to the new requirement only in 2013 and 2014 and that there were no significant impacts on trading volumes on repo markets in the euro area in the period they examined.

To further corroborate our view that regulatory aspects were substantially irrelevant in our sample period, we checked for any evidence of "window-dressing" behavior due to regulation. This behavior would affect differentially both banks and months, as riskier banks would be those having more incentives to window dress and window-dressing would be concentrated at the end of a quarter when prudential requirements are computed. We interacted accordingly variables of banks' riskiness (*Bad Loans* or *Rating*) and bilateral relationship (ICC_{jt} and IRD_{jt}) with the time-dummy variables related to the months that are quarter-ends.[34] We added these interaction terms both in the analysis of the first step (determinants of the use of CCPs) and in the second step (use of CCP by riskier borrowers). In both cases, we found that results remain unaltered, and interacted terms are hardly significant and do not present any systematic patterns. Additionally, the fact that the Tier 1 ratio,[35] included among our independent variables, was not significant suggests that in our sample period regulatory requirements were not a main driver for the use of CCPs.

[34]We thank an anonymous referee for helping us to clarify the point and suggesting the exercise.

[35]The results are not reported in the tables but are available upon request.

Sample Split and Different Starting Dates for the Two Phases of the Financial Crisis. Regarding the impact of the crisis, we have included in estimations an interaction term between the regressors and two period dummies, *CR1* and *CR2*, which take the value of 1 during the corresponding phases of the crisis and 0 otherwise. As a check, instead of the two dummies and interactions, we have used a sample split repeating the same estimations before and after the onset of each crisis (regressions were run on three subperiods: up to 2008, from 2008 to 2011, and afterward). Results remain equivalent to those obtained with the interaction terms. In addition to time fixed effects, to test the sensitivity of results to different dates and periods, we altered the dating of the two crises with slightly different starting dates, bringing it forward and postponing it by one to four months.

Nonlinear Dynamics. Some of our results could be affected by nonlinear dynamics, in particular related to central bank liquidity provisions, which have been massively used by Italian banks during the crisis. We therefore added a higher-order term to the variable *Central Bank Loans*. Both variables (*Central Bank Loans* and its square) remain statistically nonsignificant in the regression explaining participation in CCPs (first stage of the zero-inflated beta regression model) while they are both significant in the regression on the intensity of the use of CCPs (second stage of the zero-inflated beta regression model). *Central Bank Loans* is statistically positive and the squared term is significantly negative. All the other results remain unchanged in substance when the two variables are added in the estimations. Interacting *Central Bank Loans* with other covariates did not lead to significant findings.

Instrumental-Variable Estimation. A concern regards the possible presence of reverse causality between our dependent variables in both models and the key bank-level regressors. This appears a possibility when we come to interbank bilateral characteristics (while we are not aware of channels through which the use of CCPs by a bank may determine its bad loans ratio). We tested the possible presence of reverse causality in two ways. First, we used standard, although not necessarily very powerful, tests such as the Durbin and Wu test and the Hausman test. For both variables, regressors turned out not to be endogenous. Second, we reestimated our regressions through an instrumental-variable method alternating different instruments. We adopted as instruments alternatively either the respective lags of regressors or, for the $Bilateral_{jt}$ regressors, liquidity shocks correlation between interbank counterparties.[36] In all cases, results remain the same. As an example, we report (table 8, specifications (5) and (6)) the same estimation of specification (1) of table 6 while using instrumental-variable estimations.

Alternative Definitions of Variables. As mentioned, we tested different definitions of our key variable, *IRD*, which counts in each period the integer number of months elapsed since the start of an interbank relationship between each pair of banks. Allowing a maximum of, respectively, zero, one, two, or three months of continuous interruption as a precondition to consider a relationship as ongoing does not lead to differences in our results.

Alternatively to the bad loans ratio, we measured the risk of each bank also with two additional variables: *Rating*, which is coded so as to take values from 1 to 11, where 1 corresponds to the best rating class and 10 to the worst, with 11 assigned to banks with no rating; and the dummy *Banks without Rating*, which takes the value of 1 for banks with no rating and 0 otherwise.[37] Finally, for the pair of variables *Rating* and *Banks without Rating,* we used an alternative approach avoiding the imposition of a linear structure to the relationship and introducing dummies for each score using the best score as the baseline level.

[36]Following Cocco, Gomes, and Martins (2009) and Affinito (2012), liquidity shocks correlation between interbank counterparties measures the correlation between the liquidity shocks of each pair of banks, and it is computed as a correlation between the volatility of balance sheet items measuring banking liquidity. Cocco, Gomes, and Martins (2009) and Affinito (2012) show that this variable matters for the existence and persistence of interbank customer relationships.

[37]The two variables are always included simultaneously in order not to lose observations on nonrated banks while allowing the ad hoc dummy to control for nonrated banks: this setting avoids the score "11" attributed to nonrated banks implying a worse assessment than the score "10" attributed to the riskiest banks receiving a rating (e.g., Angelini, Nobili, and Picillo 2011). Credit scores are taken from Fitch, as Angelini, Nobili, and Picillo (2011) find that Fitch ratings are more informative in the assessment of banks and financial firms. All credit ratings are obtained as a monthly average of the daily overall individual rating.

8. Conclusions

During the global financial crisis Italian banks remarkably increased their use of CCPs for interbank funding, a move that lessened uncertainty and avoided the substantial freezing of the interbank market experienced in other jurisdictions. The growing role of CCPs in interbank market might, however, add a specific risk, namely to allow riskier borrowers to elude peer monitoring, recurring to anonymous transactions via CCPs, and increase the counterparty risk borne by CCPs.

We focused our analysis on this issue, and we find that both uncertainty and banks' risk were significant drivers of the increased recourse to CCPs. Our results further suggest that for the riskiest banks the recourse to the CCPs during the crisis was likely driven by difficulties in borrowing in the bilateral interbank market due to their risk.

Overall, our findings support the policy efforts to ensure that CCPs put in place adequate risk control frameworks and suggest an additional reason why this effort should remain high in the policy agenda.

References

Acharya, V., and A. Bisin. 2014. "Counterparty Risk Externality: Centralized versus Over-the-Counter Markets." *Journal of Economic Theory* 149 (January): 153–82.

Affinito, M. 2012. "Do Interbank Customer Relationships Exist? And How Did They Function in the Crisis? Learning from Italy." *Journal of Banking and Finance* 36 (12): 3163–84.

———. 2013. "Central Bank Refinancing, Interbank Markets, and the Hypothesis of Liquidity Hoarding: Evidence from a Euro-Area Banking System." Working Paper No. 928, Banca d'Italia.

Affinito, M., and A. F. Pozzolo. 2017. "The Interbank Network across the Global Financial Crisis: Evidence from Italy." *Journal of Banking and Finance* 80 (July): 90–107.

Allen F., E. Carletti, and D. Gale. 2009. "Interbank Market Liquidity and Central Bank Intervention." *Journal of Monetary Economics* 56 (5): 639–52.

Angelini P., R. Di Salvo, and G. Ferri. 1998. "Availability and Cost of Credit for Small Business: Customer Relationships and Credit Cooperatives." *Journal of Banking and Finance* 22 (6–8): 925–54.

Angelini P., A. Nobili, and M. C. Picillo. 2011. "The Interbank Market after August 2007: What Has Changed, and Why?" *Journal of Money, Credit and Banking* 43 (5): 923–58.

Arnsdorf, M. 2012. "Quantification of Central Counterparty Risk." *Journal of Risk Management in Financial Institutions* 5 (3): 273–87.

Baklanova, V., O. Dalton, and S. Tompaidis. 2017. "Benefits and Risks of Central Clearing in the Repo Market." Brief Series No. 17-04 (March 9), Office of Financial Research.

Banca d'Italia. 2013a. *Abridged Report for the Year 2012*. Rome.

———. 2013b. *Financial Stability Report*, No. 6. Rome.

Basel Committee on Banking Supervision (BCBS), Committee on Payments and Market Infrastructures (CPMI), Financial Stability Board (FSB), and International Organization of Securities Commissions (IOSCO). 2017. "Chairs' Report on the Implementation of the Joint Workplan for Strengthening the Resilience, Recovery and Resolvability of Central Counterparties." Bank for International Settlements.

Bernanke, B. S. 1990. "Clearing and Settlement during the Crash." *Review of Financial Studies* 3 (1): 133–51.

Biais, B., F. Heider, and M. Hoerova. 2012. "Clearing, Counterparty Risk, and Aggregate Risk." *IMF Economic Review* 60 (2): 193–222.

———. 2013. "Incentive Compatible Centralised Clearing." *Financial Stability Review* (Banque de France) 17 (April): 161–68.

———. 2016. "Risk-sharing or Risk-taking? Counterparty Risk, Incentives, and Margins." *Journal of Finance* 71 (4): 1669–98.

Bodenhorn, H. 2003. "Short-Term Loans and Long-Term Relationships: Relationship Lending in Early America." *Journal of Money, Credit and Banking* 35 (August): 485–505.

Bonaccorsi di Patti E. and G. Gobbi. 2007. "Winners or Losers? The Effects of Banking Consolidation on Corporate Borrowers." *Journal of Finance* 62 (2): 669–95.

Borio, C. 2004. "Market Distress and Vanishing Liquidity: Anatomy and Policy Options." BIS Working Paper No 158.

Brunnermeier, M. K. 2009. “Deciphering the Liquidity and Credit Crunch 2007–2008.” *Journal of Economic Perspectives* 23 (1): 77–100.

Bucalossi, A., and A. Scalia. 2016. “Leverage Ratio, Central Bank Operations and Repo Market.” Questioni di Economia e Finanza No. 347 (July), Banca d’Italia.

Calomiris, C. W., and C. M. Kahn. 1991. “The Role of Demandable Debt in Structuring Optimal Banking Arrangements.” *American Economic Review* 81 (3): 497–513.

Cappelletti, G., A. De Socio, G. Guazzarotti, and E. Mallucci. 2011. “The Impact of the Financial Crisis on Inter-bank Funding: Evidence from Italian Balance Sheet Data.” Questioni di Economia e Finanza No. 95, Banca d’Italia.

Cappelletti, G., and G. Guazzarotti. 2017. “The Role of Counterparty Risk and Asymmetric Information in the Interbank Market.” ECB Working Paper No. 2022 (February).

Cecchetti, S. G., J. Gyntelberg, and M. Hollanders. 2009. “Central Counterparties for Over-the-Counter Derivatives.” *BIS Quarterly Review* (September): 45–58.

Chatterjee, P., L. Embree, and P. Youngman. 2012. “Reducing Systemic Risk: Canada’s New Central Counterparty for the Fixed-Income Market.” *Financial System Review* (Bank of Canada) 43–49.

Cœuré, B. 2014. “Risks in Central Counterparties.” Speech at the policy panel during the conference “Mapping and Monitoring the Financial System: Liquidity, Funding, and Plumbing,” Office of Financial Research and Financial Stability Oversight Council, Washington D.C., January 23.

Cocco, J. F., F. J. Gomes, and N. C. Martins. 2009. “Lending Relationships in the Interbank Market.” *Journal of Financial Intermediation* 18 (1): 24–48.

Committee on the Global Financial System. 2017. “Repo Market Functioning.” CGFS Papers No. 59 (April).

Committee on Payment and Settlement Systems (CPSS) and Technical Committee of the International Organization of Securities Commissions (IOSCO). 2012. *Principles for Financial Market Infrastructures.* Bank for International Settlements.

Committee on Payments and Market Infrastructures (CPMI) and Board of the International Organization of Securities Commissions (IOSCO). 2014. *Recovery of Financial Market Infrastructures.* Bank for International Settlements.

———. 2016. *Resilience and Recovery of Central Counterparties (CCPs): Further Guidance on the PFMI — Consultative Report.* Bank for International Settlements.

Cook, D., R. Kieschnick, and B. D. McCullough. 2008. "Regression Analysis of Proportions in Finance with Self Selection." *Journal of Empirical Finance* 15 (5): 860–67.

de Haas, R., and I. van Lelyveld. 2010. "Internal Capital Markets and Lending by Multinational Bank Subsidiaries." *Journal of Financial Intermediation* 19 (1): 1–25.

Dinger, V., and J. von Hagen. 2009. "Does Interbank Borrowing Reduce Bank Risk?" *Journal of Money, Credit and Banking* 41 (2–3): 491–506.

Distinguin, I., C. Roulet, and A. Tarazi. 2013. "Bank Regulatory Capital and Liquidity: Evidence from US and European Publicly Traded Banks." *Journal of Banking and Finance* 37 (9): 3295–3317.

Duffie, D. 2019. "Prone to Fail: The Pre-Crisis Financial System." *Journal of Economic Perspectives* 33 (1, Winter): 81–106.

Duffie, D., M. Scheicher, and G. Vuillemey. 2015. "Central Clearing and Collateral Demand." *Journal of Financial Economics* 116 (2): 237–56.

Elsas, R. 2005. "Empirical Determinants of Relationship Lending." *Journal of Financial Intermediation* 14 (1): 32–57.

European Central Bank. 2007. *The Role of Central Counterparties: Issues Related to Central Counterparty Clearing.* Proceedings of the ECB-FED Chicago Conference, Frankfurt, Germany, April 3–4, 2006.

———. 2012. "The Adjustment of Monetary Statistics for Repurchase Agreement Transactions with Central Counterparties." *Monthly Bulletin* (September): 28–31 (Box 3).

European Securities and Markets Authority. 2017. *ESMA Final Report — Technical Standards under SFTR and Certain Amendments to EMIR* (March).

Financial Stability Board. 2013. *Strengthening Oversight and Regulation of Shadow Banking: Policy Framework for Addressing Shadow Banking Risks in Securities Lending and Repos* (August).

———. 2015. *OTC Derivatives Market Reforms: Tenth Progress Report on Implementation* (November).

———. 2017. *Review of OTC Derivatives Market Reforms: Effectiveness and Broader Effects of the Reforms* (June).

Financial Stability Forum. 2008. *Enhancing Market and Institutional Resilience: Follow-up on Implementation.*

Freixas, X., A. Martin, and D. Skeie. 2011. "Bank Liquidity, Interbank Markets, and Monetary Policy." *Review of Financial Studies* 24 (8): 2656–92.

Furfine, C. H. 2001. "Banks as Monitors of Other Banks: Evidence from the Overnight Federal Funds Market." *Journal of Business* 74 (1): 33–58.

———. 2004. "Public Disclosures and Calendar-Related Movements in Risk Premiums: Evidence from Interbank Lending." *Journal of Financial Markets* 7 (1): 97–116.

———. 2009. "The Liquidity of Liquid Markets during the Financial Crisis." Mimeo.

Garcia-de-Andoain C., F. Heider, M. Hoerova, and S. Manganelli. 2016. "Lending-of-Last-Resort Is as Lending-of-Last-Resort Does: Central Bank Liquidity Provision and Interbank Market Functioning in the Euro Area." *Journal of Financial Intermediation* 28 (October): 32–47.

Gobbi, G., and E. Sette. 2014. "Do Firms Benefit from Concentrating their Borrowing? Evidence from the Great Recession." *Review of Finance* 18 (2): 527–60.

———. 2015. "Relationship Lending During A Financial Crisis." *Journal of the European Economic Association* 13 (3): 453–81.

Hansen, J., and A. Moore. 2016. "The Efficiency of Central Clearing: A Segmented Markets Approach." Research Discussion Paper No. 7, Reserve Bank of Australia.

Hardouvelis, G., and S. Peristiani. 1992. "Margin Requirements, Speculative Trading, and Stock Price Fluctuations: The Case of Japan." *Quarterly Journal of Economics* 107 (4): 1333–70.

Heider F., M. Hoerova, and C. Holthausen. 2015. "Liquidity Hoarding and Interbank Market Spreads: The Role of Counterparty Risk." *Journal of Financial Economics* 118 (2): 336–54.

Heller, D., and N. Vause. 2012. "Collateral Requirements for Mandatory Clearing of Over-the-Counter Derivatives." BIS Working Paper No. 373.

Holló, D., M. Kremer, and M. Lo Duca. 2012. "CISS — A Composite Indicator of Systemic Stress in the Financial System." ECB Working Paper No. 1426.

Houston, J. F., C. James, and D. Marcus. 1997. "Capital Market Frictions and the Role of Internal Capital Markets in Banking." *Journal of Financial Economics* 46 (2): 135–64.

Huang, R., and L. Ratnovski. 2008. "The Dark Side of Bank Wholesale Funding." Working Paper No. 09-3, Federal Reserve Bank of Philadelphia.

International Monetary Fund. 2013. "European Union: Publication of Financial Sector Assessment Program Documentation—Technical Note on Financial Integration and Fragmentation in the European Union." Staff Country Report No. 13/71.

Jackwerth, J. 2000. "Recovering Risk Aversion from Option Prices and Realized Returns." *Review of Financial Studies* 13 (2): 433–51.

Jones, R., and C. Perignon. 2013. "Derivatives Clearing, Default Risk, and Insurance." *Journal of Risk and Insurance* 80 (2): 373–400.

King, T. B. 2008. "Discipline and Liquidity in the Interbank Market." *Journal of Money, Banking and Credit* 40 (2–3): 295–317.

Koeppl, T. 2012. "Central Counterparty Clearing: Incentives, Market Discipline and the Cost of Collateral." Working Paper (February), Queen's University.

Koeppl, T., and C. Monnet. 2010. "The Emergence and Future of Central Counterparties." Working Paper No. 10-30, Federal Reserve Bank of Philadelphia.

Leitner, Y. 2012. "Inducing Agents to Report Hidden Trades: A Theory of an Intermediary." *Review of Finance* 16 (4): 1013–42.

Loon, Y. C., and Z. K. Zhong. 2014. "The Impact of Central Clearing on Counterparty Risk, Liquidity, and Trading: Evidence from the Credit Default Swap Market." *Journal of Financial Economics* 112 (1): 91–115.

Mancini, L., A. Ranaldo, and I. Wrampelmeyer. 2016. "The Euro Interbank Repo Market." *Review of Financial Studies* 29 (7): 1747–79.

Papke, L., and J. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics* 11 (6): 619–32.

Petersen, M., and R. Rajan. 1994. "The Benefits of Lending Relationships: Evidence From Small Business Data." *Journal of Finance* 49 (1): 3–37.

Pirrong, C. 2011. "The Economics of Central Clearing: Theory and Practice." ISDA Discussion Paper No. 1 (May).

Rochet, J. C., and J. Tirole. 1996. "Interbank Lending and Systemic Risk." *Journal of Money, Credit and Banking* 28 (November): 733–62.

Sidanius, C., and F. Zikes. 2012. "OTC Derivatives Reform and Collateral Demand Impact." Financial Stability Paper No. 18, Bank of England.

Stephens, E., and J. Thompson. 2011. "CDS as Insurance: Leaky Lifeboats in Stormy Seas." Working Paper, University of Alberta and University of Waterloo.

Tarashev, N., K. Tsatsaronis, and D. Karampatos. 2003. "Investors' Attitude towards Risk: What Can We Learn from Options?" *BIS Quarterly Review* (June): 57–65.

Taylor, J. B., and J. C. Williams. 2009. "A Black Swan in the Money Market." *American Economic Journal: Macroeconomics* 1 (1): 58–83.

Thompson, J. 2010. "Counterparty Risk in Financial Contracts: Should the Insured Worry about the Insurer?" *Quarterly Journal of Economics* 125 (3): 1195–1252.

Pension Funds' Herding*

Dirk W.G.A. Broeders,[a,b] Damiaan H.J. Chen,[a,c]
Peter A. Minderhoud,[a] and C.J. Willem Schudel[a]
[a]De Nederlandsche Bank
[b]Maastricht University
[c]University of Amsterdam

This paper uses unique and detailed transaction data to analyze herding behavior among pension funds. We distinguish between weak, semi-strong, and strong herding behavior. Weak herding occurs if pension funds have similar rebalancing strategies. Semi-strong herding arises when pension funds react similarly to other external shocks, such as changes in regulation and exceptional monetary policy operations. Finally, strong herding means that pension funds intentionally replicate changes in the strategic asset allocation of other pension funds without an economic reason. We find empirical evidence supporting all three types of herding behavior in the asset allocation of large Dutch pension funds. Whereas weak herding can contribute to financial stability, strong herding may present a risk for financial stability.

JEL Codes: G11, G23.

1. Introduction

In this paper we use unique and detailed transaction data to analyze herding behavior among institutional investors using a rebalancing model based on Calvet, Campbell, and Sodini (2009) in combination with a spatial estimation approach. Nofsinger and Sias (1999) define herding as a group of investors trading in the same direction

*We would like to thank Jack Bekooij, David Keijzer, and Dirk van der Wal for excellent research assistance, and Roel Beetsma, Muriël van den Berg, Renske Biezeveld, Patty Duijm, Matteo Bonetti, Janko Gorter, Franc Klaassen, Ward Romp, Rutger Teulings, Sweder van Wijnbergen, and the DNB seminar participants for helpful comments and suggestions. Corresponding author (Chen): University of Amsterdam, MInt/Amsterdam School of Economics, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands; e-mail: D.H.J.Chen@UvA.nl.

over a period of time. In order to analyze this thoroughly, we distinguish between weak, semi-strong, and strong herding behavior. Weak herding is related to the information motive in the literature, semi-strong herding to the regulation motive, and strong herding to the reputation motive. We document empirical evidence to support all these types of herding in the asset allocation of large Dutch pension funds. Our findings have potential implications for policymakers who are interested in financial stability. Whereas weak herding can contribute to financial stability, strong herding is a risk for financial stability if pension funds deliberately replicate each other's investment strategies without economic reason. Furthermore, regulators need to be aware that semi-strong herding might imply that pension funds react in a similar way to regulatory changes.

Global asset portfolios of institutional investors, such as pension funds, have grown substantially over the past decades. Economic and financial policymakers around the globe have therefore become increasingly interested in the factors driving the allocation of these assets. One of the main motivations behind asset allocation decisions that receives increasing attention from global policymaking institutes is investor herding behavior. The International Monetary Fund does multiple studies on this phenomenon, e.g., Bikhchandani and Sharma (2001); Papaioannou et al. (2013); Cipriani and Guarino (2014); Jones (2015). Also the World Bank analyzes herding behavior (Raddatz and Schmukler 2011), as well as the Federal Reserve (Chari and Kehoe 2002; Cai, Han, and Li 2012; Chari and Phelan 2014) and the Bank for International Settlements (Borio, Furfine, and Lowe 2001; Nirei, Stamatiou, and Sushko 2012).

A key reason why these institutions study herding is its potential implications for financial stability. The European Insurance and Occupational Pensions Authority (EIOPA) provides evidence that pension funds contribute to financial stability as a result of rebalancing strategies (EIOPA 2016). Since most pension funds aim for a more or less fixed asset allocation within a narrow bandwidth, they typically will buy equities following a period in which the equity allocation decreased. The latter will be driven by relative price effects or exchange rate effects in the prior period(s). Also Bohl, Brzeszczynski, and Wilflin (2009) and Thomas, Spataro, and Mathew (2014) find that institutional investors such as pension funds dampen stock market volatility. The Office of Financial

Research in the United States identifies asset managers' herding as one of the key vulnerabilities to financial stability (Elliot 2014). If asset managers enter, e.g., into fire sales simultaneously, this can have an amplifying effect on asset price volatility. The Bank of England also comments on this phenomenon, relating it to the fact that more and more pension funds delegate the management of their assets to external parties (Haldane 2014). This outsourcing gives rise to the question of whether pension funds' asset allocation decisions are interdependent.

We specifically look at herding behavior among pension funds that, because of their size, are important institutional investors in financial markets. On the one hand, pension funds are long-term investors that are able to pursue an optimal long-term investment strategy to the best interest of the pension fund's beneficiaries. This may also contribute to financial market stability, as pension funds can offer liquidity in times of financial markets stress. On the other hand, pension funds are typically constraint investors, e.g., by the size and the nature of the liabilities, the risk preferences of the key stakeholders, and by external regulation. Pension funds can also feel a constraint from peer-group pressure. They may want to invest closely in line with other pension funds to avoid the reputation risk of having to report strongly deviating investment returns.

This paper distinguishes between three types of herding. We define *weak herding* as the result from the fact that pension funds have similar rebalancing strategies. Most pension funds operate in this way (Calvet, Campbell, and Sodini 2009; Bikker, Broeders, and de Dreu 2010; Gorter and Bikker 2013). This behavior is inherent to the investment strategy of pension funds, and the transactions resulting from the rebalancing strategy are not necessarily a form of herding in the sense that pension funds deliberately mimic the transactions of other pension funds. This unintentional or spurious form of herding occurs because groups face similar decision problems and information sets and make similar decisions (Bikhchandani and Sharma 2001). *Semi-strong herding* arises if pension funds react similar to external shocks, e.g., changes in pension fund regulation. Sias (2004) and Andonov, Bauer, and Cremers (2017), e.g., show that regulation can have a significant impact on pension funds' investment decisions. We define *strong herding* as a case in which pension funds intentionally copy the investment decisions of other pension funds without a distinct economic reason. This could, e.g., be the case if a group of pension funds follow changes in the strategic asset allocation of another pension fund or a group of pension funds. In this type of herding, an informed agent follows the trend even though that trend is counter to his initial information about an asset class (Avery and Zemsky 1998). Strong herding may occur through trustees, actuaries, or asset managers who provide services to multiple pension funds (Bauer, Bonetti, and Broeders 2020). Whereas weak herding can contribute to financial stability, strong herding is a risk for financial stability.

This paper seeks to shed light on herding behavior among Dutch defined-benefit funds. The Dutch pension system is an interesting case study for several reasons. First, it is relatively large in terms of its size: its total assets represent roughly twice the size of the gross domestic product (GDP) of the Netherlands. The investment behavior of these pension funds is therefore of significant importance to financial stability. Second, during the Great Financial Crisis and thereafter, most pension funds in the Netherlands suffered considerable decreases in their funding ratios. Indeed, pension funds' funding ratios (as defined by the ratio of total assets over liabilities) moved largely in tandem. This was fueled by the impact of changes in the term structure of interest rates on the value of the liabilities. But also the assets have been hit in a similar way, as pension funds all have very broadly diversified investment portfolios. Their returns will therefore be very similar.

We examine the extent to which these pension funds follow one another in terms of changing their asset allocation. We use a unique data set from De Nederlandsche Bank (DNB), containing monthly transaction data of large Dutch occupational pension funds across a period from January 2009 until January 2015. To test our hypotheses, we employ an econometric specification based on a rebalancing model in combination with a spatial estimation approach. The latter, although common in the political economy literature (see, e.g., Beck, Gleditsch, and Beardsley 2006; Franzese and Hays 2007), is to the best of our knowledge a novelty in the pension economics literature. This approach enables us to estimate the spatial dependence of pension funds' equity and bond allocations. We also check the robustness of our results using an alternative model specification based on the Error Correction Model (Engle and Granger 1987).

The remainder of this paper is organized as follows. Section 2 reviews motivations in the literature for herding behavior among asset managers. Section 3 introduces the hypotheses that we will test, while section 4 describes our data. In section 5 we lay out the model for our empirical analysis. The results are discussed in section 6. In section 7, we replicate the analysis using an alternative regression model to check for robustness of the obtained results. Section 8 concludes.

2. Motives for Herding Behavior

There is an extensive body of theoretical and empirical literature on institutional herding behavior. Institutional investors may exhibit herding behavior for a number of reasons. Bikhchandani and Sharma (2001) mention three motives for herding behavior: information-based herding, compensation-based herding, and reputation-based herding. We present an almost similar classification of motives, distinguishing between an information motive, a regulation motive, and a reputation motive. Moreover, we apply an ordering to these motives, reclassifying the information motive as weak herding, the regulation motive as semi-strong herding, and the reputation motive as strong herding behavior. Weak herding is unintentional, while strong herding is intentional. All are discussed in more detail below.

2.1 Information Motive (Weak Herding)

We define weak herding behavior as the result from the fact that pension funds have similar rebalancing strategies. Investors typically rely on similar sources of information when they make investment decisions. The information can, for instance, be market signals such as the returns on different asset classes. This can lead to herding behavior, which we classify as weak because it is an unintentional consequence of being exposed to similar information. Typically, pension funds have a rebalancing strategy, by aiming for a fixed asset allocation (Calvet, Campbell, and Sodini 2009; Bikker, Broeders, and de Dreu 2010; Rubbaniy, van Lelyveld, and Verschoor 2012; Gorter and Bikker 2013). Blake, Sarno, and Zinna (2017) report short-term mechanical portfolio rebalancing by U.K. pension funds. Also EIOPA documents that pension funds typically have rebalancing strategies (EIOPA 2016). This way, pension funds counteract changes in the asset allocation due to valuation changes in the different asset classes. Since pension funds are exposed to similar market risks, this results in trades into similar directions. Hence, this unintentional herding occurs because pension funds face similar decision problems and information sets (Bikhchandani and Sharma 2001). For example, Rauh (2006) identifies the dependence of investments for defined-benefit pension plans, particularly when they are financially constrained. Very similar, the rising popularity of "index tracking" also leads to herding behavior among institutional investors. Gleason, Mathur, and Peterson (2004); Chen et al. (2011), and Shek, Shim, and Shin (2018) document herding behavior in the market for exchange traded funds (ETFs).

2.2 Regulation Motive (Semi-Strong Herding)

Semi-strong herding arises if pension funds react similarly to external shocks, e.g., changes in pension fund regulation. Pension funds that are subject to the same regulation may choose similar asset allocations, which can result in herding. If the price of risk in regulation makes some asset classes with specific characteristics more attractive to investors, those investors may have an incentive to adjust their asset allocations in the same way (Sias 2004). On the other hand, regulation can cause investors to dislike some other asset classes with certain characteristics. These preferences or aversions for assets with specific characteristics can be measured from changes in regulation. We classify this as semi-strong herding, because in this case pension funds actively make an investment decision following specific changes in circumstances that relate to them. In the literature some examples can be found of this so-called characteristic herding. Severinson and Yermo (2012) show that the introduction of risk-based solvency standards resulted in an increased demand for government bonds by Swiss insurance companies in 2006. Another example is the shift from equities to bonds by U.K. pension funds due to the introduction of fair value accounting in Financial Reporting Standard 17 (FRS 17) in 2003 (Amir, Guan, and Oswald 2010). In addition, Andonov, Bauer, and Cremers (2017) show that Government Accounting Standards Board (GASB) regulation of U.S. public pension funds favors equity investments, as the level of the liability discount rate is derived from the expected return on assets. U.S. public pension funds can artificially improve their financial position by investing in more risky assets. Of course, the introduction of new accounting or regulatory standards does not necessarily lead to shifts in investors' allocations. For example, Amir, Guan, and Oswald (2010) also find that the introduction of fair value accounting for corporate pensions funds in the United States (Statement of Financial Accounting Standards 158 in 2006) did not have pronounced effects in asset allocations.

2.3 Reputation Motive (Strong Herding)

We define strong herding behavior as a case in which pension funds intentionally copy the investment decisions of other pension funds. Reputation-based or strong herding therefore occurs when pension funds actively react to the investment behavior of others without an economic reason. We distinguish two subclasses: career pressure and peer-group pressure. Scharfstein and Stein (1990) claim that, due to career pressure, managers will "follow the herd" if they are concerned about how others will assess their ability to make judgments. In other words, asset managers may be concerned about their labor market position and therefore may choose to mimic investing behavior of other asset managers. Prendergast and Stole (1996) show that reputation herding can be regarded as an inefficient handling of information due to concerns on the reputation of the investor himself. In an ideal world, every individual would behave like a rational Bayesian, optimally learning about the economic environment by correctly combining new information with prior knowledge and then using this information to maximize value. However, actors deviate from this efficient behavior because they care about their reputation. Moreover, Prendergast and Stole (1996) show that young investment managers want to emphasize their learning capacities by exaggerating the importance of new information, while old managers are less willing to change their behavior based on new information because they do not want to suggest their previous behavior was wrong. Dasgupta, Prat, and Verardo (2011) document that career-concerned asset managers exhibit the tendency to replicate past trades. Moreover, they prove that this has an effect on pricing: dealers take advantage of a manager's reputation motivation by offering trades above expected liquidation values based on available information. Managers typically are willing to pay excessively high prices because they expect a reputation reward. Nofsinger and Sias (1999) show that institutional investors are more prone to herding behavior than individual investors. This could indicate the presence of a labor market incentive among institutional investors.

The second subclass of reputation herding is peer-group pressure. This occurs if the risk-taking behavior of an individual asset manager is affected by the risk-taking behavior of other managers in his peer group (Graham 1999). In this case an asset manager chooses to ignore his private information and mimic the actions of another asset manager. The reputation of the other asset manager is then thought to be superior over the asset manager's private information. In following the herd and neglecting private information, reputation herding is a bit similar to herding on informational cascades. However, reputation herding models have an additional layer of mimicking which results from positive reputation externalities that can be obtained by acting as part of a group (Graham 1999). Investors can infer information from the trades of other asset managers. Banerjee (1992) describes this behavior as rational for an individual investor, as the other investors have relevant information for him. The author, however, shows that the equilibrium is inefficient if all investors use information of others instead of their own.

2.4 Risks and Costs of Herding

Herding behavior has potential consequences for market volatility. A classic example is the creation of price bubbles (Avery and Zemsky 1998; Brunnermeijer and Nagel 2004; Hott 2009). Bubbles can arise when rational investors neglect their own private information because they believe that most other traders have very accurate information, while the latter are in fact poorly informed. Jacklin, Kleidon, and Pfleiderer (1992) show that lack of perfect information by investors about the quality of the information possessed by other traders explains the stock market crash of 1987. Also Bikhchandani, Hirshleifer, and Welch (1992) explain short-term bubbles and bursts from informational cascades that occur when individuals follow the behavior of others without regarding their own information. Investors who decide early may be crucial in determining which way the majority will decide. If it turns out, e.g., when new information arrives, that investors have made a wrong decision, they are likely to start herding in the opposite direction. This increases market volatility (Bikhchandani and Sharma 2001). Hirshleifer, Subrahmanyan, and Titam (1994) analyze under which conditions investors find it more profitable to collect information on stocks that are followed by many investors, instead of comparable stocks that are being ignored by the investor community. These cases in which investors infer information from the trades of other asset managers can lead to strong herding behavior.

Herding behavior comes at a cost. Wei, Wermers, and Yao (2015) show that contrary investors benefit from providing liquidity to herding asset managers by trading against them. Froot, Scharfstein, and Stein (1992) find that in markets with short-term trading there may be information inefficiencies in which positive spillovers arise: in these cases it turns out to be rewarding for short-term investors to herd by focusing "too much" on some types of information, while neglecting other types. The reason is that if more short-term speculators study a given set of information, then more of that information disseminates in the market and, as a consequence, profits increase from obtaining a specific set of information at an early stage.

3. Testable Hypotheses

We focus our analysis on changes in equity and bond allocations of the pension funds in our sample. We test for weak, semi-strong, and strong herding in turn. Weak herding can be assessed by investigating how pension funds rebalance their asset allocation over time. Our first hypothesis is that weak herding exists. Since all pension funds will have some rebalancing policy, we expect to find a spurious relation between pension funds. In addition to that, all pension funds have well-diversified exposures on global equity and bond markets and will experience similar market returns. Rebalancing is primarily driven by past returns. Several papers describe the impact of past returns on asset allocation. Blake, Lehmann, and Timmermann (1999) find evidence of rebalancing under 300 U.K. pension funds aimed to stabilize the actual asset allocation around strategic asset allocation. Rauh (2009) finds that high past equity returns lead to higher equity allocations and consequently lower allocations to bonds and cash for U.S. corporate pension plans. However, the equity allocations do not move as far as they would if there had been no rebalancing, implying that the pension funds have some rebalancing policy. Pennacchi and Rastad (2011) report evidence that U.S. state and local government pension funds increase portfolio risk compared with the liabilities following periods of relatively poor investment performance. Mohan and Zhang (2014) also find that public pension funds take more investment risk after lower investment returns in the previous years. Obviously, rebalancing is not done continuously. In practice, the rebalancing behavior of pension funds allows for so called free-floating. Bikker, Broeders, and de Dreu (2010) describe two forms of free-floating. The first is calendar rebalancing, whereby pension funds rebalance their portfolio back to its strategic weights at regular intervals. The second refers to band rebalancing, whereby pension funds create a bandwidth around the strategic weight of each asset class and rebalance their portfolio if the weight of one asset class breaches its band.

Second, we test for semi-strong herding by testing how pension funds act upon exogenous shocks. We hypothesize that changes in regulation will affect the asset allocation of pension funds in similar directions. From the literature we know that pension fund investments are at least to some extent driven by regulation. We identify key changes in pension regulation and document the change in equity and bond allocations around (the announcement of) the change. The regulatory incentives for Dutch pension funds in our sample are mixed. First, liabilities in defined-benefit plans are valued using the term structure of risk-free market interest rates. This implicitly favors government bonds, swaps, and other fixed-income securities as appropriate asset classes. However, Dutch pension funds typically run an asset-liability mismatch by investing partially in risky assets. The risk premium on these assets can be used to index pension benefits to inflation (Broeders et al. 2014). Second, regulation allows Dutch pension funds to always rebalance their asset allocation toward their strategic portfolio weights. This also holds for pension funds with a funding shortfall, i.e., a funding ratio less than 105 percent. However, in this case pension funds are not allowed to "uprisk." They cannot increase their risk profile in excess of the risk profile of the strategic asset allocation. That would be considered a case of gambling for resurrection.[1] We therefore highlight that Dutch pension funds are not forced by regulation to "de-risk" during financial market stress.

Third, we test for strong herding. We hypothesize that pension funds do not want to underperform vis-à-vis their peers, as they are regularly exposed in the news concerning their funding ratio. Therefore they have an incentive to actively follow changes in the asset allocation of their peers. For this we test if pension funds copy the changes in the strategic investment behavior of other pension funds.

4. Data Description

In this section we first describe the structure of the data in section 4.1. Thereafter, we analyze the risk and return characteristics in section 4.2 and the proxy asset allocation and explanatory variables in section 4.3.

4.1 Structure of the Data

We use monthly transaction data that is sourced from the balance of payments statistics of DNB, which is the Dutch central bank. The primary data used are the pension fund's detailed investment holdings in individual equities and bonds. The holdings are uniquely identified according to their International Securities Identification Number (ISIN). The transaction data show the so-called direct investments of pension funds in securities. Pension funds can, however, also invest indirectly in equities and bonds through investment trusts. We also have ISIN data on the investments of these investment trusts. However, except for the two largest pension funds in the sample, we do not have information on which pension funds invest in which investment trusts. Therefore, only for the two largest pension funds can we merge the investment trusts with the pension fund data. Because of liquidations and mergers of pension funds, the length of sample period of each pension fund varies in the sample, particularly for corporate pension funds.

[1] In 2015 a new Pension Act was introduced. As part of this introduction, pension funds were allowed to increase their risk profile once, under specific solvency conditions.

We do not analyze the ISIN records directly. Instead we use aggregated transaction data for equities, bonds, and investment trusts at the pension fund level. Hence, we aggregate the data for each of the three investment classes $j = \{1, 2, 3\}$, for which the following data entries are available:

1. $PB^j_{i,t}$: position at the beginning of the month,
2. $Pur^j_{i,t}$: purchases during the month,
3. $Sal^j_{i,t}$: sales during the month,
4. $\triangle Pr^j_{i,t}$: price changes during the month,
5. $\triangle FX^j_{i,t}$: exchange rate changes during the month,
6. $\triangle OC^j_{i,t}$: other changes during the month,
7. $PE^j_{i,t}$: position at the end of the month,

with pension fund $i = \{1, 2, \ldots, I\}$ and month $t = \{1, 2, \ldots, T\}$. The data set that we analyze contains $I = 39$ large Dutch pension funds over a period that stretches across $T = 73$ months, from January 2009 until January 2015. After deleting those combinations for which we have no or imperfect data, we end up with an unbalanced panel of $N = 2{,}299$ observations.[2] The deletions are specified in appendix A. The panel covers 18 industry-wide pension funds ("bedrijfstakpensioenfondsen"), 16 corporate pension funds ("ondernemingspensioenfondsen"), and 5 professional group pension funds ("beroepspensioenfondsen"). Industry-wide pension funds provide pension services to a specific sector or industry, including public sectors. Industry-wide pension funds are typically mandatory. Corporate pension funds operate for a single company. A professional group pension fund is organized for a specific group of professions such as doctors and pharmacists. The data set covers more

[2]Both the first months and the last months contain all $I = 39$ pension funds. Hence, there is no bias concerning the existence of the pension funds in the data set that we analyze.

than 70 percent of total assets under management in the Dutch occupational pension sector.

The values in entries 1 through 7 satisfy two basic rules. First, the market value of the position at the end of this month equals the position at the beginning of the next month, so

$$PE^j_{i,t} = PB^j_{i,t+1}. \tag{1}$$

Second, the entries in 1 through 7 comply to the following identity relation for each period:

$$PE^j_{i,t} = PB^j_{i,t} + Tr^j_{i,t} + \triangle Pr^j_{i,t} + \triangle FX^j_{i,t} + \triangle OC^j_{i,t}, \tag{2}$$

where the net transactions $\left(Tr^j_{i,t}\right)$ is the difference between the sales and the purchases during the month

$$Tr^j_{i,t} = Sal^j_{i,t} - Pur^j_{i,t} \tag{3}$$

and the other changes $\triangle OC^j_{i,t}$ are reserved for reporting errors that may occur. The position in bonds includes accrued interest.

4.2 Risk, Return, and Benchmark Comparison

As a first step in our analysis, we calculate the returns and risks for the different asset classes and compare those with benchmarks. We are restricted to determining the nominal price return, as we do not have data on cash dividend receipts for equities. Cash dividends received by pension funds are either used to pay pensions or are used to invest in assets. We calculate the money-weighted return on each asset class using the Modified Dietz Method (Dietz 1966), which is given by

$$R^j_{i,t+1} = \frac{PB^j_{i,t+1} - PB^j_{i,t} - \triangle OC^j_{i,t} - Tr^j_{i,t}}{PB^j_{i,t} + w * Tr^j_{i,t}}, \tag{4}$$

whereby we set $w = 0.5$. This means that we assume that transactions are on average executed halfway during the month. Then, we calculate the average weighted return $\bar{R}$ across all pension funds as follows:

$$\bar{R}_t^j = \sum_{i=1}^{I} R_{i,t}^j q_{i,t}^j, \tag{5}$$

which takes the sum of pension funds $i = \{1, 2, \ldots, I\}$ with weights $q_{i,t}^j = \frac{PB_{i,t}^j}{\sum_{i=1}^{I} PB_{i,t}^j}$ based on the investments of pension fund i in asset class $j = \{1, 2, 3\}$ at time t. The average standard deviation of returns is derived similarly to the weighted average across pension funds.

We compare the equity portfolio return with the return on the MSCI World Price Index and the MSCI All Country World Price Index, both in euros. The bond portfolio returns are compared with the JPMorgan EMU Government Bond Index and the JPMorgan Global Bond Index. The statistics of these time series are presented in table 1.

The average monthly equity return is 0.86 percent, which corresponds to an annual price return of 10.82 percent. This shows that the period that we analyze was relatively good in terms of stock market performance. The monthly standard deviation of equity returns is 3.21 percent or about 11 percent annually.[3] The mean of the monthly returns on bonds is 0.24 percent or 2.9 percent annually. The standard deviation of the monthly bond returns is 1.81 percent or 6.27 percent on an annual basis. We find that the mean return and standard deviation of the investment trusts' returns are larger than for bonds and lower than for equity, since investment trusts have both equity and bond holdings.

The time series and corresponding correlations are shown in figure 1. The average weighted return on equity $\bar{R}^{equity}$ is about 85 percent correlated with the MSCI indexes, and the average weighted return on bonds $\bar{R}^{bonds}$ is more than 70 percent correlated with the JPMorgan indexes.

We expect the return per asset class to be closely linked to benchmark returns, as pension funds typically have broad, diversified

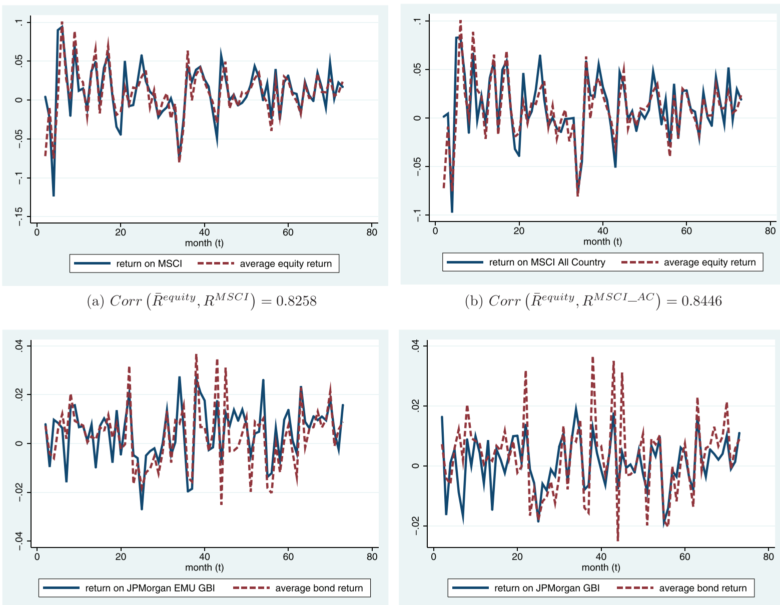
[3] We argue that the relatively low standard deviation of equity returns is a coincidence due to the short period that we analyze.

Table 1. Statistics of the MSCI, JPM EMU Government Bond Index, JPM Global Bond Index, Equity Returns, Bond Returns, and Returns Obtained from Investment Trusts

Variable	Obs.	Mean	Std. Dev.	90%–CI	Min.	Max.
R^{equity}	2,299	.0086	.0321	(−.0400, .0607)	−.1934	.1528
R^{trusts}	2,299	.0058	.0257	(−.0304, .0438)	−.1819	.2043
R^{bonds}	2,299	.0024	.0181	(−.0246, .0319	−.2205	.1518
$\bar{R}^{equity}$	72	.0092	.0317	(−.0513, .0672)	−.0808	.1014
$\bar{R}^{trusts}$	72	.0074	.0255	(−.0199, .0380)	−.1146	.0725
$\bar{R}^{bonds}$	72	.0028	.0126	(−.0188, .0314)	−.0251	.0369
R^{MSCI}	72	.0110	.0345	(−.0460, .0624)	−.1240	.0942
R^{MSCI_AC}	72	.0121	.0334	(−.0471, .0654)	−.0974	.0833
R^{JPM_EMU}	72	.0046	.0113	(−.0168, .0244)	−.0269	.0275
R^{JPM_GBI}	72	.0011	.0086	(−.0165, .0165)	−.0187	.0188

Note: MSCI denotes the MSCI World Price Index, MSCI_AC the MSCI All Country World Price Index, JPM_EMU the JPMorgan EMU Government Bond Index, and JPM_GBI the JPMorgan Global Bond Index.

Figure 1. Time Series and Correlations of the MSCI, JPM Bond Index, Equity Price Returns, and Bond Returns



(a) $Corr\left(\bar{R}^{equity}, R^{MSCI}\right) = 0.8258$

(b) $Corr\left(\bar{R}^{equity}, R^{MSCI\_AC}\right) = 0.8446$

(c) $Corr\left(\bar{R}^{bonds}, \bar{R}^{JPM\_EMU}\right) = 0.7526$

(d) $Corr\left(\bar{R}^{bonds}, \bar{R}^{JPM\_GBI}\right) = 0.7093$

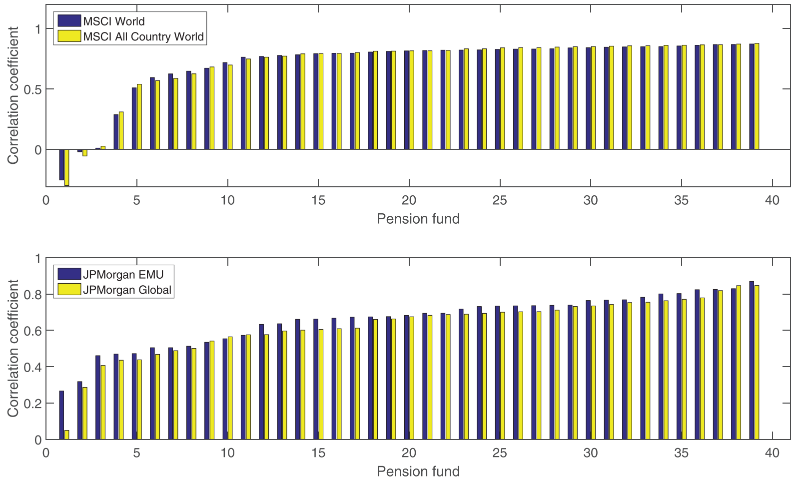
Note: MSCI denotes the MSCI World Price Index, MSCI_AC the MSCI All Country World Price Index, JPM_EMU the JPMorgan EMU Government Bond Index, and JPM_GBI the JPMorgan Global Bond Index.

portfolios and assess their performance relative to a benchmark. The correlations between individual pension fund returns and benchmark returns are shown in figure 2. For most pension funds the correlation coefficient between the price return on the equity portfolio and the MSCI World Price Index returns and the correlation coefficient between the returns on the bond portfolio and the returns on the JPMorgan Index are indeed higher than 50 percent.

4.3 Dependent and Explanatory Variables

The equity and bond allocations are the key dependent variables of interest in our analysis. Table 2 shows the summary statistics of the asset allocations of the pension funds. The mean allocation w^j is

Figure 2. Correlations of the MSCI World Indexes, JPMorgan Bond Indexes, Equity Price Returns, and Bond Returns per Pension Fund



calculated as the equally weighted average direct equity allocation across all pension funds and across time,

$$w^j = \frac{1}{N} \sum_{i=1}^{I} \sum_{t=1}^{T} w_{i,t}^j, \tag{6}$$

for asset class $j = \{1, 2, 3\}$. The mean direct equity allocation is 27.04 percent. This is a proxy for the true equity allocation for two reasons. First, our ISIN data do not include information on pension funds' investments in other asset classes, which are mainly alternative asset classes, such as private equity, direct real estate, hedge funds, and commodities. Second, pension funds can also have indirect equity exposure through investment trusts. The true asset allocation will therefore deviate from the proxy asset allocation presented in table 2. The mean direct allocation to bonds is 46.43 percent. Also this will deviate from the true bond allocation because of the two reasons mentioned before. By construction the three weights add up to one.

If we turn to the explanatory variables, we observe the following. The variable log $(Assets)$ denotes the natural logarithm of the

Table 2. Summary Statistics

Variable	**Obs.**	**Mean**	**Std. Dev.**	**90%–CI**	**Min.**	**Max.**
w^{equity}	2,299	.2704	.1441	(.0127, .5150)	0	.8476
w^{trusts}	2,299	.2653	.2011	(.0280, .9263	0	.9560
w^{bonds}	2,299	.4643	.1481	(.0570, .6647)	0	.8128
Log $(Assets)$	2,299	15.6764	1.1638	(14.1344, 18.3561)	13.2588	19.7443
$\frac{Actives}{AllParticipants}$	2,299	.3210	.1383	(.0991, .5377)	0	.6528
FR	2,299	1.0910	.1179	(.919, 1.310)	.8	1.57

total assets. This number is below the true log of assets, as again not all asset classes are included in our sample. The ratio of active participants over all participants is an indicator of the maturity of a pension fund. The active participants are the participants that pay contributions to the pension fund. The inactive participants are the retirees plus the so-called dormant members.[4] A dormant or former member is entitled to future pension benefits but is no longer in the service of the employer and therefore does not contribute to the pension fund. The funding ratio FR is the ratio of a pension fund's assets to its liabilities. The latter is the total marked-to-market value of accrued benefit obligations. The minimum required funding ratio by Dutch legislation is roughly 105 percent. However, 37.76 percent of the observations do not satisfy this requirement, due to the weak financial positions of pension funds during the Great Financial Crisis.

5. The Model

In this section we describe the benchmark model of our analysis. The rebalancing model for the asset allocation is introduced in section 5.1. Section 5.2 discusses the changes in the strategic asset allocation. In section 5.3 we extend the benchmark model by a variable which measures the strategic deviations in the asset allocation with respect to other pension funds, depending on their interconnectivity, i.e., we add a spatial estimation approach to our benchmark model.

5.1 Rebalancing Regression Model

Over time, a pension fund's asset allocation will fluctuate around its strategic level. We perform an analysis based on the method applied by Calvet, Campbell, and Sodini (2009). They show that the allocation of a specific asset class can be decomposed into a passive and

[4]We have the data on the number of participants on a yearly basis only. However, the ratio of active participants over all participants is rather stable over time for each pension fund. Therefore, we interpolate the data to approximate this variable on a monthly basis. Furthermore, the data set contains one so-called closed pension fund, which means that no new participants enter the pension fund. The $\min\left(\frac{Actives}{AllParticipants}\right) = 0$ obtained from our data set concerns this closed pension fund, with non-active participants only.

an active share. The current month's passive share in asset class j is the hypothetical share that would have been obtained if the pension fund had not traded during the last month,

$$w_{i,t}^{j,p} = \frac{w_{i,t-1}^{j}\left(1 + R_{i,t}^{j}\right)}{\sum_{k=1}^{3} w_{i,t-1}^{k}\left(1 + R_{i,t}^{k}\right)}. \tag{7}$$

Then, we derive the passive change as the difference between the current passive share and the last month's actual share,

$$P_{i,t}^{j} = w_{i,t}^{j,p} - w_{i,t-1}^{j}. \tag{8}$$

The active change is given by the actual change minus the passive change,

$$A_{i,t}^{j} = w_{i,t}^{j} - w_{i,t-1}^{j} - P_{i,t}^{j}. \tag{9}$$

Then, we explore to what extent the passive changes explain the active changes, as an estimation for pension funds' rebalancing within a month. However, the returns of the different asset classes determine the asset allocation, not only in the corresponding month but also thereafter. We capture this effect by including the lagged asset allocation $w_{i,t-1}^{j}$ in the model. Hence, we apply the following benchmark equation for pension fund $i \in \{1, 2, \ldots, I\}$, for month $t \in \{1, 2, \ldots, T\}$ and asset class $j \in \{1, 2, 3\}$:

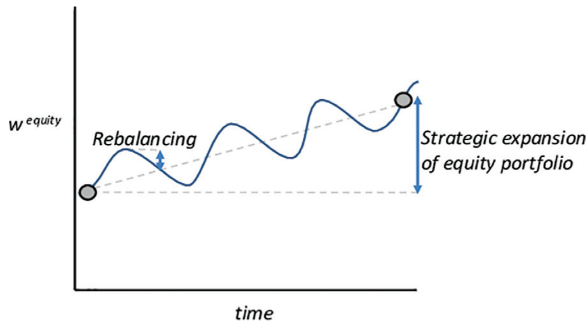
$$A_{i,t}^{j} = \beta_1 P_{i,t}^{j} + \beta_2 w_{i,t-1}^{j} + \beta_3 d\left(Act_{i,t}\right) + \beta_4 d\left(FR_t\right) + \alpha_i + \theta_t + \varepsilon_{i,t}. \tag{10}$$

In this model $d\left(Act\right)$ is the change in the pension fund's share of active participants,[5] $d\left(FR\right)$ is the change in the pension fund's funding ratio,[6] α_i is the pension fund fixed effect, θ_t is the time fixed effect, and $\varepsilon_{i,t}$ is a random error term.

[5]The share of active participants is defined as the number of active members divided by the total number of participants, being active members, dormant members, and pensioners.

[6]There is one missing observation for the funding ratio, which we replace with an approximated value using interpolation.

Figure 3. Graphical Illustration of Rebalancing and Strategic Deviations from an Equity Allocation over Time



5.2 Rebalancing and Changes in the Strategic Asset Allocation

The asset allocations fluctuate over time because of two reasons: (i) pension funds (partially) rebalance in response to the returns of the different asset classes, and (ii) the pension fund's strategic asset allocation changes over time. Figure 3 provides a graphical illustration of the rebalancing effects and the strategic deviations. When the returns on equity are, e.g., relatively high compared with the return on other asset classes, the pension fund can sell equities to buy other asset classes. This process is referred to as rebalancing. If pension funds continuously rebalance their portfolio, the effect under (i) will be completely offset. Continuously rebalancing, however, is costly, and it is not always possible and necessary to immediately respond to fluctuations in the asset returns. Therefore, most pension funds allow the asset allocation to drift between certain limits. For example, a pension fund might allow the equity allocation to fluctuate between 40 and 50 percent. In practice, rebalancing will therefore only be partial. According to Bikker, Broeders, and de Dreu (2010), rebalancing accounts for 39 percent of the portfolio changes. All pension funds are expected to have a rebalancing strategy; otherwise, the actual asset allocation will drift away from the strategic asset allocation. When rebalancing, pension funds make active investment decisions based on similar market information. Rebalancing can therefore be interpreted as a form of weak herding.

It is hard to disentangle the strategic deviations from the rebalancing effects, which are the two effects that cause the changes in the equity allocation. Over the long run, however, deviations in the equity allocation can be considered as a strategic decision of the pension fund's management—see figure 3. Hence, we disentangle changes in the strategic asset allocation from the rebalancing effects by tracking the changes over a long time period. Our measure for changes in the strategic asset allocation is denoted by $Z^j_{i,t} = \frac{w^j_{i,t} - w^j_{i,t-\tau}}{\tau}$. For a large enough time span τ, the fluctuations due to volatile asset returns are smoothed out, such that we mainly measure the changes in the strategic equity allocation. Typically, pension funds review and adjust their strategic asset allocation every three years, with a midpoint of 18 months. We therefore look at τ ranging from 12 to 24 months. If we extend τ further, we would lose too many observations.

5.3 Interconnectivity

The final step in our model is to apply spatial econometric analysis to determine the interconnectivity between pension funds to test for strong herding behavior. For that we use a weighting matrix W of size $[IT \times IT]$ that denotes the spatial distance between pension funds. We define different matrix specifications in order to test herding between pension funds with specific characteristics. For example, we assign weights equal to one in case pension funds are of similar type, have similar share of active participants, or are of similar size. Alternatively we can test whether, for example, the three largest pension funds are market leaders, which holds when they are followed by all others. Hence, for measuring the connectivity of pension funds to their competitors' deviations in the equity and bond allocation, we extend our benchmark model with a spatial relation toward Z, as follows:

$$A^j_{i,t} = \beta_1 P^j_{i,t} + \beta_2 w^j_{i,t-1} + \beta_3 d\left(Act_{i,t}\right) + \beta_4 d\left(FR_t\right) + \beta_5 W_i Z^j_{t-1} + \alpha_i + \theta_t + \varepsilon_{i,t}, \tag{11}$$

whereby W_i denotes the (spatial) weighting matrix, which relates to the changes in strategic asset allocation of the different pension funds.[7] We argue that it is plausible that pension funds observe each other's asset weights, e.g., by quarterly and annual reports.

6. Results

This section discusses the main results from our empirical analysis. First, section 6.1 discusses the results with respect to weak herding. Second, section 6.2 provides a discussion about the findings for semi-strong herding. Finally, we investigate the results for strong herding in section 6.3.

6.1 Weak Herding (Information Motive)

In this section we discuss the results of weak herding. This is based on similar rebalancing strategies across pension funds. The motive for weak herding is based on the fact that pension funds have the same market information and will react similar to this information, as they want stay close to their strategic asset allocation over time. Table 3 presents the results for two specifications of our benchmark model, for both equities and bonds. The first and third column exclude the control variables for the change in active participants and the change in the funding ratio from equation (10). Both models have been specified using a within regression with clustered (by pension fund) standard errors. A Hausman test indicates that a model using unit random effects does not satisfy the corresponding assumptions.

The key observation from table 3 is that the coefficient estimates in the first two rows support rebalancing strategies of pension funds. First, approximately 20 percent of the passive changes in the equity allocation is offset by active changes, while for the bond allocation the active changes offset almost 25 percent of the passive changes. Hence, this implies that pension funds rebalance 20–25 percent of the passive changes during the month by active buying and selling in the asset classes. Second, the coefficient estimates for the asset

[7]We row standardize W, such that the weights per pension fund i at time t add up to one. This means that when pension funds consider the competitors' deviations, they have to divide their attention among the number of competitors. Hence, the assigned weight attributed to each competitor reduces as a pension fund is connected to more competitors.

Table 3. Coefficient Estimates Based on Regression Equation (10)

Dependent Variable $A^j_{i,t}$	***j*: Equity**		***j*: Bonds**	
$P^j_{i,t}$	$-.2053^{***}$	$-.2029^{***}$	$-.2455^{***}$	$-.2454^{***}$
	(.0538)	(.0539)	(.0543)	(.0544)
$w^j_{i,t-1}$	$-.0171^{***}$	$-.0170^{***}$	$-.0211^{***}$	$-.0211^{***}$
	(.0032)	(.0032)	(.0040)	(.0040)
$\mathrm{d}(Act_{i,t})$	—	.0347	—	−.0627
		(.0722)		(.0870)
$\mathrm{d}(FR_{i,t})$	—	−.0110	—	.0054
		(.0109)		(.0131)
Number of Observations	2,149	2,149	2,149	2,149
R^2 – Within	.0737	.0743	.0827	.0831
R^2 – Between	.0097	.0097	.0053	.0034
R^2 – Overall	.0355	.0362	.0381	.0388
Wald Test: Prob. $> \chi^2$	0.0000	0.0000	0.0000	0.0000

Note: Robust standard errors are in parentheses; $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

allocation in the previous period $w^i_{i,t-1}$ is around −2 percent and statistically negative at the 1 percent significance level. Since a high asset allocation in the previous month implies a decline in the corresponding asset allocation in the current month, this finding also shows the tendency of pension funds to rebalance their asset allocation. Both results suggest that pension funds on average rebalance their asset allocation towards a strategic level.

This rebalancing strategy of pension funds contributes to financial market stability, as this implies a buy-low-and-sell-high strategy. If the return on equities is relatively low compared with bonds (and other asset classes), pension funds will buy additional equities. And reversely, if equities performed relatively well, they will sell equities.

Moving on to the two additional explanatory variables in the second and fourth column, we observe that neither the change in the share of active participants nor the change in the funding ratio of pension funds significantly affects equity allocation changes. Since these variables are slowly moving and are likely to exert an effect on the dependent variable over the long term, the monthly deviations are not significantly affected by these effects.

6.2 Semi-strong Herding (Regulation Motive)

Next we turn to the results for semi-strong herding. Changes in regulation can affect the asset allocation of pension funds. This type of herding takes place when investors' preferences (risk appetite) toward asset classes with specific characteristics change following new regulation. We test the prevalence of semi-strong herding among Dutch pension funds by investigating monthly dummy variables. Table 4 shows the dummy variables for which the specified model produces statistically significant coefficients. The cases listed are significant changes in equity or bond allocation simultaneous to or directly following a regulatory change. According to our knowledge, it is in many instances not a priori clear whether it would be optimal to expand or contract the equity or bond allocation as a result of the corresponding event. Also we cannot be sure that the significant time effect comes from the economic and regulatory event around that date. However, on average pension funds appear to react in similar ways, as is demonstrated by the significant time effects around the date of the economic and regulatory event, for which we find multiple examples. Hence, we consider these findings as semi-strong herding, which we discuss below. Notice that the sign of the coefficient, even if significant, does not necessarily indicate whether the corresponding asset allocation on average expands or contracts. It is the average net active change in the asset allocation after correcting for the other variables presented in equation (10).

The main results concern changes in Dutch pension regulation and developments in the Dutch pension system. The first significant time dummy is obtained for May 2009. On May 25, 2009, the Ministry for Social Affairs and Employment (MSAE; this is the ministry responsible for pension fund legislation) announced broad measures in order to tackle the many financial challenges that Dutch pension funds were facing following the financial crisis. It also announced an independent enquiry into pension fund's risk-taking in asset management. When the crisis hit, many pension funds had to incur losses on their investment portfolios, forcing some of them to temporarily cut (previously defined) retirement benefits. It is not unlikely that pension funds viewed the May 2009 announcement as a starting point for regulations that favored de-risking, which would reduce potential losses but also decrease the likelihood that retirees be

Table 4. Coefficient Estimates for Monthly Period Dummy Variables with January 2015 as Reference Date Based on Regression Equation (10)

Year	Month	Equity Allocation	Bond Allocation	Relevant Economic and Regulatory Event(s)
2009	May	.0016	−.0047*	Ministry of Social Affairs and Employment announces broad measures to tackle financial challenges of the Dutch pension system
		(.0022)	(.0026)	
	Jul	.0042**	−.0055**	
		(.0021)	(.0025)	
	Aug	.0064***	−.0083***	
		(.0023)	(.0027)	
2010	Feb	−.0008	−.0064**	Publication of the report of Commission Goudswaard and Commission Frijns
		(.0020)	(.0025)	
	Sep	.0018	−.0042*	European Parliament approved legislation allowing establishment of European Supervisory Authorities
		(.0021)	(.0025)	
	Oct	−.0012	−.0047*	
		(.0021)	(.0025)	
2011	Jan	−.0011	−.0045*	
		(.0020)	(.0025)	
	Feb	.0037*	−.0027	EIOPA established and EIOPA regulation enters into force
		(.0020)	(.0024)	
	Mar	−.0043**	.0039	
		(.0020)	(.0024)	
2013	Mar	.0030	−.0061**	Benefit reductions for insolvent pension funds
		(.0021)	(.0025)	
2014	Mar	−.0045**	.0031	
		(.0022)	(.0025)	Commission Parameters publishes second report
	Apr	.0025	−.0042*	
		(.0021)	(.0025)	
	Dec	.0004	−.0044*	Dutch legislation on adjustment of the financial assessment framework for pension funds adopted
		(.0021)	(.0025)	

Note: Robust standard errors are in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

compensated for inflation. In this regard, the equity allocation hike in July and August might be in anticipation of stricter regulation of risky investments.

In February 2010, the report of Commission Goudswaard on the long-run financial sustainability of Dutch occupational funded pensions and the report of Commission Frijns on pension funds' investment were published. Also, the so-called Commission Parameters (an independent advisory committee established by the MSAE) published its second report in March 2014. One of the changes in this second report was a reduction in the expected return on equities. These parameters are used by pension funds in making long-term stochastic projections of their funding ratios. They are also used in setting the contribution policy. In April 2013, many pension funds were forced to reduce the pension rights of their participants to fulfill the recovery requirements, which is followed by a significant change in the bond allocations in March 2013. Finally, in December 2014, some adjustments in the financial assessment framework for pension funds were adopted. EIOPA is the supervisory authority for Institutions for Occupational Retirement Provision (IORP). We observe significant changes in the asset allocations during January 2011 to March 2011, which is immediately after the establishment of EIOPA and after its regulation entered into force. Also in September and October 2010, we obtain a significant change in the bond allocations, around September 22, 2010, when the European Parliament approved the legislation allowing the establishment of the European Supervisory Authorities.

Finally, there are some periods in which relevant changes in regulation did not lead to significant time effects in the aggregate asset allocation of Dutch pension funds. For example, the ultimate forward rate (UFR) for pension funds, affecting the discount rates for long-term liabilities, was introduced in October 2012. Nonetheless, no significant changes in equity or bond allocations are found around that introduction.

6.3 Strong Herding (Reputation Motive)

A final motive driving institutional herding behavior is reputation. Following the argumentation of peer-group pressure, we would expect the risk-taking behavior of a pension fund to be partly dependent on the risk-taking behavior of other pension funds. In other words, pension funds follow the asset allocation of one another. We call this strong herding, as this motive suggests a direct link between the behavior of different actors, rather than an indirect one through common exposure to information or regulation.

We test the hypothesis of the reputation motive by identifying the existence of spatial correlation between changes in pension funds' strategic allocations in asset class j, which is measured by $Z^j_{i,t} = \frac{w^j_{i,t} - w^j_{i,t-\tau}}{\tau}$ for a sufficiently large time span τ. Hence, we take $Z^j_{i,t-1}$ as our measure for strategic changes in the equity or bond allocation of pension funds, which may potentially be followed by other pension funds. Choosing an appropriate time frame to test the spatial effect of the asset allocation is key. Typically pension funds review and adjust their strategic asset allocation every three years, with a midpoint of 18 months. We therefore capture the strategic deviations in the equity and bond portfolio of a pension fund by tracking the changes over 12, 15, 18, 21, and 24 months. In addition, we specify four different connectivity matrices, which allows us to test alternative channels (based on different ways to measure similarity between funds) of herding between pension funds in our data set.

A complicating factor in establishing a relationship between active changes in the asset allocation (our dependent variable) and the change in strategic asset allocation of other pension funds is the fact that pension funds tend to rebalance their asset portfolios over time. A change in the composition of asset portfolios may therefore be the result of the fact that a pension fund is merely rebalancing its portfolio to align it with a strategically chosen asset mix. We have no strong prior as to the length of the time horizon across which rebalancing is the strongest. However, we consider it unlikely that this time horizon exceeds 12 months given the regulatory cycle to which Dutch pension funds are exposed. Still, even when some funds rebalance over a longer period of time, this effect should diminish the spatial effect (which is positive according to our hypothesis), not strengthen it.

Table 5 contains the results of this analysis, which are based on the model as described in equation (11). Hence, we use the

Table 5. Coefficient Estimates for the Spatial Lags $Z_{i,t-1}^{j} = \frac{w_{i,t-1}^{j} - w_{i,t-1-\tau}^{j}}{\tau}$ Based on Regression Equation (11)

	Connected with Three Largest Funds	Connected for Similar Fund Size	Connected for Similar Fund Type	Connected for Similar Fund Age
		j = *Equity*		
$\tau = 12$	1.8909**	.0593	−.1052	.1461
	(.8608)	(.1173)	(.1199)	(.0931)
$\tau = 15$	−.2084	.3499**	−.1116	.0956
	(.9013)	(.1479)	(.1380)	(.1118)
$\tau = 18$	1.9129**	.4667**	−.0775	.0953
	(.8512)	(.1937)	(.1510)	(.1240)
$\tau = 21$	.9180	.3556	−.0988	.0659
	(1.2872)	(.2535)	(.1831)	(.1601)
$\tau = 24$	1.3997	−.0220	.0623	−.0467
	(1.2098)	(.2734)	(.2164)	(.1766)
		j = *Bonds*		
$\tau = 12$	.2903	.2788**	−.0659	.0107
	(.3536)	(.1123)	(.1391)	(.1460)
$\tau = 15$	−.0366	.1088	−.0314	−.2267
	(.3884)	(.1312)	(.1694)	(.1768)
$\tau = 18$	.1122	.0273	.1536	−.2266
	(.5047)	(.1590)	(.1942)	(.2291)
$\tau = 21$	1.3159**	−.1168	−.0710	−.3791
	(.6345)	(.1878)	(.2208)	(.2405)
$\tau = 24$	.9294*	−.2910	.3958	−.4286
	(.5636)	(.2165)	(.2506)	(.2827)

Note: Robust standard errors are in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

same estimator (fixed-effects regression with cluster-robust standard errors) and include all explanatory variables included in that model. Yet, for the sake of parsimony, only the spatial lag coefficients are displayed in the table. The columns feature four spatial lags based on the following connectivity matrices. In the first column, all pension funds are connected to the three largest pension funds in terms of assets under management. In the second column, pension funds are only connected to other pension funds when they are of similar size (also measured by assets under management). We distinguish between small, medium-sized, and large pension funds, where the thresholds between these categories are at 3 billion and 9 billion euros, respectively. This way, each of the three categories represent roughly a third of the data set. In the third column, pension funds are connected only to the same "type" of pension funds. We distinguish between three types of pension funds: industry-wide, professional group, and corporate pension funds. The fourth column connects pension funds only to other funds when they have a similar share of active (still working) participants as opposed to retired participants. We distinguish three categories with thresholds at 25 percent and 40 percent active participants. Again, this results in roughly equally sized categories. Finally, none of the connectivity matrices allow for pension funds to be connected to themselves, which is indicated by setting the corresponding weights in W equal to zero. To the extent that pension funds "follow themselves" (i.e., demonstrate path dependence in their asset allocation), this effect is captured by the lagged asset allocation and pension fund fixed effect which are included in all models as is done in the benchmark model.

Moving to the results, we observe that two of the four columns generate some significant coefficients. Column 2, which contains a spatial lag that is based on fund size similarity, suggests that there is a positive effect over a time horizon of 15 and 18 months for which we find the most robust evidence of strong herding behavior. If pension funds increase their equity allocation over the last 15–18 months with 1 percentage point on average, then pension funds with a similar size typically increase their equity allocation by 0.35 to 0.47 percentage point as well. Both in terms of significance and size, the effect diminishes when the time horizon moves away from these 15–18 months. As discussed above, this could be partly due to rebalancing, but we find it equally likely that pension funds do not change their strategic asset allocation over a shorter period of time.

There is also some (although less robust) evidence that pension funds follow the equity allocation of the three largest pension funds. Given the spatial effects of similarly sized pension funds discussed above, this result is perhaps not surprising. In terms of time horizon, the evidence is found at 12 months, but also at 18 months, as shown in column 2. The significant coefficient estimates are almost equal to 2, meaning that when the three largest pension funds increase their strategic equity allocation by 1 percentage point, the other pension funds overreact with an increase of their equity allocation by almost 2 percentage points.

We found less statistical evidence concerning bond allocations. However, the two cases for which we found strong herding at 5 percent significance level are similar to the cases for the equity allocation.

These results need to be interpreted with care. As already mentioned, it is not possible to perfectly disentangle changes in the strategic asset allocation from the rebalancing effect. Furthermore, it strongly depends on the specification of the connectivity whether strong herding can be identified. This appears not to be the case for the connectivity among pension funds with similar type or similar share of active participants.

7. Robustness Checks

As a robustness check, we perform an alternative analysis in this section. First, we explain the alternative model in section 7.1. Second, we discuss the results with respect to weak herding, semi-strong herding, and strong herding in section 7.2, section 7.3, and section 7.4, respectively.

7.1 Error Correction Model for Changes in the Asset Allocation

We perform an alternative analysis using a slight adoption of the Error Correction Model (Engle and Granger 1987). The asset allocations are again the key interest in our analysis. We cannot reject that the asset allocation is a stationary variable. The test results for unit root of equity and bond allocations are shown in appendix B. This could lead to biased results when left unattended in the analysis. To tackle this issue, we take the changes in the asset allocation $d\left(w_{i,t}\right) \equiv w_{i,t} - w_{i,t-1}$ as the dependent variable, which does satisfy stationarity. The returns of the different asset classes determine the asset allocation, not only in the corresponding month but also thereafter. For the changes in the asset allocation in the corresponding month, we include the returns of the three asset classes, while for the changes thereafter we again include the lagged asset allocation $w_{i,t-1}^{j}$ in the model. Hence, we specify the following model that has similarities with the Error Correction Model:

$$d\left(w_{i,t}^{j}\right) = \sum_{j=1}^{3} \beta_j R_{i,t}^{j} + \beta_4 w_{i,t-1}^{j} + \beta_5 d\left(Act_{i,t}\right) + \beta_6 d\left(FR_{i,t}\right) + \alpha_i + \theta_t + \varepsilon_{i,t}. \tag{12}$$

To replicate the analysis of section 5.3, we also test for strong herding, by extending the regression with a spatial relation to

$$d\left(w_{i,t}^{j}\right) = \sum_{j=1}^{3} \beta_j R_{i,t}^{j} + \beta_4 w_{i,t-1}^{j} + \beta_5 d\left(Act_{i,t}\right) + \beta_6 d\left(FR_{i,t}\right) + \beta_7 W_i Z_{t-1}^{j} + \alpha_i + \theta_t + \varepsilon_{i,t}. \tag{13}$$

7.2 Weak Herding

Table 6 presents the results of our alternative regression model equation (12). Reading the table from top to bottom, the change in equity allocation is obviously positively related to equity returns. This result simply points toward the fact that the equity allocation increases by construction if equity returns are positive. Conversely, and following the same line of reasoning, equity allocation reacts negatively to positive bond and trust returns.

The key insight from table 6 is that the coefficient estimates in the first four rows support rebalancing strategies of pension funds. First, the coefficients of the returns from the three asset classes are

Table 6. Coefficient Estimates of the Benchmark Model Based on Regression Equation (12)

Dependent Variable $d(w^j_{i,t})$:	***j*: Equity**		***j*: Bonds**	
$R^{equity}_{i,t}$	.1063***	.1063***	−.0410***	−.0414***
	(.0116)	(.0116)	(.0140)	(.0140)
$R^{trusts}_{i,t}$	−.0309***	−.0307***	−.0700***	−.0704***
	(.0085)	(.0085)	(.0102)	(.0102)
$R^{bonds}_{i,t}$	−.0508***	−.0506***	.1163***	.1154***
	(.0153)	(.0154)	(.0185)	(.0185)
$w^j_{i,t-1}$	−.0155***	−.0155***	−.0225***	−.0225***
	(.0032)	(.0032)	(.0041)	(.0041)
d($Act_{i,t}$)	—	−.0055	—	−.0420
		(.0741)		(.0892)
d($FR_{i,t}$)	—	−.0058	—	.0100
		(.0112)		(.0135)
Number of Observations	2,149	2,149	2,149	2,149
R^2 – Within	.3090	.3091	.2601	.2604
R^2 – Between	.0912	.0905	.0010	.0015
R^2 – Overall	.2547	.2549	.2009	.2016
Wald Test: Prob. > χ^2	0.0000	0.0000	0.0000	0.0000

Note: Robust standard errors are in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

lower in absolute terms than what we would expect from a "passive strategy," whereby the pension fund does not rebalance, such that the asset allocations are fully determined by the past returns.[8] Hence, the coefficient estimates of the returns from the three asset classes imply that pension funds rebalance during the month by offsetting part of the returns, as confirmed by our results in section 6.1.

[8]Consider the following numerical example. Suppose the equity allocation equals $w^{equity}_{t-1} = 25\%$ and the monthly returns are $R^{equity}_t = 1\%$, $R^{bonds}_t = 0\%$, and $R^{trusts}_t = 0\%$. Then, ceteris paribus, we would obtain $w^{equity}_t = \frac{101\%*0.25}{101\%*0.25+100\%*0.75} = 25.19\%$. Hence, we might expect a coefficient for R^{equity}_t roughly equal to $\frac{25.19\%-25\%}{1\%} = .19$. However, we find a substantial lower coefficient for R^{equity}_t, namely .1063. This means that we need to take all four coefficients into account when we quantify the average extent of rebalancing, as we have done under our benchmark model in section 6.1. The same holds for the other coefficients and for the bond allocation.

Second, we observe that the larger last month's equity or bond allocation is, the more the current month's share is reduced on average. Both results suggest that pension funds on average rebalance their asset allocation toward a desired level, which is in line with our results on weak herding obtained in section 6.1. For the two additional variables $d\,(Act_{i,t})$ and $d\,(FR_t)$, we again obtain no significant effect on the dependent variable. Hence, the change in the funding ratio and the change in the share of active members do not affect the changes in the monthly asset allocations.

7.3 Semi-strong Herding

Next we turn to the discussion of the results for semi-strong herding, which are presented in table 7. We find more significant month effects under our alternative model than under our benchmark model. Since there is quite some overlap with the results obtained in section 6.1, we mainly discuss the newly obtained significant time effects.

First, we obtain a significant time effect for March 2009, when pension funds with insufficiently high funding ratios received instructions from the regulator for filing recovery plans. Also, the "Commission Parameters" published its first report defining new parameters in September 2009. Their second report, published in March 2014, again significantly affected asset allocations, with lower equity and higher bond allocations.

Furthermore, several developments in the financial assessment framework for Dutch pension funds, the so-called FTK, took place. For example, in April 2010, a report on the evaluation of the FTK was published, while in May 2012 a letter on the revision of the FTK was released. Both events resulted in significant changes in the next month's asset allocations. In September 2011, MSAE published a report which announced a revision of the standard method for the calculation of risk-based buffers for pension funds. Next, in September 2011, a "Pension Deal" was accepted, which includes an agreement among social partners and MSAE concerning the future of the Dutch occupational pension system.

Unlike the benchmark model, we now find several examples in which the European Central Bank's (ECB's) exceptional monetary

Table 7. Coefficient Estimates for Monthly Period Dummy Variables with January 2015 as Reference Date Based on Regression Equation (12)

Year	Month	Equity Allocation	Bond Allocation	Relevant Economic and Regulatory Event(s)
2009	Mar	−.0032	.0049*	Instructions for recovery plans
		(.0024)	(.0029)	
	Jul	.0051**	−.0051**	Ministry of Social Affairs and Employment announces broad measures to tackle financial challenges of the Dutch pension system
		(.0021)	(.0026)	
	Aug	.0050**	−.0066**	
		(.0023)	(.0028)	
	Sep	.0000	−.0050*	Commission Parameters publishes first report and ECB's launch of the Covered Bond Purchase Program (CBPP1)
		(.0024)	(.0028)	
2010	Feb	.0006	−.0081***	Publication of the report of Commission Goudswaard and Commission Frijns
		(.0021)	(.0026)	
	May	.0021	−.0051**	Publication of the report on evaluation of FTK and announcement of SMP by ECB
		(.0021)	(.0025)	
	Aug	−.0030	.0051*	No major regulatory event observed
		(.0022)	(.0026)	
	Oct	−.0040*	−.0020	European Parliament approved legislation allowing establishment of European Supervisory Authorities
		(.0022)	(.0026)	
2011	Jan	−.0010	−.0057**	
		(.0021)	(.0026)	
	Feb	.0049**	−.0044*	EIOPA established and EIOPA regulation enters into force
		(.0021)	(.0025)	
	Mar	−.0042*	.0042	
		(.0021)	(.0026)	
	Oct	−.0057**	.0069**	"Pension Deal" and revision of risk-based capital buffers
		(.0025)	(.0030)	
	Dec	.0055***	−.0056**	Launch of the CBPP2
		(.0021)	(.0025)	

(*continued*)

Table 7. (Continued)

Year	Month	Equity Allocation	Bond Allocation	Relevant Economic and Regulatory Event(s)
2012	Jul	−.0046**	.0065**	Letter on revision of the FTK
		(.0023)	(.0028)	
	Aug	.0032	−.0050*	OMT announced by ECB
		(.0022)	(.0026)	
2013	Mar	.0045**	−.0077***	Benefit reductions for insolvent pension funds
		(.0022)	(.0027)	
2014	Mar	−.0077***	.0065**	
		(.0022)	(.0027)	Commission Parameters publishes second report
	Apr	.0032	−.0049*	
		(.0021)	(.0026)	

Note: Robust standard errors are in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

policy affect pension funds' asset allocations. First, the bond allocation is negatively affected by the ECB's first Covered Bond Purchase Program (CBPP1), which started in September 2009. The second Covered Bond Purchase Program (CBPP2), launched in December 2011, again resulted in significantly lower bond allocations. In May 2010, the ECB's Securities Markets Program (SMP) started with purchasing securities. Finally, the Outright Monetary Transactions (OMT) was announced in August 2012. All these programs resulted in a significant contraction of the pension funds' bond allocation.

In addition, we find one example of the situation in which significant changes in equity or bond allocation did not concur with relevant changes in regulation or exceptional monetary policy operations, which holds for the time effect in August 2010. However, this case is only weakly significant.

7.4 Strong Herding

Table 8 presents the results of the spatial analysis under our alternative regression model equation (13). We use the same estimator (fixed-effects regression with cluster-robust standard errors) and include all explanatory variables included in that model. Again, only the coefficients of the spatial lags are presented.

The first column, which contains spatial lags with the three largest pension funds, and the second column, which contains spatial lags based on fund size similarity, provide the only statistically significant evidence on strong herding. For almost all cases which are significant in table 5, we again obtain significant coefficient estimates for the spatial lag at 5 percent significance level under our alternative regression model. Moreover, the strongest evidence is again obtained for the equity allocation over 15 to 18 months for pension funds with similar size. From this result we can conclude that when pension funds increase their equity allocation over the last 15–18 months with 1 percentage point on average, then pension funds with a similar size typically expand their equity holdings by 0.36 to 0.49 percentage point. We can conclude that our results on strong herding are robust to the type of regression model, as we obtain qualitatively the same results as the ones we have obtained in section 6.3.

Table 8. Coefficient Estimates for the Spatial Lags $Z_t = \frac{w^j_{i,t} - w^j_{i,t-\tau}}{\tau}$ Based on Regression Equation (13)

	Connected with Three Largest Funds	Connected for Similar Fund Size	Connected for Similar Fund Type	Connected for Similar Fund Age
		j = *Equity*		
$\tau = 12$	2.0870**	.0388	−.1272	.1033
	(.8900)	(.1213)	(.1239)	(.0963)
$\tau = 15$	−.5247	.3617**	−.1226	.0232
	(.9248)	(.1519)	(.1418)	(.1147)
$\tau = 18$	1.7152*	.4941**	−.1304	.0225
	(.8790)	(.1995)	(.1555)	(.1277)
$\tau = 21$	.4566	.4287	−.1604	−.0444
	(1.3285)	(.2612)	(.1892)	(.1650)
$\tau = 24$	1.1445	.0390	.0322	−.1171
	(1.2454)	(.2814)	(.2228)	(.1818)
		j = *Bonds*		
$\tau = 12$	.2859	.2727**	−.1059	−.0214
	(.3639)	(.1155)	(.1430)	(.1502)
$\tau = 15$	−.0167	.0837	−.0012	−.1765
	(.3941)	(.1331)	(.1719)	(.1797)
$\tau = 18$	.1316	−.0051	.1893	−.2365
	(.5157)	(.1624)	(.1983)	(.2341)
$\tau = 21$	1.4222**	−.1187	−.0571	−.2976
	(.6467)	(.1918)	(.2252)	(.2455)
$\tau = 24$	.9357	−.3027	.3809	−.3704
	(.5721)	(.2203)	(.2543)	(.2878)

Note: Robust standard errors are in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

8. Conclusion

In this paper we use unique and detailed transaction data to analyze herding behavior among pension funds. We distinguish between weak, semi-strong, and strong herding behavior. Weak herding occurs if pension funds have similar rebalancing strategies. This is unintentional herding based on the fact that pension funds act similar upon market information. Semi-strong herding arises if pension funds react similar to other external shocks, e.g., changes in pension fund regulation. Herding has a regulation motive in this case. Finally, strong herding occurs if pension funds intentionally replicate changes in the strategic asset allocation of other pension funds. In this case herding has a reputation motive. Pension funds may adjust their investment strategy as a result of peer-group pressure without an economic reason.

We find empirical evidence for all three types of herding. In doing so, we use monthly holdings and transaction data of 39 large Dutch pension funds over the period from January 2009 until January 2015. The primary data used are pension funds' detailed investment holdings in bonds, equities, and trusts. These holdings are uniquely identified according to their International Securities Identification Number (ISIN). We aggregate the holdings and transaction data for these three asset classes. We focus the empirical analysis on the equity and bond allocations. We apply a rebalancing regression model to track changes in the equity and bond allocation over time and to measure the spatial distance between pension funds.

Our key findings are the following. Pension funds exhibit weak herding behavior. Pension funds rebalance their asset allocation in the short run and, hence, they react similar to market information. We find robust evidence that more than 20 percent of the passive changes in the equity allocation are offset by active changes during the month. For bonds this rebalancing of the asset allocation accounts for almost 25 percent. Since rebalancing implies a buy-low-and-sell-high strategy, pension funds contribute to financial market stability.

In addition, pension funds demonstrate semi-strong herding behavior. We find multiple examples where pension funds adjust their equity and bond allocations around (the announcements of) changes in pension fund regulation.

Finally, pension funds also display strong herding behavior. The most robust evidence of strong herding is observed for pension funds of similar size over a 15- to 18-month period. If pension funds increase their equity allocation with 1 percentage point on average, then pension funds with a similar size typically increase their equity allocation by 0.35 to 0.47 percentage points with a lag of 15–18 months. The 18-month period is halfway the typical three-year cycle at which the strategic asset allocation is reviewed and adjusted.

As such, our results indicate support for the information, regulation, and reputation motives of herding. We find that our results are robust by replicating the analysis using an alternative regression model. The results from this confirm that pension funds rebalance their asset allocations. Also there is quite some overlap with the results on semi-strong herding. However, we also document evidence of (small) changes in asset allocations in response to exceptional monetary policy operations. Furthermore, we obtain the same qualitative results on strong herding from an expanded model with spatial lags.

Our findings have potential implications for regulators and policymakers who are interested in safeguarding financial stability. Whereas weak herding can contribute to financial stability, strong herding behavior is a risk for financial stability. Regulators need to be aware that semi-strong herding behavior might imply that pension funds react in a similar way to regulatory changes. To prevent a large impact on asset allocations, the regulatory price of risk for different asset classes should be balanced.

Havine said this, there are some points to consider when interpreting the results. First, our holdings and transactions data represent the majority of pension fund investments but exclude alternative asset classes, such as private equity, direct real estate, hedge funds, and commodities. Second, pension funds can also have equity and bond exposures indirectly through the investment trusts. Since we have no detailed information on the holdings and transactions data of the investment trusts, we cannot offer the complete picture on changes in the true asset allocation. In our sample roughly 26.5 percent is allocated to investment trusts. For future research we could extend our analysis by researching herding behavior in specific segments of the equity market, or even in specific stocks and the deployment of derivatives to hedge risks.

Appendix A. Deleted Observations

The raw data contain 2,567 observations. After cleaning the data, the remaining number of observations is 2,299. The following steps show the procedure we followed:

- We drop outliers which do not satisfy the rules from equation (1) and equation (2) with an error over more than 5 percent of the corresponding value (42 observations deleted).
- We drop excessive monthly returns, specifically if they exceed 25 percent (7 observations deleted).
- We drop observations when in a single month the equity or bond allocation sharply increases (> 0.1), while the allocation to investment trusts sharply decreases (< -0.1), and vice versa (22 observations deleted).
- We drop observations when the change in equity and bond allocation ($d\left(w^{equity}\right)$ or $d\left(w^{bond}\right)$) are missing (100 observations deleted).
- We drop outliers for the change in the equity allocation or bond allocation, which holds for $\frac{abs\left\{d\left(w^{j}\right)-mean\left[d\left(w^{j}\right)\right]\right\}}{3*std[d(w^{j})]} > 1$ (92 observations deleted).

Appendix B. Testing for Unit Roots

Since we have a fixed number of pension funds ($I = 39$) and we assume that pension funds have an infinite horizon ($T \rightarrow \infty$), we apply the Fisher-Dickey-Fuller test for a unit root. To control for time effects, we subtract the cross-sectional means. The model we test, the corresponding hypotheses, and the test results are shown in table B.1, for which we specified six lags. The results are robust for the specification of the number of lags. Hence, we have no evidence to reject the null hypothesis, so we conclude that the panels for the equity and bond allocation contain unit roots.

Table B.1. Fisher-Type Unit-Root Test for w^j Based on Augmented Dickey-Fuller Tests

$d(w^j_{i,t}) = \alpha + \beta d(w^j_{i,t-1})$
H_0: All panels contain unit roots ($\alpha, \beta = 0$)
H_a: At least one panel is stationary ($\alpha, \beta \neq 0$)

	p-value	
Test	**j = Equity**	**j = Bond**
Inverse χ^2	0.9139	0.9655
Inverse Normal	0.9413	0.9981
Inverse Logit t	0.9424	0.9987
Modified Inverse χ^2	0.9056	0.9546

No Evidence to Reject H_0

References

Amir, E., Y. Guan, and D. Oswald. 2010. "The Effect of Pension Accounting on Corporate Pension Asset Allocation." *Review of Accounting Studies* 15 (2): 345–66.

Andonov, A., R. M. Bauer, and K. Cremers. 2017. "Pension Fund Asset Allocation and Liability Discount Rates." *Review of Financial Studies* 30 (8): 2555–95.

Avery, C., and P. Zemsky. 1998. "Multidimensional Uncertainty and Herd Behavior in Financial Markets." *American Economic Review* 88 (4): 724–48.

Banerjee, A. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* 107 (3): 797–817.

Bauer, R., M. Bonetti, and D. Broeders. 2020. "Common Advisor Effect in Strategic Asset Allocations." Available on SSRN.

Beck, N., K. S. Gleditsch, and K. Beardsley. 2006. "Space is More than Geography: Using Spatial Econometrics in the Study of Political Economy." *International Studies Quarterly* 50 (1): 27–44.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5): 992–1026.

Bikhchandani, S., and S. Sharma. 2001. "Herd Behavior in Financial Markets." *IMF Staff Papers* 47 (3): 279–310.

Bikker, J. A., D. W. G. A. Broeders, and J. de Dreu. 2010. "Stock Market Performance and Pension Fund Investment Policy: Rebalancing, Free Float, or Market Timing?" *International Journal of Central Banking* 6 (2, June): 53–79.

Blake, D., B. Lehmann, and A. Timmermann. 1999. "Asset Allocation Dynamics and Pension Fund Performance." *Journal of Business* 72 (4): 429–61.

Blake, D., L. Sarno, and G. Zinna. 2017. "The Market for Lemmings: The Herding Behavior of Pension Funds." *Journal of Financial Markets* 36 (November): 17–39.

Bohl, M., J. Brzeszczynski, and B. Wilflin. 2009. "Institutional Investors and Stock Returns Volatility: Empirical Evidence from a Natural Experiment." *Journal of Financial Stability* 5 (2, June): 170–82.

Borio, C., C. Furfine, and P. Lowe. 2001. "Procyclicality of the Financial System and Financial Stability: Issues and Policy Options." *BIS Papers* 1 (March): 1–57.

Broeders, D., P. Hilbers, D. Rijsbergen, and N. Shen. 2014. "What Drives Pension Indexation in Turbulent Times? An Empirical Examination of Dutch Pension Funds." *De Economist* 162 (1): 41–70.

Brunnermeijer, M., and S. Nagel. 2004. "Hedge Funds and the Technology Bubble." *Journal of Finance* 59 (5): 2013–40.

Cai, F., S. Han, and D. Li. 2012. "Institutional Herding in the Corporate Bond Market." International Finance Discussion Paper No. 1071, Board of Governors of the Federal Reserve System.

Calvet, L., J. Campbell, and P. Sodini. 2009. "Fight or Flight? Portfolio Rebalancing by Individual Investors." *Quarterly Journal of Economics* 124 (1): 301–48.

Chari, V., and P. Kehoe. 2002. "On the Robustness of Herds." Working Paper No. 622, Federal Reserve Bank of Minneapolis.

Chari, V., and C. Phelan. 2014. "Too Correlated to Fail." Economic Policy Paper No. 14-3, Federal Reserve Bank of Minneapolis.

Chen, H., J. Ho, C. Lai, and E. Morales-Camargo. 2011. "Institutional Trading Behavior in the ETF Market." Midwest Finance Association 2012 Annual Meetings Paper.

Cipriani, M., and A. Guarino. 2014. "Estimating a Structural Model of Herd Behavior in Financial Markets." *American Economic Review* 104 (1): 224–51.

Dasgupta, A., A. Prat, and M. Verardo. 2011. "The Price Impact of Institutional Herding." *Review of Financial Studies* 24 (3): 892–925.

Dietz, P. 1966. *Pension Funds: Measuring Investment Performance.* New York: The Free Press.

Elliot, D. 2014. "Systemic Risk and the Asset Management Industry." Economic Studies at Brookings.

Engle, R. F., and C. W. J. Granger. 1987. "Co-integration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55 (2): 251–76.

European Insurance and Occupational Pensions Authority (EIOPA). 2016. "IORPs Stress Test Report 2015." http://www.eiopa.europa.eu.

Franzese, R., and J. Hays. 2007. "Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data." *Political Analysis* 15 (2): 140–62.

Froot, K., D. Scharfstein, and J. Stein. 1992. "Herd on the Street: Informational Inefficiencies in a Market with Short-term Speculation." *Journal of Finance* 47 (4): 1461–84.

Gleason, K., I. Mathur, and M. Peterson. 2004. "Analysis of Intraday Herding Behavior among the Sector ETFs." *Journal of Empirical Finance* 11 (5): 681–94.

Gorter, J., and J. Bikker. 2013. "Investment Risk Taking by Institutional Investors." *Applied Economics* 45 (33): 4629–40.

Graham, J. 1999. "Herding among Investment Newsletters: Theory and Evidence." *Journal of Finance* 54 (1): 237–68.

Haldane, A. G. 2014. "The Age of Asset Management?" Speech at the London Business School, London, April 4.

Hirshleifer, D., A. Subrahmanyam, and S. Titam. 1994. "Security Analysis and Trading Patterns when Some Investors Receive Information before Others." *Journal of Finance* 49 (5): 1665–98.

Hott, C. 2009. "Herding Behavior in Asset Markets." *Journal of Financial Stability* 5 (1, January): 35–56.

Jacklin, J., A. Kleidon, and P. Pfleiderer. 1992. "Underestimation of Portfolio Insurance and the Crash of October 1987." *Review of Financial Studies* 5 (1): 35–63.

Jones, B. 2015. "Asset Bubbles: Re-thinking Policy for the Age of Asset Management." Working Paper No. 15/27, International Monetary Fund.

Mohan, N., and T. Zhang. 2014. "An Analysis of Risk-taking Behavior for Public Defined Benefit Pension Plans." *Journal of Banking and Finance* 40 (March): 403–19.

Nirei, M., T. Stamatiou, and V. Sushko. 2012. "Stochastic Herding in Financial Markets: Evidence from Institutional Investor Equity Portfolios." BIS Working Paper No. 371.

Nofsinger, J., and R. Sias. 1999. "Herding and Feedback Trading by Institutional and Individual Investors." *Journal of Finance* 54 (6): 2263–95.

Papaioannou, M., J. Park, J. Pihlman, and H. Van der Hoorn. 2013. "Procyclical Behavior of Institutional Investors during the Recent Financial Crisis: Causes, Impacts, and Challenges." IMF Working Paper No. 13/193.

Pennacchi, G., and M. Rastad. 2011. "Portfolio Allocation for Public Pension Funds." *Journal of Pension Economics and Finance* 10 (2): 221–45.

Prendergast, C., and L. Stole. 1996. "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning." *Journal of Political Economy* 104 (6): 1105–34.

Raddatz, C., and S. Schmukler. 2011. "Deconstructing Herding: Evidence from Pension Fund Investment Behavior." Policy Research Working Paper No. 5700, World Bank.

Rauh, J. 2006. "Investment and Financing Constraints: Evidence from the Funding of Corporate Pension Plans." *Journal of Finance* 61 (1): 33–71.

———. 2009. "Risk Shifting versus Risk Management: Investment Policy in Corporate Pension Plans." *Review of Financial Studies* 22 (7): 2687–2733.

Rubbaniy, G., I. van Lelyveld, and W. Verschoor. 2012. "Herding Behavior and Trading among Dutch Pension Funds." Mimeo, Erasmus School of Economics.

Scharfstein, D., and J. Stein. 1990. "Herd Behavior and Investment." *American Economic Review* 80 (3): 465–79.

Severinson, C., and J. Yermo. 2012. "The Effect of Solvency Regulations and Accounting Standards on Long-Term Investing." OECD Working Paper on Finance, Insurance and Private Pensions No. 30.

Shek, J., I. Shim, and H. S. Shin. 2018. "Investor Redemptions and Fund Manager Sales of Emerging Market Bonds: How Are They Related?" *Review of Finance* 22 (1): 207–41.

Sias, R. 2004. "Institutional Herding." *Review of Financial Studies* 17 (1): 165–206.

Thomas, A., L. Spataro, and N. Mathew. 2014. "Pension Funds and Stock Market Volatility: An Empirical Analysis of OECD Countries." *Journal of Financial Stability* 11 (1, April): 92–103.

Wei, K. D., R. Wermers, and T. Yao. 2015. "Uncommon Value: The Characteristics and Investment Performance of Contrarian Funds." *Management Science* 61 (10): 2394–2414.