Essays on Three Operational and Strategic Problems of Central Banks in a World of Low Interest Rates or with Interbank Market Frictions

by

Thomas Link

DISSERTATION

Submitted to the Faculty of Business Administration and Economics
Heinrich Heine University Düsseldorf
on December 5, 2019,

in Partial Fulfillment of the Requirements for the Degree of

doctor rerum politicarum (Dr. rer. pol.)

Supervisor and Referee: Prof. Dr. Ulrike Neyer

Referee: Prof. Dr. Hans-Theo Normann

To Ela, my great love

Acknowledgments

I am grateful for the privilege that has been granted to me having been allowed to experience the intellectually and personally enriching time I have had as a doctoral student in Düsseldorf.

Ulrike Neyer has substantially shaped this time. She has been an inspiring teacher, a deeply committed and exceptional supervisor, and a tireless discussion partner. I have learned a lot from her in countless hours of intense discussions and from a vast number of valuable remarks she made on my work. I am deeply indebted to Ulrike Neyer for her encouragement, support, guidance as well as for the extraordinary amount of trust and freedom she gave me.

I also would like to thank Hans-Theo Normann for his helpful comments and for his readiness to review this dissertation.

At the faculty, and in particular at the chair of Monetary Economics, I benefited from having colleagues who contributed to a productive and to an uncomplicated environment and I am thankful for their steady cooperation and support.

I gratefully acknowledge financial support from the Heinrich Heine University Düsseldorf and from the state of North Rhine-Westphalia which funded my position as Ulrike Neyer's research and teaching assistant. I also obtained funding from the Konrad-Henkel-Stiftung, the University of Augsburg, and the University of Vienna which is gratefully acknowledged, too.

***

My parents Monika and Dieter have always backed me up with a haven of home and I only recently started to fully realize what a great job they have done from the beginning of my life since I am a father myself. I am forever grateful for their unconditional love, their support, and devotion which they now grant to my own family as well. I am also grateful and I owe a lot to my grandparents Uschi, Erhard, and Hilde who certainly would have been happy if she could have read these lines. In addition, I would like to thank my parents-in-law Gerlinde and Wilfried for their support especially during the past year.

My son Maximilian has given me more moments of happiness since he has entered our lives than I would ever be able to count. His smile has often made me take a more relaxed view of many things.

My wife Michaela has shown me that the light sometimes can start to shine in the blink of an eye. Her love, her companionship, her kindness, her modesty – everything else she is – has kept me on track during the past years. I am grateful to her for loving me, for staying with me, for experiencing the biggest of all adventures together with me, for supporting me, for taking care of Max in a way that is wonderful beyond all comparison, and for having patience until we will have found a place for us to settle down. Merci. I dedicate this dissertation to Michaela.

Thomas Link
Düsseldorf, December 2019

Contents

Preface

This dissertation includes three essays, henceforward referred to as "papers", on operational and strategic problems of central banks in a world of low interest rates or with interbank market frictions. The first paper, co-authored with Ulrike Neyer, deals with the implementation of monetary policy. We consider a central bank that aims at steering the interest rate in the unsecured overnight segment of the interbank money market (the "interbank rate") under an interest corridor regime. Controlling the interbank rate was a standard approach to implement monetary policy until the global financial crisis started.[1] Since then, and in light of the unconventional measures that major central banks have implemented, there is a debate on whether, respectively how, monetary policy implementation frameworks or procedures need to be refined or changed.[2] One of the various problems that central banks have to consider in this respect is the past and potential future emergence of "frictions" in the interbank money market. For example, the interbank rate could become unstable and harder to control in the future due to new banking regulations, as discussed, for instance, by Jackson and Noss (2015); Committee on the Global Financial System and Markets Committee (2015); Bindseil (2016). Frictions in the interbank money market of a different kind materialized in the course of the global financial crisis. For instance, as discussed by Bucher, Hauck, and Neyer (2020, section 1), banks had to make higher efforts to find trading counterparties or to overcome problems of asymmetric information about counterparty credit risks. Taken altogether, this is the context in which our paper is placed. We introduce interbank market frictions in the seminal model of monetary policy implementation presented by Whitesell (2006). Our goal is to explore whether, respectively how, a central bank can control friction-induced volatility of the interest rate in the unsecured overnight interbank market. While a wider interest corridor is typically associated with higher levels of interbank rate volatility, our main finding is that, under certain circumstances, widening an implemented interest corridor can be a measure to reduce the extent of friction-induced interbank rate volatility.

Of course, for central banks, a relatively rigorous consequence of market frictions could be to simply turn away from targeting the interest rate in the unsecured overnight interbank market (as insinuated, for instance, by Bech and Monnet, 2013, pp. 147–148). At

[1] See, for instance, Bech and Monnet (2013, p. 147). For a larger discussion on the concept and the choice of an appropriate "operational target" of monetary policy see, for instance, Bindseil (2004).

[2] See, for instance, Bindseil (2016) and Potter (2016).

the time we wrote our paper, the question of which variable should become the future "key operational target of monetary policy" (for a proposal in this respect see Bindseil, 2016, p. 45, point 4) was still unanswered. By now, at least as far as the European Central Bank (ECB) is concerned, the fact that the reference rate "EONIA" will be (or de facto recently has been) replaced by the "euro short-term rate (€STR)" indicates that the ECB, in fact, might consider to target an interest rate that does not only refer to transactions in the interbank money market.[3] As a consequence, should the ECB try to implement its monetary policy primarily by steering the €STR in the future, it would have to take into account that this reference rate can lie outside the interest corridor formed by its deposit and marginal lending facility.[4] To capture such a world, the theoretical framework we apply in our paper would have to be extended by also considering agents that are counterparties of banks in the unsecured overnight segment of the money market but lack access to the central bank's standing facilities.

The second and third paper in this dissertation have a different subject matter. Both papers are tied to the problem for central banks that, as long as cash exists, it will be hard for them to implement significantly negative policy rates.[5] So, at some point, expansionary monetary policy cannot be implemented by simply lowering policy rates.[6] The notion of an "effective lower bound (ELB)" on monetary policy rates refers to this constraint. With regard to the historically low level of the ECB's key interest rates (at the end of 2019) it is clear that the ELB-constraint is an acute problem which could be intensified should inflation rates in the euro area decline further. As a means to relax the ELB-constraint, Kenneth Rogoff proposes to "phase out" large-denomination banknotes (see, for instance, Rogoff, 2017). Rogoff's reasoning (see Rogoff, 2017, pp. 59–60) is based on the natural assumption that a flight from negative monetary policy rates to cash would be harder if the costs associated with large cash hoardings were higher – and hoarding costs, in turn, for instance including transportation or storage costs, would be be higher in a world where only small-denomination banknotes were available.

[3]According to the European Central Bank (2019), the "euro short-term rate" captures the price that euro-area located banks pay for funds borrowed in the unsecured overnight segment of the wholesale euro money market. Crucially, as pointed out in Deutsche Bundesbank (2019, section 3.1), also transactions between a bank and a non-bank "financial counterparty" in this market segment are considered in the computation of the euro short-term rate. As pointed out in Deutsche Bundesbank (2019, section 3.3), some of these counterparties might lack access to the monetary policy operations of the Eurosystem.

[4]For a short remark on the possibility that the euro short-term rate can lie outside the ECB's interest corridor see, for instance, Deutsche Bundesbank (2019, section 3.3).

[5]See, for instance, Rogoff (2017, p. 47).

[6]See, for instance, Buiter and Rahbari (2015, pp. 3–4) in a paper that addresses a broader audience.

It is an open research question why major central banks like the ECB and the Federal Reserve are still issuing large banknotes and accept to be constrained by a higher effective lower bound – especially with regard to their experiences during the global financial crisis or, as far as the ECB is concerned, with regard to the environment of persistently low interest rates in the euro area. One reason in this respect that is frequently brought forward is the loss of seignorage revenues a central bank had to accept if it took any measures to make large cash holdings unattractive or virtually impossible. Rogoff (2015, pp. 452) argues that losing seignorage revenues, in turn, would weaken a central bank's ability to stay independent from external financing and thus to keep its operational independence.[7] So, if this is actually the case, there is a trade-off for central banks between relaxing the ELB-constraint and shielding their independence by keeping a source of seignorage revenues. In the second and third paper in this dissertation, I take up Rogoff's proposal to lower the ELB and start from this trade-off. The assumption that seignorage matters is a strong one. However, as it will become apparent, the implications this assumption has in the theoretical framework I employ are consistent with major central banks' behaviors observed by now. The models I employ can also explain why it can be rational to keep issuing large banknotes even though the net benefits from removing them immediately would already be greater than zero.

My starting point in the second and third paper is that the aforementioned trade-off is state-dependent. The state-dependency, in turn, creates a need to time the removal of large-denomination banknotes optimally. In the second paper, I consider a central banker in a one-country setting with the option to "call in" large-denomination banknotes, i.e., with the option to stop the issuance and to remove the legal tender status of large notes. I assume that the net benefits from calling-in large notes depend on the natural rate of interest which is uncertain and governed by a stochastic process. I treat the central banker's problem of when to optimally call in large notes as a problem of when to optimally exercise a perpetual American option, respectively, a "real option".[8] The option structure of the central banker's timing problem can explain why a major central bank like the ECB is still issuing banknotes as large as the 100- or 200-euro notes.

[7]See Rogoff (2015, pp. 450–452). See also Rogoff (2016, chapter 6), Buiter (2009, p. 224), Thiele, Niepelt, Krüger, Seitz, Halver, and Michler (2015, p. 10), and Krüger and Seitz (2017, chapter 4.1).

[8]See Dixit and Pindyck (1994, pp. 3–25) for an introduction into the idea of treating irreversible investment decisions under uncertainty as ("real") option exercise problems.

The third paper in this dissertation builds on the second paper and addresses the question, recently raised by Rogoff (2016, chapter 13), of whether there is a need to coordinate the elimination of large banknotes internationally. I consider a strategic setting with two central bankers and assume that calling in large banknotes involves international spillovers on the central bankers' seignorage revenues. I show that strategic considerations determine the central bankers' timing decisions and make them end up in a prisoner's dilemma or face a coordination problem. Which situation will arise in this regard depends on the substitutability of banknotes of different denominations and currencies. Altogether, strategic interactions are a further explanation of why central banks could possibly have the tendency to delay the elimination of large banknotes.

This dissertation is organized as a "collection" of papers. To acknowledge the independence of each paper, I decided to deviate from the "traditional" form of corresponding theses where each essay or paper would be assigned to a consecutively numbered chapter. Instead, each of the three papers in the following is included in its original form with the original section numbering having been maintained, respectively. Each paper includes an abstract, a table of contents, a statement on my own contributions, acknowledgments, and published paper versions, a list of figures, a list of tables (except for the first paper), and a bibliography, respectively. The papers included in this dissertation are, in that order, "Controlling Friction-Induced Interbank Rate Volatility under Symmetric and Asymmetric Interest Corridor Systems", co-authored with Ulrike Neyer, "Optimal Timing of Calling In Large-Denomination Banknotes under Natural Rate Uncertainty", and "International Coordination and Optimal Timing of Calling In Large-Denomination Banknotes in a Two-Player Game".

References

Bech, M. L., and C. Monnet (2013): "The Impact of Unconventional Monetary Policy on the Overnight Interbank Market," in *Liquidity and Funding Markets*, ed. by A. Heath, M. Lilley, and M. Manning, RBA Annual Conference Volume (Discontinued), pp. 147–177. Reserve Bank of Australia.

Bindseil, U. (2004): "The Operational Target of Monetary Policy and the Rise and Fall of Reserve Position Doctrine," Working Paper Series No. 372, European Central Bank.

——— (2016): "Evaluating Monetary Policy Operational Frameworks," Paper for presentation at the Jackson Hole Economic Policy Symposium 2016, August 2016, Federal Reserve Bank of Kansas City, available at `https://www.kansascityfed.org/~/media/files/publicat/sympos/2016/econsymposium-bindseil-paper2.pdf?la=en`.

Bucher, M., A. Hauck, and U. Neyer (2020): "Interbank Market Friction-Induced Holdings of Precautionary Liquidity: Implications for Bank Loan Supply and Monetary Policy Implementation," *Economic Theory*, 70(1), 165–222.

Buiter, W. H. (2009): "Negative Nominal Interest Rates: Three Ways to Overcome the Zero Lower Bound," *The North American Journal of Economics and Finance*, 20(3), 213 – 238.

Buiter, W. H., and E. Rahbari (2015): "High Time to Get Low: Getting Rid of the Lower Bound on Nominal Interest Rates," Global Economics View, Citi Research, Citigroup Global Markets Inc., 9 April 2015, available at https://willembuiter.com/ELB.pdf.

Committee on the Global Financial System and Markets Committee (2015): "Regulatory Change and Monetary Policy," CGFS Papers No. 54, Bank for International Settlements, available at https://www.bis.org/publ/cgfs54.htm.

Deutsche Bundesbank (2019): "Fragen und Antworten zur Euro Short-Term Rate," Deutsche Bundesbank, Zentralbereich Statistik, Abteilung Wertpapier- und Geldmarktstatistiken, Frankfurt am Main, 2 October 2019, available at https://www.bundesbank.de/resource/blob/785114/fae0faabda527de30021a2be0d0bdc67/mL/faq-euro-short-term-rate-data.pdf, accessed on 12 Novemeber 2019.

Dixit, A. K., and R. S. Pindyck (1994): *Investment under Uncertainty*. Princeton University Press, Princeton, New Jersey.

European Central Bank (2019): "Overview of the Euro Short-Term Rate (€STR)," entry on the ECB's website, available online at https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_short-term_rate/html/eurostr_overview.en.html, accessed on 12 Novemeber 2019.

Jackson, C., and J. Noss (2015): "A Heterogeneous Agent Model for Assessing the Effects of Capital Regulation on the Interbank Money Market under a Corridor System," Staff Working Paper No. 548, Bank of England.

Krüger, M., and F. Seitz (2017): *The Benefits of Cash: Costs and Benefits of Cash and Cashless Payment Instruments (Module 2)*. Fritz Knapp Verlag, Frankfurt am Main.

Potter, S. (2016): "Implementing Monetary Policy Post-Crisis: What Have We Learned? What Do We Need to Know?," Remarks at a workshop organized by Columbia University SIPA and the Federal Reserve Bank of New York, New York City, 4 May 2016, available at https://www.newyorkfed.org/newsevents/speeches/2016/pot160504 and https://www.bis.org/review/r160510a.htm.

Rogoff, K. (2015): "Costs and Benefits to Phasing out Paper Currency," *NBER Macroeconomics Annual*, 29, 445–456.

——— (2016): *The Curse of Cash*. Princeton University Press, Princeton and Oxford.

——— (2017): "Dealing with Monetary Paralysis at the Zero Bound," *Journal of Economic Perspectives*, 31(3), 47–66.

Thiele, C.-L., D. Niepelt, M. Krüger, F. Seitz, R. Halver, and A. F. Michler (2015): "Diskussion um das Bargeld: Hätte eine Abschaffung von Banknoten und Münzen wirklich Vorteile?," *ifo Schnelldienst*, 68(13), 3–18.

WHITESELL, W. (2006): "Interest Rate Corridors and Reserves," *Journal of Monetary Economics*, 53(6), 1177–1195.

Controlling Friction-Induced Interbank Rate Volatility under Symmetric and Asymmetric Interest Corridor Systems

Thomas Link　　Ulrike Neyer

Abstract

Standing facilities offered by the central bank are usually considered an effective instrument to control interbank rate volatility. Narrowing the width of the respective interest corridor or installing an asymmetric corridor are seen as appropriate measures to reduce "liquidity shock-induced" volatility. However, since the outbreak of the global financial crisis, interbank market frictions have gained a growing importance. We employ a theoretical model to show that the control of "friction-induced" interbank rate volatility can require different measures. For instance, if friction-induced volatility emerges under an asymmetric corridor, it can be controlled by increasing the corridor width – which is the inversion of the traditional principle.

JEL classification: E43, E52, E58, G21

Keywords: interbank market, monetary policy implementation, interest corridor, floor system, transaction costs, excess reserves

Contents

Contributions, Acknowledgments, and Published Paper Versions

The paper "Controlling Friction-Induced Interbank Rate Volatility under Symmetric and Asymmetric Interest Corridor Systems" is co-authored with Ulrike Neyer. We thank Monika Bucher for her contributions to the first draft of section 3 "Model Setup" and many precious comments. We also thank Jennifer Rontganger for proofreading the paper. Moreover, we thank the anonymous referees in the peer review processes for their comments and suggestions. Furthermore, we thank seminar and conference participants at the Heinrich Heine University Düsseldorf, the University of Gießen, the University of Marburg, the University of Augsburg, the University of Macedonia (Thessaloniki), the Università della Svizzera italiana (Lugano), the University of Auvergne (Clermont-Ferrand), the University of Portsmouth, the University of Vienna, the Technical University of Dortmund, and in Vandans (Austria) for the fruitful discussions and the precious remarks.

Earlier versions of the paper "Controlling Friction-Induced Interbank Rate Volatility under Symmetric and Asymmetric Interest Corridor Systems" have been published as:

Link, T., and U. Neyer (2017): "Friction-Induced Interbank Rate Volatility under Alternative Interest Corridor Systems," DICE Discussion Paper, No. 259, Düsseldorf Institute for Competition Economics (DICE), Düsseldorf, available at `https://www.econstor.eu/bitstream/10419/162837/1/89391472X.pdf`.

Link, T., and U. Neyer (2016): "Transaction Cost Heterogeneity in the Interbank Market and Monetary Policy Implementation under alternative Interest Corridor Systems," Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Monetary Policy, Banks, and Mortgage Markets, No. G12-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz Informationszentrum Wirtschaft, Kiel und Hamburg, available at `https://www.econstor.eu/bitstream/10419/145853/1/VfS_2016_pid_6931.pdf`.

Contributions of Thomas Link

Thomas Link has contributed to all parts and sections of the paper "Controlling Friction-Induced Interbank Rate Volatility under Symmetric and Asymmetric Interest Corridor Systems". In particular, he has contributed in

- doing literature research and doing the theoretical groundwork,
- analyzing the model and conducting comparative statics,
- performing model simulations as well as producing the figures,
- and in writing the discussion.

____

Prof. Dr. Ulrike Neyer

List of Figures

1 Introduction

The design of optimal "post-crisis" monetary policy implementation frameworks remains an open task (see, for instance, Bindseil, 2016; Potter, 2016). A number of issues have to be addressed, like the robustness of any future implementation schemes against recent and coming regulatory reforms.[1] Proposals to adjust implementation frameworks to a new regulatory environment range from varying the structure of open market operations (Bech and Keister, 2017) to changing the symmetry of an established interest corridor system (Jackson and Noss, 2015). This paper addresses the issue of how to design a monetary policy implementation framework that allows for an effective control of interbank rate volatility when volatility stems from market frictions brought about, for instance, by new banking regulations.

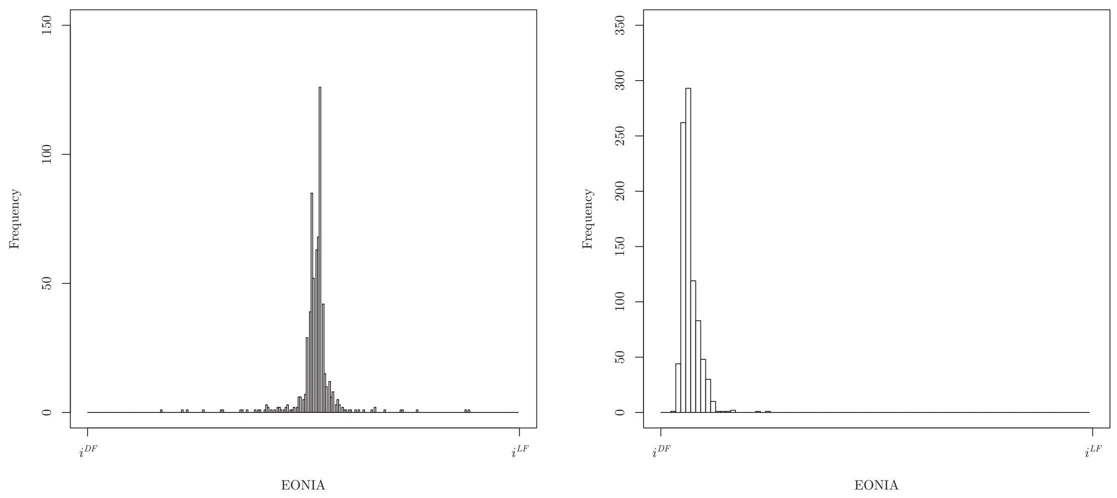
Usually, a monetary policy implementation framework that includes standing facilities is regarded to be effective in controlling the interbank rate. The underlying principle is simple: By providing two standing facilities a central bank creates outside options for banks to using the interbank market.[2] The existence of these options dampens the interest rate effects triggered by liquidity shocks to the banking system. The more attractive these options are made to banks (relative to using the interbank market), the stronger their stabilizing effect on the interbank rate. There are several ways to reach higher attractiveness, such as (1) narrowing the width of the corridor, i.e., the spread between the rates on the deposit and the lending facility or (2) installing an asymmetric corridor system by driving the interbank rate either down (floor system) or up (ceiling system) close to one of the facility rates which makes recourse to that facility highly attractive.[3] The European Central Bank (ECB) has already made experience with both corridor schemes: Figure 1 shows the distribution of daily EONIA rates during two exemplary periods. The figure illustrates the difference between a symmetric corridor and a floor system (which

[1]See Bindseil (2016) for a systematic overview of requirements for future monetary policy implementation frameworks.

[2]In this paper, the term "interbank market" always refers to the unsecured overnight segment of interbank money markets.

[3]The notion of "symmetry" in this context refers to the spreads between the central bank's target interbank rate and the two facility rates. A corridor scheme is symmetric if the target rate is located in the midpoint of the interest corridor. A floor, respectively a ceiling, system is an asymmetric corridor scheme. The target rate in such a system corresponds to the rate on the deposit, respectively the lending, facility. It is implemented by an ample, respectively an insufficient, provision of liquidity. See, for instance, Whitesell (2006), Bindseil, Camba-Mendez, Hirsch, and Weller (2006), Berentsen and Monnet (2008), Bindseil and Jabłecki (2011), Bech and Monnet (2013); for an overview of several options for the design of corridor schemes see United States – Federal Open Market Committee (2008).

the ECB – de facto – is still implementing to date) with respect to the resulting interbank rates. Figure 1 also indicates that during both periods the volatility of the EONIA was, on average, relatively low.



(a) 6 June 2003 to 5 December 2005, 645 observations, deposit rate = 1.00%, lending rate = 3.00%.

(b) 16 March 2016 to 17 September 2019, 897 observations, deposit rate = -0.4%, lending rate = 0.25%.

Figure 1: Distribution of daily EONIA rates under (a) a symmetric corridor system and (b) an asymmetric (floor) system. Horizontal axis: EONIA rate in percentage points. Vertical axis: number of observations. *Data:* European Central Bank – Statistical Data Warehouse (2019).

In particular, figure 1(b) suggests that, until recently, also interbank rate volatility from market frictions was relatively low in the euro area. But there are a couple of potential sources that could account for an increase in friction-induced volatility in the future. In times of crises, information asymmetries about counterparty credit risks or market fragmentation could impair interbank trading and lead to fluctuations of the interbank rate.[4] Also new regulatory burdens can make transactions in the unsecured overnight interbank market less profitable such that banks shift to other options in order to balance their short-term liquidity needs. Theoretically, such shifts can lead to higher interbank rate volatility as already pointed out for the case of new banking regulations by Jackson and Noss (2015), Committee on the Global Financial System and Markets Committee (2015),

[4]See Bucher, Hauck, and Neyer (2020, footnote 9) for a short discussion of why asymmetric information about counterparty credit risks can increase the costs of participating in the interbank market (Bucher, Hauck, and Neyer (2020) mention the costs of signaling the own creditworthiness and the costs of checks for the creditworthiness of counterparties). See also the literature on interbank market frictions that Bucher, Hauck, and Neyer (2020) review. For an analysis of the effects of fragmentation of the banking system respectively the interbank market see, for instance, Vari (2014).

and Bindseil (2016). These studies consider the effects of new banking regulations and argue that concrete measures as a regulatory leverage ratio, large exposure limits, a liquidity coverage ratio, a net stable funding ratio, or suggested risk-based capital requirements for interbank exposures will have an impact on interbank liquidity demand and supply and will lead to higher interbank rate volatility.[5] One rationale for this increase in volatility is that regulatory burdens on banks are bank- and time-specific. Over time, this leads to demand and supply fluctuations in the interbank market that are transmitted into volatility of the interbank rate. This kind of volatility is the starting point of the present paper. With regard to the role of market frictions in a "post-crisis" world the question is whether, and if so how, a central bank can control friction-induced interbank rate volatility.

The aforementioned rules for the control of interbank rate volatility (narrowing the width of the interest corridor, installing an asymmetric corridor) hold if volatility arises from aggregate liquidity shocks. However, we argue that the control of volatility that arises from interbank market frictions is subject to different rules. This is shown in a theoretical analysis which is based on the seminal model of monetary policy implementation under an interest corridor regime presented in Whitesell (2006). We introduce frictions into the Whitesell-model in the form of broadly defined transaction costs that alter the relative attractiveness of outside options for banks to using the interbank market. Transaction cost heterogeneity across banks captures that banks differ in the degree to which they prefer other options than using the interbank market. Ultimately, transaction cost heterogeneity in two dimensions (cross-section and time) explains interbank rate volatility.

In the frictionless benchmark scenarios the model results are in line with those of the existing literature on volatility control under interest corridor systems. Accordingly, the central bank is able to control volatility that arises from aggregate liquidity shocks by increasing the attractiveness of outside options for banks to using the interbank market – concretely, by narrowing the interest corridor or by implementing an asymmetric corridor scheme. However, to control volatility that stems from market frictions, the central bank must create an unattractive outside option to using the interbank market for friction-affected banks (for instance potential lenders which are constrained by capital or liquidity requirements) while maintaining or improving the availability of an attractive outside

[5] Jackson and Noss (2015) consider the effects of a minimum leverage ratio and risk-based capital requirements on the cross-section dispersion of market rates in a multi-agent framework that accounts for the over-the-counter character of interbank markets and, crucially, for the different weights of regulatory burdens for individual banks.

option for banks that are not affected by frictions (for example potential borrowers). Under an initially implemented symmetric corridor scheme this can only be achieved by switching to an asymmetric scheme. Under an initially implemented asymmetric scheme, volatility control requires the central bank to increase the width of the interest corridor.

The rationale for these results is that transaction costs make the use of the interbank market less attractive, interbank market activities decline, and banks fall back on using outside options, that is, on using the standing facilities. The drop in interbank demand/supply is transmitted to the interbank rate. Over time, transaction cost heterogeneity leads to demand, respectively supply, fluctuations that cause interbank rate volatility. While the decline in market activity is stronger the more attractive the outside options for friction-affected banks are, the impact on the interbank rate is stronger the lower the attractiveness of outside options for their potential interbank counterparties is. Thus, interbank rate volatility is higher the higher the attractiveness is of outside options to using the interbank market for friction-affected banks, and the lower the attractiveness of outside options for their potential interbank counterparties. The reason behind these relationships lies in the interest sensitivity of interbank liquidity supply and demand which increases in the attractiveness of outside options for the respective market side. Any measures to reduce interbank rate volatility rely on the exploitation of these properties. Therefore, the control of friction-induced volatility will be possible if the central bank is able to systematically manipulate the attractiveness of outside options for potential lenders and borrowers to a different extent or in opposite directions. Under an initially implemented symmetric corridor system this cannot be achieved by simply changing the corridor width: The symmetry of this scheme implies that any corridor width adjustment will have an equal effect on the attractiveness of outside options for friction-affected banks and their counterparties. The attenuating and the dampening effect on interbank rate volatility cancel each other out. In contrast, the corridor width can be perfectly used as an instrument to reduce friction-induced volatility if the central bank implements an asymmetric corridor scheme, i.e., a floor or a ceiling system. Under a floor system demand-side frictions can no longer be responsible for significant interbank rate volatility and volatility due to supply-side frictions can be controlled by increasing the width of the interest corridor. This measure leads to a stabilization of interbank liquidity supply and therewith of the interbank rate by reducing the attractiveness of potential lenders' outside option to using the interbank market (which is the deposit facility). The asymmetry of a floor system thereby guarantees that the lending banks' potential counterparties still have no attractive outside option available. Analogously, under a ceiling system supply-side frictions can no longer be responsible for interbank rate volatility and widening the corridor is a way of making a ceiling system more robust to demand-side frictions.

Section 2 reviews the related literature. Section 3 presents the model setup. Section 4 derives optimal bank behavior with respect to the banks' use of the central bank's standing facilities and their interbank market activities. This allows for an in-depth analysis of the determinants of banks' liquidity needs, as well as of interbank loan supply and demand. The interbank market equilibrium is identified in Section 5. Then, the implications for the control of interbank rate volatility under a symmetric corridor system (Section 6) and under an asymmetric corridor system (Section 7) are discussed. Section 8 contains some concluding remarks.

2 Related Literature

The model of monetary policy implementation employed in this paper is based on the seminal model of interest corridor systems proposed in Whitesell (2006). Extended versions of that framework have been introduced in several other works, the two closest to the model presented in this paper are those by Bech and Klee (2012) and Jackson and Noss (2015). Whitesell (2006), in turn, is part of a large body of research that refers to the seminal model of an overnight interbank market in Poole (1968). Poole models a representative commercial bank's reserve management and liquidity demand to describe the price formation in the interbank market in the presence of a central bank that provides outside options for banks to using the interbank market. Poole's starting point is that uncertainty about actual liquidity needs during the day explains a precautionary motive behind bank demand for liquidity. This precautionary liquidity demand serves as an explanatory variable for interbank market activities and therefore plays an important role in the analysis of the interbank market equilibrium (see also, for instance, Baltensperger, 1980; Clouse and Dow, 1999; Bech and Monnet, 2013; Bucher, Hauck, and Neyer, 2020). Factors that determine bank demand for precautionary liquidity have been used to explain or predict movements, volatility or observed patterns of the overnight interbank rate, for instance, over reserve maintenance periods or on reserve settlement days. Such determinants are the level of daily interbank payment volumes (Furfine, 2000), lending constraints for banks (Cassola and Huetl, 2010), credit constraints particularly for small banks (Ashcraft, McAndrews, and Skeie, 2011), credit risk (Bech and Klee, 2012), fragmentation of the interbank market (Vari, 2014), regulatory capital requirements (Jackson and Noss, 2015), or broadly defined transaction costs (Bucher, Hauck, and Neyer, 2020). Other determinants with fundamental implications for the optimal design of monetary policy implementation frameworks are reserve requirement schemes (Whitesell, 2006; Gaspar, Pérez Quirós, and Rodríguez Mendizábal, 2008) and specifications of the interest corridor like its width (Woodford, 2001; Bindseil and Jabłecki, 2011) or symmetry (Pérez Quirós and Rodríguez Mendizábal, 2012; Jackson and Noss, 2015).

Similar to Bech and Klee (2012) and Jackson and Noss (2015), this paper starts with the introduction of interbank market transaction costs in the Whitesell-model. Like Jackson and Noss (2015), we explore the transaction cost effect on the price formation in the interbank market and the options a central bank has to react to this effect by adjusting its policy implementation framework. However, we address some issues that have been left open by Jackson and Noss (2015). For instance, we propose rules for the control of friction-induced volatility under a symmetric as well as an asymmetric corridor with a special focus on the width of a respective corridor. In contrast to our approach, Jackson and Noss (2015) conduct their analysis on the basis of a multi-agent framework that captures a crucial institutional detail of the Bank of England's monetary policy implementation framework, namely a system of minimum reserve requirements where banks must meet their reserve requirements within a "target range". One major focus of Jackson and Noss (2015) is on how this reserve target range should be adjusted in the presence of new banking regulations. A second major focus of Jackson and Noss (2015) (and a one that is more closely related to our research question) lies on the symmetry of a corridor system and in particular on the advantages an asymmetric corridor has in the presence of new banking regulations. However, Jackson and Noss (2015) do not place a major focus on the role of the width of the interest corridor (although they make a few comments in this regard: see, for instance, Jackson and Noss, 2015, p. 22). Bech and Klee (2012) put a stronger focus explicitly on the corridor width (in addition to other determinants like transaction costs) as a determinant of the interbank rate respectively of banks' demand for central bank reserves. However, Bech and Klee (2012) are primarily interested in the effect of the corridor width and other factors like transaction costs on the level rather than on volatility of the interbank rate.[6] In our analysis, we are especially interested in the effect of transaction cost heterogeneity across banks and time on the interbank rate in a time dimension.

3 Model Setup

The model introduced in this section is a one-period model based on the framework proposed by Whitesell (2006) and contains some elements of the model presented by Bindseil and Jabłecki (2011).[7] There is a large number of commercial banks and a central bank. The central bank provides settlement accounts for banks, conducts open market operations, and operates two standing facilities. Banks are subject to liquidity shocks and can balance their individual liquidity needs by using the interbank market or the central bank's standing facilities.[8]

Figure 2 illustrates the sequence of events. *At the beginning* of the period under consideration (henceforth called "day"), banks settle their due claims and liabilities from the previous period (for instance, these might stem from overnight interbank loans or from previous recourse to the central bank's facilities). Banks which have insufficient reserve balances for this purpose are allowed to overdraw their settlement accounts during the course of the day. After claims/liabilities are settled, the central bank conducts open market operations and thus injects or withdraws liquidity to/from the banking system. The resulting aggregate liquidity position of the banking sector at that time is denoted by $\bar{\Xi}$. Subsequently, an aggregate liquidity shock α occurs with $\tilde{\alpha} \sim \mathcal{N}(0, \sigma_{AS}^2)$ (with $\tilde{\alpha}$ denoting a random variable, α denoting its realization). Positive values of α indicate liquidity inflows to the banking system, negative values of α liquidity outflows, so banks' aggregate liquidity position after the occurrence of the shock is $\bar{\Xi} + \alpha =: \Xi$. Eventually,

[6] For instance, Bech and Klee (2012) are also interested in the effect of "credit risk" on the interbank rate.

[7] We thank Monika Bucher for her contribution to the first draft of this section at an early stage of the paper.

[8] The elements we take from Bindseil and Jabłecki (2011) are (see Bindseil and Jabłecki, 2011, pp. 14–15): (1) the explicit consideration of two types of banks (and interbank trading), (2) a central bank open market operation at the beginning of the "day", (3) an aggregate liquidity shock directly after the central bank's open market operation, and (4) reserve shifts between banks before interbank trading takes place (Bindseil and Jabłecki (2011) assume that deposits of households are shifted between banks). Note, that Bindseil and Jabłecki (2011) also introduce interbank market transaction costs in their model but our paper is more closely related to Jackson and Noss (2015) and Bech and Klee (2012) in this respect.

bank customers make bank transfer payments which reshuffle reserves within the banking sector.[9]

These activities imply that the banking sector's aggregate liquidity position *at noon*, Ξ, as well as an individual bank's liquidity position *at noon*, denoted by ξ, might be positive or negative. There are two types of commercial banks $i \in \{1, 2\}$: Letting $\xi_1 > 0$ and $\xi_2 < 0$, bank 1 is assumed to have a liquidity surplus at noon, bank 2 a liquidity deficit. If the banking sector's aggregate liquidity position at noon $\Xi = \xi_1 + \xi_2$ is strictly positive (negative), the banking sector as a whole exhibits a liquidity surplus (deficit) vis-à-vis the central bank. Banks have to balance their reserve accounts with the central bank overnight. This setting thus describes an arbitrary day in a world where banks are subject to reserve requirements which have to be precisely fulfilled each day (with end-of-day required reserves being normalized to zero). Alternatively, the period might be interpreted as the last day of a reserve maintenance period where banks are allowed to make use of averaging provisions over the reserve maintenance period. Accordingly, with hypothetical reserve requirements bank 1 would have over-fulfilled reserve requirements at noon to the amount of ξ_1. Bank 2 would not have met reserve requirements but would exhibit a reserve deficiency at noon of $|\xi_2|$.

To balance their liquidity positions *at noon*, banks can use an (overnight) interbank market for central bank reserves. A bank's position in this market is b_i. If $b_i > 0$ ($b_i < 0$), the bank will borrow (lend) the amount $|b_i|$ at rate i^{IBM}. In both cases, transaction costs $\gamma_i |b_i|$ accrue, with $\gamma_i \geq 0$.[10] Following Whitesell (2006), the level of reserve account balances bank i wishes to hold after the closure of the interbank market, its *"target reserve account balance,"* is denoted by T_i with $T_i := \xi_i + b_i$. As intra-day overdrafts are allowed, T_i might be positive or negative.

In the evening, once the interbank market is closed, bank i is hit by an idiosyncratic reserve account shock (a *"late payment shock"*) ϵ_i. The shock ϵ_i is the realization of the random variable $\tilde{\epsilon}_i \sim \mathcal{N}(0, \sigma_i^2)$ with the publicly observable probability density function f_i and the cumulative distribution function F_i.[11] If $\epsilon_i > 0$ ($\epsilon_i < 0$) there will be an inflow

[9] Note that, as mentioned above, in part, this setup is following Bindseil and Jabłecki (2011).

[10] With respect to borrower and lender specific transaction costs our model is most closely related to Bech and Klee (2012) (see, for instance, ibid. pp. 13–18) and Jackson and Noss (2015) (see, for instance, ibid. pp. 26–27).

[11] As argued by Whitesell (2006, p. 1179), a bank does not know its actual liquidity needs for the period under consideration at the time it can trade on the interbank market because it is *"[...] subject to unexpected late payments or delayed accounting information [...]"*. The term *"late payment shock"* in this

(outflow) of funds. The shocks $\tilde{\epsilon}_1$ and $\tilde{\epsilon}_2$ are independent and identically distributed with $f \equiv f_1 \equiv f_2$ and $F \equiv F_1 \equiv F_2$.

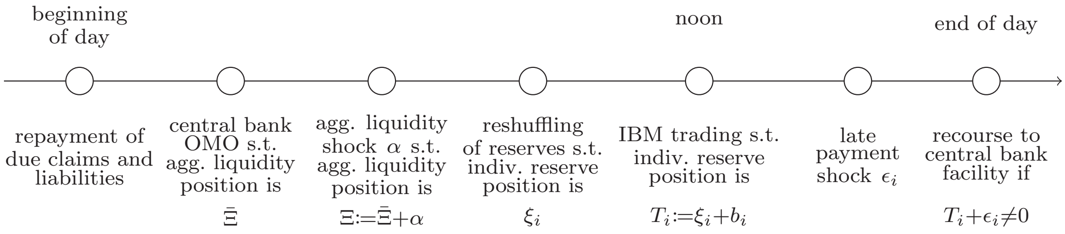


Figure 2: Sequence of events within the period under consideration. To see to what extent we took model ingredients from Bindseil and Jabłecki (2011), see the respective figure in (ibid., p. 16).

A bank's actual *end-of-day* liquidity position is $T_i + \epsilon_i$. Bank i will face an end-of-day deficit if $T_i + \epsilon_i < 0$ and an end-of-day surplus if $T_i + \epsilon_i > 0$. Banks have to balance their reserve accounts with the central bank overnight. Bank i thus has to take recourse to the central bank's lending facility at rate i^{LF} in case of an end-of-day deficit, respectively, to the central bank's deposit facility at rate i^{DF} in case of an end-of-day surplus (banks can obtain/place liquidity from/at the central bank on an overnight basis unlimitedly and without any restrictions). Eventually, at this point in time, bank i learns its actual liquidity costs K_i being the realization of the random variable $\tilde{K}_i$ where

$$\begin{aligned} \tilde{K}_i \; &= i^{IBM} \cdot b_i + \gamma_i \cdot |b_i| \\ &- \; \left(i^{LF} \cdot (T_i + \tilde{\epsilon}_i)\right) \cdot 1_{\{\epsilon_i \leq -T_i\}}(\epsilon_i) \\ &- \; \left(i^{DF} \cdot (T_i + \tilde{\epsilon}_i)\right) \cdot 1_{\{\epsilon_i > -T_i\}}(\epsilon_i) \end{aligned} \tag{1}$$

with $T_i = \xi_i + b_i$.

Actual liquidity costs include the bank's interest costs, resp. revenues, and its transaction costs that accrue when using the interbank market at noon (first line of equation (1)) and – depending on which of the central bank's facilities the bank uses – the interest costs, resp. revenues, of taking recourse to the lending facility (second line), resp. deposit facility (third line). Which of the facilities bank i uses ultimately depends on the late payment shock and finds expression in the values of the indicator functions $1_{\{\epsilon_i \leq -T_i\}}(\epsilon_i)$ and $1_{\{\epsilon_i > -T_i\}}(\epsilon_i)$.

context is used, for instance, in Bech and Monnet (2013), Bindseil, Camba-Mendez, Hirsch, and Weller (2006), and Jackson and Noss (2015).

The rationale behind a bank's interbank market activities is the minimization of its expected liquidity costs. A bank's objective function, yielding its optimal position in the interbank market and therewith its optimal target reserve account balance, is thus given by

$$\begin{aligned} \mathbb{E}[\tilde{K}_i] = - & \left(i^{IBM} + \frac{b_i}{|b_i|} \cdot \gamma_i\right) \cdot \xi_i \\ & - \int_{-\infty}^{-T_i} \left[i^{LF} - \left(i^{IBM} + \frac{b_i}{|b_i|} \cdot \gamma_i\right)\right] \cdot (T_i + \tilde{\epsilon}_i)\, dF \\ & + \int_{-T_i}^{\infty} \left[\left(i^{IBM} + \frac{b_i}{|b_i|} \cdot \gamma_i\right) - i^{DF}\right] \cdot (T_i + \tilde{\epsilon}_i)\, dF \\ & \rightarrow \min_{T_i}! \end{aligned} \tag{2}$$

with $T_i = \xi_i + b_i$.[12]

The bank's optimization problem above differs from the problem in Whitesell (2006, p. 1180) in two respects: Firstly, a bank's pre-trade liquidity endowment ξ_i as well as its position in the interbank market b_i are explicitly considered. This allows for an analysis of the liquidity redistribution within the banking sector via the interbank market.[13] Secondly, interbank market transaction costs γ_i are considered. Bank-specific transaction costs $\gamma_1 \lesseqgtr \gamma_2$ capture the cross-section dimension of transaction cost heterogeneity (the time dimension is considered in sections 6 and 7). In effect, transaction costs increase the relative attractiveness of banks' outside options to using the interbank market, as the expression in the two square brackets formally shows. In the remainder of this paper, the term $\left[i^{LF} - \left(i^{IBM} + \frac{b_i}{|b_i|} \cdot \gamma_i\right)\right]$ will be referred to as "*effective* marginal deficit costs," the term $\left[\left(i^{IBM} + \frac{b_i}{|b_i|} \cdot \gamma_i\right) - i^{DF}\right]$ as "*effective* marginal surplus costs." Therewith, line 2 in equation (2) captures bank i's "expected effective deficit costs" and line 3 its "expected effective surplus costs."

With these extending features, equation (2) states the following:[14] The first line reveals the lower bound for a bank's expected liquidity costs. This lower bound is determined by the bank's pre-trade liquidity endowment. This liquidity would be fully traded in the interbank market at any interbank rate $i^{IBM} > i^{DF} + \gamma_i$ as a lending bank, respectively,

[12] The expression $\frac{b_i}{|b_i|}$ simply captures whether a bank acts as a lender ($\frac{b_i}{|b_i|} = -1$) or borrower ($\frac{b_i}{|b_i|} = 1$) in the interbank market. Of course, optimizing over T_i is equivalent to the optimization over b_i. However, in the following, T_i, which captures a bank's precautionary liquidity demand, is of special interest.

[13] See also, for instance, Bindseil and Jabłecki (2011) or Jackson and Noss (2015) who consider interbank trading explicitly, too.

[14] See also Whitesell (2006, p. 1180) for the original framework.

at any $i^{IBM} < i^{LF} - \gamma_i$ as a borrowing bank if there was no late payment shock. For the interpretation of lines two and three, which reflect the expected liquidity costs due to the late payment shock, it is useful to distinguish between lending and borrowing banks in the interbank market. The second line reveals the expected effective deficit costs of a bank. These are the costs of balancing the expected end-of-day deficit by taking recourse to the lending facility at rate i^{LF}. However, for a bank that has *lent* to the interbank market, and which at the end of the day learns of having placed too much liquidity in the interbank market at noon, the costs of using the lending facility are *effectively* reduced by the interest revenues (minus transaction costs) of the bank's excessive interbank lending. For a borrowing bank, the *effective* costs of using the lending facility are the *additional* costs of using the lending facility instead of the interbank market at noon. The third line captures the expected effective surplus costs which are the effective costs of placing the expected end-of-day liquidity surplus in the deposit facility at rate i^{DF}. For a lending bank, these costs are the (net) *opportunity costs* of using the deposit facility instead of the interbank market. Analogously, for a borrowing bank, the third line of equation (2) reveals the costs of "over-funding" in the interbank market at noon.

4 Optimal Bank Behavior

4.1 Optimal Target Level of Reserve Balances

The first-order condition for optimal, i.e., for expected cost-minimizing, borrowing/lending in the interbank market and thus for the optimal target reserve account balance T_i is

$$\left[i^{LF} - \left(i^{IBM} + \frac{b_i}{|b_i|} \cdot \gamma_i\right)\right] \cdot F(-T_i) \stackrel{!}{=} \left[\left(i^{IBM} + \frac{b_i}{|b_i|} \cdot \gamma_i\right) - i^{DF}\right] \cdot (1 - F(-T_i)) \, . \quad (3)$$

Crucially, as expected liquidity needs due to the late payment shock are zero ($\tilde{\epsilon}_i \sim \mathcal{N}(0, \sigma_i^2)$), T_i represents a bank's demand for precautionary liquidity. With probability $F(-T_i)$, which is decreasing in T_i, (respectively $1 - F(-T_i)$, which is increasing in T_i) the bank faces an end-of-day liquidity deficit (surplus) and has to take recourse to the lending facility (deposit facility). The first-order condition thus implies that the expected marginal return on precautionary liquidity, in the form of avoided illiquidity costs (given by the LHS of (3)), must equal the expected marginal costs of precautionary liquidity (given by the RHS of (3)). Expected marginal costs are in the form of opportunity costs for a bank that lends to the interbank market and in the form of interest costs for a bank that borrows from the interbank market.

For the rest of the paper, it is assumed that the liquidity surplus bank 1 always acts as a lender, whereas the liquidity deficit bank 2 always acts as a borrower in the interbank market. Accordingly, bank 1 increases its target level of reserve balances, T_1, by cutting down its liquidity supply to the interbank market. Respectively, bank 2 increases T_2 by increasing its interbank liquidity demand. This yields a lower bound, respectively an upper bound, for the interbank rate beyond which no interbank trading takes place:

$$\underline{i^{IBM}} := i^{LF} \cdot F(-\xi_1) + i^{DF} \cdot (1 - F(-\xi_1)) + \gamma_1, \tag{4}$$

$$\overline{i^{IBM}} := i^{LF} \cdot F(-\xi_2) + i^{DF} \cdot (1 - F(-\xi_2)) - \gamma_2. \tag{5}$$

4.2 Optimal Precautionary Demand for Reserves in a Frictionless World

In the absence of transaction costs ($\gamma_1 = \gamma_2 = 0$), the case that is discussed by Whitesell (2006), the target reserve account balance that minimizes a bank's expected funding costs is a function of the interbank rate, the rates on the standing facilities, and the parameters of the distribution underlying the late payment shock. T_i is derived from the first-order condition (3) and has the following representation (for an illustration see figure 3):[15]

$$T_i(\cdot) = \begin{cases} -F^{-1}\left(\frac{i^{IBM} - i^{DF}}{i^{LF} - i^{DF}}\right) & \text{if } \left\{i = 1 \wedge i^{IBM} > \underline{i^{IBM}}\right\} \vee \left\{i = 2 \wedge i^{IBM} < \overline{i^{IBM}}\right\} \\ \xi_i & \text{otherwise.} \end{cases} \tag{6}$$

Under the assumption of ϵ_i being distributed symmetrically around zero, the sign of T_i depends only on whether i^{IBM} is above or below the corridor midpoint rate. This is a crucial result in Whitesell (2006, p. 1180): If i^{IBM} lies in the midpoint of the interest corridor, that is, if effective marginal deficit costs just equal the effective marginal surplus costs, optimal bank demand for precautionary liquidity is always zero.

Explicitly considering bank i's pre-trade liquidity endowment ξ_i in addition to its precautionary liquidity demand allows for the analysis of the bank's activity in the interbank market: Accordingly, bank i's interbank liquidity demand (resp. supply) is the sum of a precautionary component T_i and an exogenous component ξ_i:

$$b_i(\cdot) = T_i(\cdot) - \xi_i. \tag{7}$$

[15] F^{-1} denotes the inverse cumulative distribution function.

This decomposition also illustrates that endogenous bank behavior in the interbank market is fully explained by banks' precautionary liquidity demand $T_i(\cdot)$. Crucially, it is this precautionary demand which reflects the degree to which the redistribution of liquidity via the interbank market is inhibited in the presence of market frictions. The remainder of this section discusses the determinants of T_i as an explanatory variable of banks' interbank market activities in more detail. First, we will have a closer look at the interbank rate i^{IBM}, including the respective interest sensitivity of T_i. Then, we will comment on the width w of the interest corridor formed by the rates on the facilities i^{DF} and i^{LF}, and on a bank's pre-trade reserve account balance ξ_i. The discussion serves as the theoretical base for the analysis in Sections 6 and 7.

Interbank Rate

Equation (6) reveals that a bank's precautionary liquidity demand decreases in i^{IBM}. Obviously, an increase in i^{IBM} makes precautionary liquidity holdings relatively less attractive: Liquidity surplus banks want to place a higher amount in the interbank market, whereas liquidity deficit banks are willing to cover a higher portion of a potential deficit by borrowing from the central bank's lending facility. Formally, this reads:[16]

$$\frac{\partial T_i}{\partial i^{IBM}} = \frac{\partial b_i}{\partial i^{IBM}} = -\frac{1}{f(-T_i)(i^{LF} - i^{DF})} \leq 0. \tag{8}$$

With respect to the interest sensitivity of a bank's demand for precautionary liquidity, it is crucial how strongly the probability of facing an end-of-day deficit $F(-T_i)$ reacts to changes in T_i (resp. to changes in b_i) as formally revealed by the first-order condition (3). If there is only a weak response, interest sensitivity (in absolute value) will be high because then there must be a relatively strong increase or decrease in T_i to have a sufficiently high impact on $F(-T_i)$ to restore optimality after a change in i^{IBM}. As $\tilde{\epsilon}_i \sim \mathcal{N}(0, \sigma_i^2)$, the impact of a change in T_i on $F(-T_i)$ is lower the more T_i deviates from 0 in either

[16] Equation (8) can be derived explicitly by differentiating (6) or by using (3) and applying the implicit function theorem.

direction, i.e., the more precautionary liquidity in absolute terms bank i holds. Formally, this is reflected by[17]

$$\frac{\partial^2 T_i}{\partial (i^{IBM})^2} = \frac{\partial^2 b_i}{\partial (i^{IBM})^2} = \frac{f'(-T_i)}{(i^{LF} - i^{DF})^2 \cdot (f(-T_i))^3} \begin{cases} < 0 & \text{if } T_i < 0 \\ = 0 & \text{if } T_i = 0 \\ > 0 & \text{if } T_i > 0. \end{cases} \tag{9}$$

However, for the interest sensitivity of a bank's precautionary liquidity demand it is also decisive how strongly the expected marginal return/costs of precautionary liquidity react to changes in $F(-T_i)$. This is determined by the width of the interest corridor formed by i^{DF} and i^{LF}. The wider the interest corridor is, the more pronounced the expected marginal return/costs of precautionary liquidity will react to changes in $F(-T_i)$, that is, the lower is the interest sensitivity of precautionary liquidity demand. This is because the wider the interest corridor is, the larger the spreads between the interbank rate and the facility rates might possibly become. For an individual bank, such large spreads imply relatively high effective marginal deficit or surplus costs. Accordingly, only a relatively small change in T_i, and therewith in the probabilities of using the facilities, is needed to have a sufficiently strong effect on the expected marginal return on or the expected marginal costs of precautionary liquidity to restore optimality after a change in i^{IBM}, as shown formally by (3). Considering symmetric changes of the interest corridor around some given corridor midpoint rate i^{MR}, with $i^{DF} \equiv i^{MR} - w$ and $i^{LF} \equiv i^{MR} + w$, it is

$$\frac{\partial^2 T_i}{\partial w \partial i^{IBM}} = \frac{\partial^2 b_i}{\partial w \partial i^{IBM}} = \frac{1}{2w^2 \cdot f(-T_i)} - \frac{f'(-T_i) \cdot (2 \cdot F(-T_i) - 1)}{4w^2 \cdot (f(-T_i))^3} \geq 0, \tag{10}$$

which formally shows that the interest sensitivity of bank i's precautionary liquidity demand (in absolute value), and therewith the interest sensitivity of interbank demand (for $i = 2$) and supply (for $i = 1$), decreases in the width of the corridor.[18]

[17] To check the sign of $\frac{\partial^2 T_i}{\partial (i^{IBM})^2}$, recall that $f(\cdot)$ is the probability density of the normal distribution with $f'(-T_i) < 0$ for $T_i < 0$, $f'(-T_i) > 0$ for $T_i > 0$, and $f'(T_i) = 0$ for $T_i = 0$.

[18] Recall, that $F(\cdot)$ is the cumulative distribution function of the normal distribution. To check the sign of (10), recall the property discussed in fn. 17 and recall that $F(-T_i) > 0.5$ for $T_i < 0$, $F(-T_i) < 0.5$ for $T_i > 0$, and $F(-T_i) = 0.5$ for $T_i = 0$.

Width of the Interest Corridor

In general, the width of the interest corridor is a crucial determinant of a bank's precautionary liquidity demand and thus its interbank liquidity demand/supply: A symmetric increase in the corridor width leads to an increase in a bank's effective marginal deficit and surplus costs and therewith to an increase in the expected marginal return on and the expected marginal costs of precautionary liquidity. The increase in the expected marginal return will outweigh the increase in the expected marginal costs if the bank targets a negative reserve account balance ($T_i < 0$), which implies that the probability of using the lending facility at the end of the day is greater than 0.5. Consequently, the bank will increase the level of its precautionary liquidity holdings. Analogously, if the bank targets a positive reserve account balance, an increase in the corridor width will induce the bank to decrease its target reserve account balance. This formally reads (for an illustration see figure 3):

$$\frac{\partial T_i}{\partial w} = \frac{\partial b_i}{\partial w} = \frac{2 \cdot F(-T_i) - 1}{f(-T_i) \cdot 2w} \begin{cases} > 0 & \text{for } T_i < 0 \\ = 0 & \text{for } T_i = 0 \\ < 0 & \text{for } T_i > 0. \end{cases} \tag{11}$$

The effect of a change in the width of the corridor on a bank's precautionary liquidity demand (resp. on interbank liquidity demand/supply) is stronger the more T_i deviates from zero. The more T_i deviates from zero, the higher the probability is that one of the facilities will be used after the occurrence of the late payment shock, hence the larger the difference in the changes in the expected marginal return on and the expected marginal costs of precautionary liquidity implied by a change in w. Consequently, as formally reflected by (10), a relatively pronounced change in T_i is needed to restore optimality after a change in the corridor width.

Pre-trade Reserve Account Balance

As a bank's precautionary liquidity demand is independent from its pre-trade reserve account balance per construction, a change in the bank's pre-trade reserve account balance is reflected completely in its interbank liquidity demand/supply:

$$\frac{\partial b_i}{\partial \xi_i} = -1. \tag{12}$$

Equations (8) to (12) illustrate that the surplus bank's precautionary liquidity demand, T_1, as well as the deficit bank's precautionary liquidity demand, T_2, are qualitatively affected in the same way by changes in i^{IBM}, w, and ξ. In contrast, as discussed in the next section, interbank market transaction costs will have opposing effects on T_i, for $i = 1, 2$.

4.3 Optimal Precautionary Demand for Reserves in the Presence of Transaction Costs

The impact of interbank market transaction costs on bank i's precautionary demand for reserves depends on whether bank i acts as a lender or as a borrower in the interbank market. Explicit representations of banks' target reserve account balances that minimize their expected funding costs in the presence of transaction costs are given by

$$T_1(\cdot) = \begin{cases} -F^{-1}\left(\frac{i^{IBM}-\gamma_1-i^{DF}}{i^{LF}-i^{DF}}\right) & \text{if } i^{IBM} > \underline{i^{IBM}} \\ \xi_1 & \text{if } i^{IBM} \leq \underline{i^{IBM}}, \end{cases} \tag{13}$$

$$T_2(\cdot) = \begin{cases} -F^{-1}\left(\frac{i^{IBM}+\gamma_2-i^{DF}}{i^{LF}-i^{DF}}\right) & \text{if } i^{IBM} < \overline{i^{IBM}} \\ \xi_2 & \text{if } i^{IBM} \geq \overline{i^{IBM}}. \end{cases} \tag{14}$$

For the surplus bank 1, an increase in γ_1 implies that holding precautionary liquidity becomes more attractive as the alternative of placing excess liquidity in the interbank market becomes more expensive. Formally, transaction costs lead to an increase in effective marginal deficit costs (the term in square brackets on the LHS of equation (3)), and to a decrease in effective marginal surplus costs (the term in square brackets on the RHS of equation (3)). Consequently, bank 1 reduces its interbank liquidity supply. For the deficit bank 2, an increase in γ_2 implies that holding precautionary liquidity becomes less attractive, as borrowing the respective liquidity from the interbank market becomes more expensive. As a result, the deficit bank 2 reduces its liquidity demand in the interbank market. Formally, this reads (for an illustration see figure 3):

$$\frac{\partial T_1}{\partial \gamma_1} = \frac{\partial b_1}{\partial \gamma_1} = \frac{1}{f(-T_1)\cdot(i^{LF}-i^{DF})} > 0, \tag{15}$$

$$\frac{\partial T_2}{\partial \gamma_2} = \frac{\partial b_2}{\partial \gamma_2} = \frac{-1}{f(-T_2)\cdot(i^{LF}-i^{DF})} < 0. \tag{16}$$

Analogously to the interest sensitivity, the transaction cost sensitivity of bank i's precautionary liquidity demand is higher (in absolute value) the less $F(-T_i)$ reacts to changes in T_i (resp. to changes in b_i) and the less the expected marginal return on or marginal costs of precautionary liquidity react to changes in $F(-T_i)$. Thus, the transaction costs sensitivity of banks' precautionary liquidity demand (in absolute value) is higher the more T_i deviates from zero, and the narrower the interest corridor is. Formally, this is captured by equations (17) to (20):[19]

$$\frac{\partial^2 T_1}{\partial i^{IBM} \partial \gamma_1} = \frac{\partial^2 b_1}{\partial i^{IBM} \partial \gamma_1} = \frac{-f'(-T_1)}{(i^{LF} - i^{DF})^2 \cdot (f(-T_1))^3} \begin{cases} > 0 & \text{if } T_1 < 0 \\ = 0 & \text{if } T_1 = 0 \\ < 0 & \text{if } T_1 > 0, \end{cases} \tag{17}$$

$$\frac{\partial^2 T_2}{\partial i^{IBM} \partial \gamma_2} = \frac{\partial^2 b_2}{\partial i^{IBM} \partial \gamma_2} = \frac{f'(-T_2)}{(i^{LF} - i^{DF})^2 \cdot (f(-T_2))^3} \begin{cases} < 0 & \text{if } T_2 < 0 \\ = 0 & \text{if } T_2 = 0 \\ > 0 & \text{if } T_2 > 0, \end{cases} \tag{18}$$

$$\frac{\partial^2 T_1}{\partial w \partial \gamma_1} = \frac{\partial^2 b_1}{\partial w \partial \gamma_1} = \frac{-1}{2w^2 \cdot f(-T_1)} + \frac{f'(-T_1) \cdot (2 \cdot F(-T_1) - 1)}{4w^2 \cdot (f(-T_1))^3} \leq 0, \tag{19}$$

$$\frac{\partial^2 T_2}{\partial w \partial \gamma_2} = \frac{\partial^2 b_2}{\partial w \partial \gamma_2} = \frac{1}{2w^2 \cdot f(-T_2)} - \frac{f'(-T_2) \cdot (2 \cdot F(-T_2) - 1)}{4w^2 \cdot (f(-T_2))^3} \geq 0. \tag{20}$$

5 Interbank Market Equilibrium

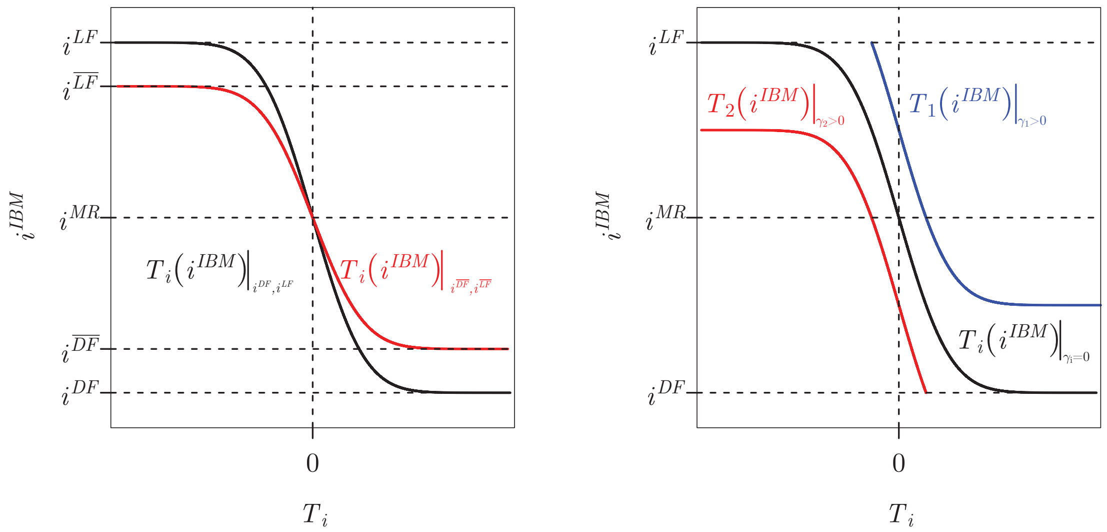
Indicating the equilibrium variables with the superscript *, the interbank market clearing condition reads

$$\sum_i b_i^*(\cdot) = 0. \tag{21}$$

Considering (7) and denoting the aggregate of banks' precautionary liquidity demand with $T \coloneqq \sum_i T_i$ and the banking sector's aggregate liquidity endowment with $\Xi = \sum_i \xi_i$, the market clearing condition (21) can be rewritten as

$$T^* \left(i^{IBM^*}, i^{DF}, i^{LF}, \gamma_1, \gamma_2, \sigma_i \right) = T_1^* \left(i^{IBM^*}, \gamma_1, \cdot \right) + T_2^* \left(i^{IBM^*}, \gamma_2, \cdot \right) \stackrel{!}{=} \Xi. \tag{22}$$

[19] To check the signs of equations (17) to (20), recall the properties discussed in footnotes 17 and 18.



(a) Corridor width effect.

(b) Transaction cost effect.

Figure 3: Individual banks' precautionary liquidity demand T_i for $i = 1, 2$ and its determinants. Horizontal axis: Quantity of precautionary liquidity demanded. Vertical axis: Interbank rate. (a) Impact of corridor width (see also Whitesell, 2006, p. 1181). (b) Impact of borrowers' (red line) and lenders' (blue line) transaction costs on their individual precautionary liquidity demand, respectively (see also Bech and Klee, 2012, p. 14). Note that the figures were drawn by assuming that $|\xi_i|$ is sufficiently large for both $i = 1, 2$ such that the lower and the upper bound given by equations (4) and (5) do not become binding in the region of the parameter space considered in the figures.

Equation (22) illustrates that the interbank market will clear at an interbank rate at which the banking sector's aggregate precautionary liquidity demand T is equal to its pre-trade liquidity endowment Ξ. Hence, the liquidity in the banking sector is redistributed via the interbank market such that each bank ends up with its optimal level of precautionary liquidity holdings.

However, a crucial distinguishing feature of the model presented in this paper as compared to the Whitesell (2006) framework is that individual banks might differ in their precautionary liquidity demand because of different interbank market transaction costs. Crucially, the difference in the quantities of precautionary liquidity demanded by the two types of banks (T_1^* and T_2^* in equation (22)) will reflect the *degree* to which the liquidity redistribution via the interbank market is inhibited (i.e., the extent to which transaction costs reduce the interbank market transaction volume). The larger this difference is, the smaller is the extent to which banks use the interbank market to balance their pre-trade liquidity surplus/deficit, respectively, the heavier is their reliance on the standing facilities to balance their reserve accounts at the end of the day. Clearly, if transaction costs are prohibitively high, no interbank trade will take place at all. But since we want to focus on interbank rate volatility rather than on a complete breakdown of the interbank market, we rule out the case that transaction costs can become prohibitively high for the remainder of the paper. With respect to the equilibrium interbank rate i^{IBM^*} this implies that i^{IBM^*} will always be located between the lower and the upper bound given by equations (4) and (5) such that interbank trade will take place.

The equilibrium interbank rate i^{IBM^*} is implicitly given by equation (23) which is obtained by inserting (13) and (14) into (22):

$$F^{-1}\left(\frac{i^{IBM^*} - \gamma_1 - i^{DF}}{i^{LF} - i^{DF}}\right) + \xi_1 + F^{-1}\left(\frac{i^{IBM^*} + \gamma_2 - i^{DF}}{i^{LF} - i^{DF}}\right) + \xi_2 \overset{!}{=} 0. \qquad (23)$$

The equilibrium level of individual banks' precautionary liquidity demand T_i^* and therewith the equilibrium interbank transaction volume $b^* := b_2^* = -b_1^*$ is implicitly given by

$$F(-T_1^*) - F(-T_2^*) + \frac{\gamma_1 + \gamma_2}{i^{LF} - i^{DF}} \overset{!}{=} 0, \qquad (24)$$

which is obtained from the first-order condition (3) for bank $i = 1, 2$.

6 Volatility Control Under a Symmetric Corridor System

This section derives specific rules for the control of friction-induced interbank rate volatility under a symmetric corridor system. First, in Section 6.1, a frictionless benchmark scenario is considered. Within this scenario, a comparative static analysis of the interbank market equilibrium is conducted, and model simulations illustrate the dispersion of interbank rates (as a proxy for volatility). Then, in Section 6.2, an analogous analysis is made considering interbank market frictions.

6.1 Frictionless World

6.1.1 Comparative Statics

The frictionless scenario ($\gamma_1 = \gamma_2 = 0$) considered in this section is the benchmark scenario for the subsequent section. The main results are in line with the respective findings of Whitesell (2006). The two standing facility rates form a symmetric corridor around the central bank's targeted interbank rate i^{target} such that $i^{DF} = i^{target} - w$ and $i^{LF} = i^{target} + w$. The target rate thus corresponds to the mid-point rate of the interest corridor $i^{MR} := \frac{1}{2} \cdot (i^{DF} + i^{LF})$.

A crucial feature of this implementation scheme with regard to the central bank's steering of the interbank rate is that banks' precautionary liquidity demand at the target rate $T(i^{target})$ is zero (Whitesell, 2006; Woodford, 2001). This property holds independently of the absolute level of the facility rates, the width of the interest corridor, and the level of banks' pre-trade liquidity endowments $\xi_1,\ \xi_2$:

Property 1 (Demand for Precautionary Liquidity):

$$T(i^{target}) = 0 \ \ \textit{for any} \ \ i^{target} = i^{MR}, \ \textit{and for any} \ \ w,\ \xi_1,\ \xi_2. \tag{25}$$

Formally, Property 1 follows directly from the first-order condition (3). In the absence of transaction costs the condition will be satisfied at $i^{IBM} = i^{MR}$ if bank i targets a reserve account balance of zero. In particular, $T_i = 0$ implies that the bank will face an end-of-day liquidity deficit and surplus with the same probability; and exactly this is what optimality requires when the effective marginal deficit and surplus costs are of equal height, i.e., when $i^{LF} - i^{IBM} = i^{IBM} - i^{DF}$ which is the case at $i^{IBM} = i^{MR}$.[20]

With a predictable aggregate demand for precautionary liquidity (equal to zero at i^{target}), the only source of deviations of the interbank rate from the central bank's target is the central bank's inability to perfectly control the liquidity conditions in the banking system. In this paper, such an aggregate liquidity shock is captured by the realization of the random variable $\tilde{\alpha}$. In the absence of transaction costs, the interbank market will clear at $i^{target} = i^{MR}$ if the banking sector's aggregate liquidity position $\Xi = \bar{\Xi} + \alpha = 0$. This means that also its aggregate precautionary liquidity demand at i^{target} must be zero, as revealed by equation (22). However, with $\tilde{\alpha} \sim \mathcal{N}(0, \sigma^2_{AF})$ and a central bank that therefore chooses $\bar{\Xi} = 0$, the banking sector's pre-trade liquidity position after the occurrence of the shock, at noon, is $\Xi = \alpha$. The implicit differentiation of (23) formally shows the interest rate effects that are produced by any liquidity imbalances:[21]

[20] See also Whitesell (2006), p. 1180-1181.

[21] Recall that $\Xi = \xi_1 + \xi_2$.

Property 2 (Liquidity Effect):

$$\frac{\partial i^{IBM^*}}{\partial \xi_1} = \frac{\partial i^{IBM^*}}{\partial \xi_2} = - \frac{\left(i^{LF} - i^{DF}\right) \cdot f\left(-\xi_1 - b_1^*\right) \cdot f\left(-\xi_2 - b_2^*\right)}{f\left(-\xi_1 - b_1^*\right) + f\left(-\xi_2 - b_2^*\right)} < 0. \tag{26}$$

It is conventional wisdom that these effects (and thus the effect of an aggregate liquidity shock on the interbank rate) are weaker the narrower the interest corridor is. Equation (11) reveals this property: The narrower the interest corridor is, the more attractive the facilities are as outside options for banks to using the interbank market and thus the larger the interest sensitivity (in absolute value) of banks' precautionary liquidity demand is (if $\gamma_1 = \gamma_2 = 0$). This leads to[22]

Property 3 (Corridor Width Effect): *Narrowing the corridor width reduces possible deviations of the interbank rate from its target, i.e.,*

$$\frac{\partial i^{IBM^*}}{\partial w} = \frac{i^{IBM^*} - i^{MR}}{w} \begin{cases} > 0 & for\ \Xi < 0 \\ = 0 & for\ \Xi = 0 \\ < 0 & for\ \Xi > 0. \end{cases} \tag{27}$$

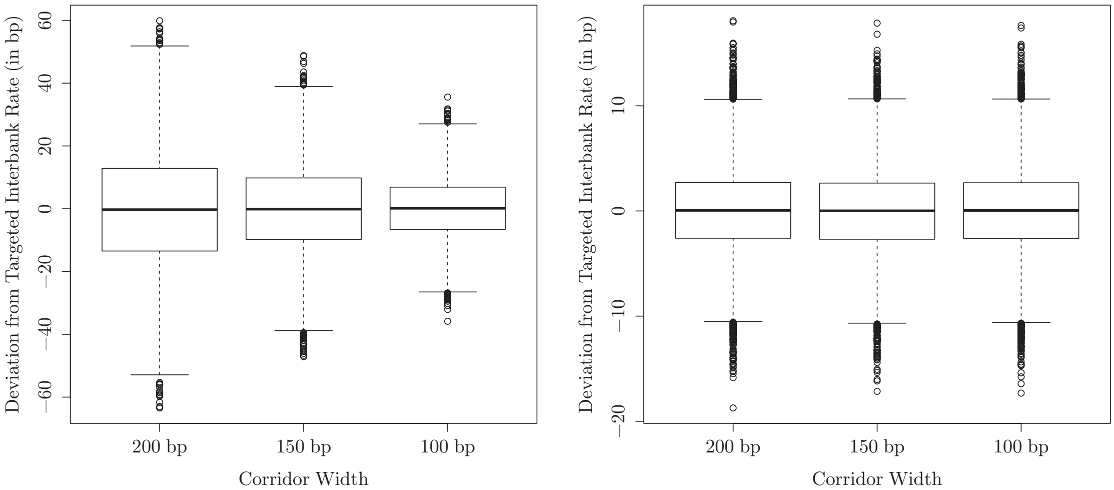
6.1.2 Distribution of Interbank Rates and Model Simulations

Of course, the employed one-period model does not explain the evolution of interbank rates over time but it does predict how a time series of interbank rates consistent with the model parameters would be distributed. The dispersion of this distribution is then a proxy for interbank rate volatility. Thus, implications for the sources of interbank rate volatility and for the measures to control this volatility can be drawn from the results of the comparative static analysis by mapping them into a parameter space with a time dimension. In this regard, the employed model yields some empirically testable hypotheses, formulated in the following as *"Implications."* Properties 1–3 imply, respectively:

Implication 1 (Source of Interbank Rate Volatility): *The only source of interbank rate volatility is the aggregate liquidity shock α. Banks' precautionary liquidity demand does not cause any interbank rate volatility since at the target rate $T(i^{target})$ is zero with certainty and thus stable over time, i.e., from period to period or from "day to day."*[23]

[22] To check the sign of equation (27) note that in the absence of transaction costs, the interbank market will clear at $i^{IBM^*} = i^{MR}$ if the banking sector's aggregate liquidity position is balanced, i.e., if $\Xi = 0$. An aggregate liquidity deficit ($\Xi < 0$) will drive the interbank rate above i^{MR}, an aggregate liquidity surplus will drive the interbank rate below i^{MR}.

[23] See also Whitesell (2006), Woodford (2001).



(a) Aggregate liquidity shock effects. (b) Transaction cost heterogeneity effects.

Figure 4: Distribution of simulated interbank rates under a symmetric corridor system for different values of the corridor width visualized with boxplots (lines of the box: first, second, and third quartile; whiskers: max. 1.5 inter-quartile range). Basic parameter values are: $\tilde{\epsilon}_i \sim \mathcal{N}(0,1)$, $\bar{\Xi} = 0$. Subfigure (a) shows the results of 10,000 draws of $\tilde{\alpha}$ for each corridor width where $\tilde{\alpha} \sim \mathcal{N}(0, 0.5^2)$, $\gamma_1 = \gamma_2 = 0$. Subfigure (b) shows the results of 10,000 independent draws of γ_1 and γ_2 (for each corridor width) from the truncated normal distribution $\mathcal{N}(0, 0.1^2)|_0^{0.4}$ with $\tilde{\alpha}$ kept constant at zero. The simulated interbank rates have been obtained in MATLAB (see The MathWorks, Inc., 2016) by using the function "`fsolve`" to solve equation (22) numerically. Note, to rule out the possibility of a breakdown of the interbank market in the simulations, we implicitly assume that $|\xi_i|$ is sufficiently large for both $i = 1, 2$. This guarantees that the interbank rate will be located between the lower and the upper bound given by equations (4) and (5) for the region of the parameter space considered in the simulations.

Implication 2 (Distribution of Interbank Rates): *The distribution of a time series of interbank rates consistent with the model is determined by the distribution of aggregate liquidity shocks.*

Implication 3 (Dispersion of Interbank Rates and Corridor Width Effect): *Regarded over time, the dispersion of interbank rates is lower the smaller the width of the interest corridor set by the central bank is. Thus, the corridor width can be systematically used to reduce interbank rate volatility.*

Figure 4(a) illustrates the relationship between the corridor width and interbank rate volatility that stems from aggregate liquidity shocks. For specifically chosen parameter values, the model was solved for 10,000 draws of $\tilde{\alpha}$. The dispersion of simulated interbank rates (a proxy for interbank rate volatility) is increasing in the corridor width.[24]

6.2 Consideration of Transaction Cost Heterogeneity

6.2.1 Comparative Statics

Interbank market transaction costs increase the relative attractiveness of outside options for banks to using the interbank market. Thus, transaction costs induce banks to substitute away from the use of interbank loans to balance their reserve accounts at noon toward an increased reliance on the central bank's standing facilities at the end of the day. Such shifts are reflected in the levels of banks' precautionary liquidity demand:[25]

Property 4 (Transaction Cost Effect on Precautionary Liquidity Demand): *In the presence of transaction costs banks will target a higher (if they are potential interbank lenders), resp. lower (if they are potential interbank borrowers), level of precautionary liquidity holdings than in the frictionless case. Accordingly, the interbank market transaction volume will be lower, i.e., the liquidity redistribution via the interbank market will be inhibited, formally stated by*

$$\frac{\partial b^*}{\partial \gamma_1} = \frac{\partial b^*}{\partial \gamma_2} = -\frac{1}{(i^{LF} - i^{DF}) \cdot (f(-T_1^*) + f(-T_2^*))} < 0, \tag{28}$$

$$\frac{\partial T_1^*}{\partial \gamma_i} = -\frac{\partial b^*}{\partial \gamma_i} > 0 \textit{ for } i = 1, 2, \tag{29}$$

$$\frac{\partial T_2^*}{\partial \gamma_i} = \frac{\partial b^*}{\partial \gamma_i} < 0 \textit{ for } i = 1, 2. \tag{30}$$

The existence of transaction costs leads to the following key property of a symmetric corridor system:

Property 5 (Demand for Precautionary Liquidity): *Transaction costs imply that the banking sector's aggregate demand for precautionary liquidity at the target rate may differ from zero, i.e.,*

$$T(i^{target}) \lesseqgtr 0 \quad \textit{if } \gamma_1, \gamma_2 \geq 0. \tag{31}$$

[24] This simulation approach follows Whitesell (2006).

[25] Recall that with the convention $b^* := b_2^* = -b_1^*$ it is $T_1^* = \xi_1 - b^*$ and $T_2^* = \xi_2 + b^*$.

Formally, Property 5 is implied by equations (15) and (16) which show that the banking sector's aggregate precautionary liquidity demand T is an increasing function of potential lenders' transaction costs and a decreasing function of potential borrowers' transaction costs. The quantities of precautionary liquidity demanded by banks thereby depend on the width of the interest corridor. A narrow corridor leads to relatively large deviations from zero of banks' precautionary liquidity demand at i^{target} (as discussed in section 4.3 and as formally captured by equations (19) and (20)).

So, with regard to the central bank's liquidity management there are two cases that have to be distinguished: (1) There is no heterogeneity in the cross-section dimension, i.e., $\gamma_1 = \gamma_2$. This implies that $T_1(i^{target}) = -T_2(i^{target})$ so that $T(i^{target}) = 0$, independent of which banks will be active on which side of the interbank market. This means that there is no uncertainty about $T(i^{target})$ and also no need for the central bank to accommodate any demand for precautionary liquidity by the banking sector as a whole. (2) There is transaction cost heterogeneity in the cross-section dimension, i.e., $\gamma_1 \neq \gamma_2$, implying that $T_1(i^{target}) \neq -T_2(i^{target})$ and $T(i^{target}) \neq 0$. Since bank customer payments reshuffle reserves within the banking sector after the central bank has conducted its open market operations and before interbank trading takes place, the central bank does not know "in the early morning" which banks will be active on which interbank market side "at noon." Hence, the banking sector's aggregate precautionary liquidity demand is uncertain and the establishment of adequate liquidity conditions in the early morning to hit the targeted interbank rate requires the central bank to estimate $T(i^{target})$. Forecast errors result in deviations of the equilibrium interbank rate from the target level. Formally, the interest rate effects of such unobservable transaction cost heterogeneity are captured by:

Property 6 (Pass-through of Transaction Costs on the Interbank Rate):

$$\frac{\partial i^{IBM*}}{\partial \gamma_1} = \frac{f(-T_2^*)}{f(-T_1^*) + f(-T_2^*)} > 0, \tag{32}$$

$$\frac{\partial i^{IBM*}}{\partial \gamma_2} = \frac{-\ f(-T_1^*)}{f(-T_1^*) + f(-T_2^*)} < 0. \tag{33}$$

As formally stated by the following Property 7, the magnitude of these effects depends on the width of the interest corridor:

Property 7 (Corridor Width Effect): *Possible deviations of the interbank rate from its target are reduced either by a widening or a narrowing of the corridor width, i.e.,*

$$\frac{\partial^2 i^{IBM^*}}{\partial w \partial \gamma_1} = \frac{\frac{\partial b^*}{\partial w} \cdot \left(\frac{-\xi_1 - \xi_2}{\sigma^2} \cdot f(-\xi_1 + b^*) \cdot f(-\xi_2 - b^*) \right)}{\left(f(-\xi_1 + b^*) + f(-\xi_2 - b^*)\right)^2} \begin{cases} > 0 & \text{for } \Xi < 0 \\ = 0 & \text{for } \Xi = 0 \\ < 0 & \text{for } \Xi > 0, \end{cases} \tag{34}$$

$$\frac{\partial^2 i^{IBM^*}}{\partial w \partial \gamma_2} = \frac{\frac{\partial b^*}{\partial w} \cdot \left(\frac{-\xi_1 - \xi_2}{\sigma^2} \cdot f(-\xi_1 + b^*) \cdot f(-\xi_2 - b^*) \right)}{\left(f(-\xi_1 + b^*) + f(-\xi_2 - b^*)\right)^2} \begin{cases} > 0 & \text{for } \Xi < 0 \\ = 0 & \text{for } \Xi = 0 \\ < 0 & \text{for } \Xi > 0, \end{cases} \tag{35}$$

where

$$\frac{\partial b^*}{\partial w} = \frac{\gamma_1 + \gamma_2}{2w^2 \left(f(-\xi_1 + b^*) + f(-\xi_2 - b^*)\right)} \geq 0. \tag{36}$$

Equations (34) and (35) give the formal description that the corridor width cannot be used systematically to make the interbank rate robust to bank-specific transaction costs under a symmetric corridor system. The intuition is simple. While, as argued in section 4.2, the high interest sensitivity of interbank liquidity demand and supply under a narrow corridor from the central bank's perspective is desirable in a frictionless world, it is ambivalent in the presence of transaction costs. The following considerations for the case of supply-side transaction costs illustrate this ambivalence: Lending transaction costs lead to a drop in interbank liquidity supply. This drop is larger the more interest-sensitive the supply is. Hence, the upward pressure on the interbank rate implied by a transaction cost-induced drop in supply is larger the more interest-sensitive the supply is. This is the case under a narrow corridor where banks have relatively attractive outside options available and depend less on the interbank market. Now, the ambivalence of a narrow corridor in this respect is revealed when the demand side of the interbank market is considered. If demand is highly interest-sensitive, the equilibrium interbank rate is relatively robust to transaction-cost induced changes in supply. So, in a comparative static view, a reduction in the corridor width, which makes both the interbank demand and supply more interest-sensitive, has two opposing effects on the interbank rate and on the magnitude of the lending-transaction-cost effect on the interbank rate. The sign of the overall effect depends on the extent to which a corridor-width reduction increases the interest sensitivity of demand compared to the extent to which a corridor-width reduction increases the interest sensitivity of supply. Demand effects will dominate if there is a scarcity of aggregate liquidity, $\Xi < 0$, supply effects will dominate if there is an excess of aggregate liquidity, $\Xi > 0$. Both effects will be of the same magnitude if aggregate liquidity conditions are balanced and, in this case, the lending-transaction-cost effect on the interbank rate will even be independent of the corridor width (see equation (34)).

So, in the presence of transaction cost heterogeneity ($\gamma_1 \neq \gamma_2$), whether a relatively wide or narrow corridor is suitable for minimizing the deviations of the interbank rate from its target thus depends on the sign of the banking sector's pre-trade liquidity position Ξ. However, the sign of Ξ under a symmetric corridor system is determined by the aggregate liquidity shock. Therefore, there is no general rule the central bank could follow in order to implement a symmetric corridor system that is relatively "robust" to lending transaction cost effects and – with an analogous argumentation – to borrowing transaction cost effects.[26]

6.2.2 Distribution of Interbank Rates and Model Simulations

Again, implications for the sources of and the measures to control the volatility of a time series of interbank rates in a multi-period world consistent with the model can be drawn by mapping the comparative static results into a parameter space that has a time dimension. Now, the interesting case is the one where interbank rate volatility stems from market frictions. This might be the case in a world where new banking regulations are fully phased in, as discussed in Bindseil (2016), Committee on the Global Financial System and Markets Committee (2015), or Jackson and Noss (2015). The ultimate rationale for the increase in volatility caused by banking regulations is that the financial weights of regulatory burdens that banks have to carry are bank- and time-specific. Transaction cost heterogeneity, as introduced in this paper, captures the nature of such frictions in the cross-section dimension and can easily be thought further into a time dimension:

Definition (Transaction Cost Heterogeneity in Two Dimensions): *Transaction cost heterogeneity in the two dimensions cross-section and time is present if, regarded over time, potential interbank lenders' and borrowers' transaction costs γ_1 and γ_2 change independently from period to period (or from "day to day").*

[26] These results can be derived formally from (17), (18), (19), and (20)

In a world with transaction cost heterogeneity in two dimensions, regarded over time, Properties 4 to 7 have the following implications:

Implication 4 (Two Sources of Interbank Rate Volatility): *Transaction cost heterogeneity in the two dimensions cross-section and time is a source of interbank rate volatility in addition to the first source that lies in the aggregate liquidity shock. This is because banks' precautionary liquidity demand at the target rate $T(i^{target})$ will be unstable over time, i.e., from period to period or from "day to day," if γ_1 and γ_2 change independently over time. Moreover, $T(i^{target})$ is uncertain at the time the central bank conducts open market operations. Hence, the central bank cannot perfectly offset daily fluctuations in $T(i^{target})$ by adequate provision of liquidity. The daily fluctuations in $T(i^{target})$ cause fluctuations in interbank liquidity demand/supply that are transmitted into the interbank rate.*

Implication 5 (Distribution of Interbank Rates): *The distribution of a time series of interbank rates consistent with the model is determined by the distribution of the time series of potential lenders' and borrowers' transaction costs and by the distribution of aggregate liquidity shocks.*

Implication 6 (Dispersion of Interbank Rates and Corridor Width Effect): *The dispersion of a time series of interbank rates that is explained by transaction cost heterogeneity under a symmetric corridor system cannot be systematically lowered by adjusting the width of the interest corridor. This is a direct implication of Property 7. Thus, the width of the interest corridor is not an instrument for the systematic control of interbank rate volatility that stems from transaction cost heterogeneity. For the special case in a world without an aggregate liquidity shock, such that the banking sector's aggregate liquidity position at noon is always balanced ($\Xi = 0$), volatility stemming from frictions is not even correlated with the corridor width.*

Figure 4(b) illustrates the neutral relationship between the corridor width and volatility that stems from transaction cost heterogeneity for the special case of a world without an aggregate liquidity shock, i.e., of a world in which $\alpha = 0$. The crucial point is that the dispersion of a time series of interbank rates in this special case is determined only by the dispersion of the time series of transaction costs and is thus independent of the corridor width. In order to illustrate this relationship, the model was solved for 10,000 draws of lending and borrowing transaction costs from a truncated normal distribution.

In summary, the results above suggest that a central bank which chooses to operate a symmetric corridor system in the presence of transaction cost heterogeneity will be confronted with a kind of “white noise” volatility stemming from frictions that cannot be controlled through adjustments in the corridor width.[27]

7 Volatility Control Under an Asymmetric Corridor System

This section conducts the same analysis as the previous section assuming that the central bank implemented an asymmetric corridor system. Section 7.1, considering a frictionless world, provides some comparative statics with respect to the interbank rate and model simulations which illustrate the dispersion of the interbank rate in a floor system. Section 7.2 does the same analysis considering interbank market frictions. Section 7.3 uses these results to draw some conclusions for a ceiling system.

7.1 Frictionless World

7.1.1 Comparative Statics

A floor system is an asymmetric corridor scheme where the central bank's targeted interbank rate corresponds to the rate on the deposit facility (for analytical traceability let $i^{target} = i^{DF} + \delta$ for some small $\delta > 0$).[28] The implementation of this scheme by itself – through an ample central bank provision of liquidity – produces a relatively stable interbank rate that will fluctuate only marginally around the target rate. The corridor width of a floor system as an instrument to control interbank rate volatility therefore plays a less relevant role – at least in the frictionless benchmark scenario considered in this subsection.

The basic idea when implementing a floor system is to exploit the following two properties of banks' aggregate precautionary liquidity demand:

Property 8 (Demand for Precautionary Liquidity):

$$T(i^{target}) >> 0 \quad for \quad |i^{target} - i^{DF}| \approx 0. \tag{37}$$

[27] With regard to the control of volatility stemming from aggregate liquidity shocks, the model implies a further property that also is in line with conventional wisdom (see, for instance, Bindseil and Jabłecki, 2011, section 4). Considering $\frac{\partial b^*}{\partial w} = \frac{\gamma_1+\gamma_2}{2w^2\left(f(-T_1^*)+f(-T_2^*)\right)} \geq 0$ for the case of $\gamma_1 + \gamma_2 > 0$ reveals that, if transaction costs are present, the central bank must take into account that narrowing the interest corridor leads to a reduction of trading activity in the interbank market.

[28] See, for instance, United States – Federal Open Market Committee (2008). The “asymmetry” of this scheme lies in the difference of the spreads between i^{target} to i^{DF} and to i^{LF}.

Property 9 (Interest Sensitivity of T): *Using equation* (9) *for* $\frac{\partial^2 T_i}{\partial (i^{IBM})^2}$ *it is*

$$\frac{\partial^2 T}{\partial (i^{IBM})^2} = \frac{\partial^2 T_1}{\partial (i^{IBM})^2} + \frac{\partial^2 T_2}{\partial (i^{IBM})^2} > 0 \quad \textit{if} \;\; T_{1,2} > 0. \tag{38}$$

Eventually, the interbank market will clear at the targeted rate if there is virtually zero risk for banks to become illiquid at the end of the day due to the late payment shock, that is, if $F(-T_i^*) \approx 0$.[29] This will be the case if the banking sector's aggregate liquidity endowment at noon after the realization of the aggregate liquidity shock, Ξ, still sufficiently exceeds its expected liquidity needs (which are zero), that is, if $\Xi = T(i^{target}) >> 0$.

Thus, in order to implement an interbank rate close to i^{DF} the central bank must use its open market operations in the early morning to provide an ample amount of liquidity $\bar{\Xi} >> 0$ such that only extreme left-tail events described by $\alpha << 0$ could increase the probability of and end-of-day deficit for banks significantly above zero. So, with $\bar{\Xi} \to \infty$, the liquidity risk posed by left-tail events converges to zero, that is, $F(-T_i^*)$ will remain close to zero and will be insensitive even to relatively large pre-trade aggregate liquidity shocks. The first-order condition (3) illustrates that this insensitivity of the cumulative distribution function $F(\cdot)$ translates into a high interest sensitivity of demand for precautionary liquidity, which in turn translates into a high interest sensitivity of interbank demand and supply.[30]

Therewith, the interbank rate will be insensitive to aggregate liquidity shocks if the interbank liquidity demand and supply curves always intersect at their highly interest-sensitive regions even after large aggregate pre-trade liquidity drains. This will be the case if the banking sector's aggregate liquidity endowment $\bar{\Xi}$ (which is the central bank's choice) is sufficiently large. Thus, the principle of tight interbank rate control under a floor system with $\Xi > 0$ relies on a relatively weak liquidity effect (as implied by Property 9).

7.1.2 Distribution of Interbank Rates and Model Simulations

Mapping the comparative static results for the frictionless benchmark scenario under a floor system into a parameter space with a time dimension yields the same implications for the source of interbank rate volatility and the distribution of interbank rates as in

[29] This property is implied by equation (8) in section 4.2.

[30] Formally, this is captured by Property 9 which is implied by equation (9) in section 4.2. See also Poole (1968, p. 774).

the benchmark scenario under a symmetric corridor system. Thus, the only source of interbank rate volatility is the aggregate liquidity shock $\tilde{\alpha}$ with the distribution of a time series of interbank rates consistent with the model being determined by the distribution of aggregate liquidity shocks. However, with regard to the role played by the corridor width in attenuating the effects of aggregate liquidity shocks on the interbank rate, there is the following:

Implication 7 (Dispersion of Interbank Rates and Corridor Width Effect): *Although, regarded over time, the dispersion of interbank rates is lower the smaller the width of the interest corridor is, a key feature of a floor system is that this effect of the corridor width on the dispersion of interbank rates is negligible. The corridor width as an instrument to control interbank rate volatility is less relevant.*

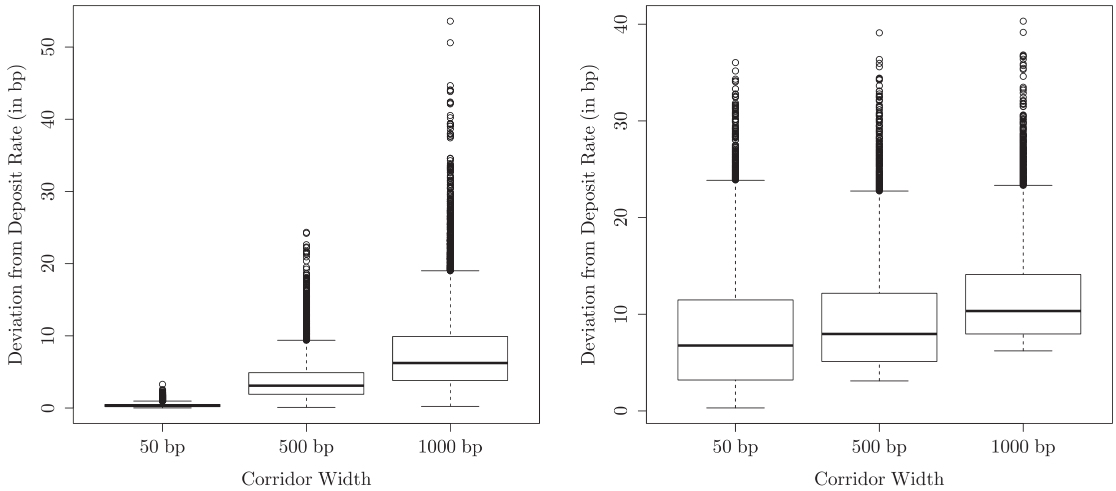
This crucial property is illustrated by figure 5(a) which shows that the dispersion of simulated interbank rates (as a proxy for interbank rate volatility) is relatively low even under a relatively wide interest corridor.

7.2 Consideration of Transaction Cost Heterogeneity

7.2.1 Comparative Statics

Transaction cost heterogeneity in the two dimensions cross-section and time can make the interbank rate more volatile – as is the case under a symmetric corridor. However, the central bank is able to exploit some of the properties implied by the characteristic asymmetry of a floor system to reduce friction-induced interbank rate volatility in a systematic manner. The control of volatility that stems from supply-side transaction costs even involves the implementation of a relatively wide interest corridor.

The mechanisms at work are tied to the asymmetry in the pass-through rates of lending and borrowing transaction costs that exists under a floor system. Under permanent excess liquidity conditions established in a floor system, $\Xi > 0$, the banking sector as a whole has to – and in particular the liquidity surplus banks have to – rely more on the deposit facility, implying that the interest sensitivity of interbank liquidity supply is always greater than or equal to the interest sensitivity of demand. So the pass-through rate of lending transaction costs is always greater than or equal to the pass-through rate of borrowing transaction costs. The crucial point is that the central bank can systematically use the corridor width to reduce the pass-through rate of one market side's transaction costs –



(a) Aggregate liquidity shock effects.

(b) Transaction cost heterogeneity effects.

Figure 5: Distribution of simulated interbank rates under a floor system for different values of the corridor width visualized with boxplots (lines of the box: first, second, and third quartile; whiskers: max. 1.5 inter-quartile range). Basic parameter values are: $\tilde{\epsilon_i} \sim \mathcal{N}(0, 1)$, $\bar{\Xi} = 5$ (we have also used other parameter constellations but decided to choose this parameter constellation as it makes the effect of a wide corridor on the dispersion of interbank rates well visible; the effect is harder to spot for other values of $\bar{\Xi}$). Subfigure (a) shows the results of 10,000 draws of $\tilde{\alpha}$ for each corridor width where $\tilde{\alpha} \sim \mathcal{N}(0, 0.5^2)$, $\gamma_1 = \gamma_2 = 0$. Subfigure (b) shows the results of 10,000 draws of γ_1 (for each corridor width) from the truncated normal distribution $\mathcal{N}(0, 0.1^2)|_0^{0.4}$ with γ_2 and $\tilde{\alpha}$ kept constant at zero. The simulated interbank rates have been obtained in MATLAB (see The MathWorks, Inc., 2016) by using the function “`fsolve`” to solve equation (22) numerically. Note, to rule out the possibility of a breakdown of the interbank market in the simulations, we implicitly assume that $|\xi_i|$ is sufficiently large for both $i = 1, 2$. This guarantees that the interbank rate will be located between the lower and the upper bound given by equations (4) and (5) for the region of the parameter space considered in the simulations.

although at the expense of the other side's pass-through rate of transaction costs. If the corridor width is increased, it is the pass-through rate of lending transaction costs that decreases because it is the surplus banks which react most strongly to the decline in the attractiveness of outside options – arguing analogously to the case considered in section 6.2. So, under permanent excess liquidity conditions Property 7 is reduced to the special case of

Property 10 (Corridor Width Effect):

$$\frac{\partial^2 i^{IBM*}}{\partial w \partial \gamma_1} = \frac{\frac{\partial b^*}{\partial w} \cdot \left(\frac{-\xi_1 - \xi_2}{\sigma^2} \cdot f(-\xi_1 + b^*) \cdot f(-\xi_2 - b^*) \right)}{\left(f(-\xi_1 + b^*) + f(-\xi_2 - b^*) \right)^2} < 0 \; for \;\; \Xi > 0, \tag{39}$$

$$\frac{\partial^2 i^{IBM*}}{\partial w \partial \gamma_2} = \frac{\frac{\partial b^*}{\partial w} \cdot \left(\frac{-\xi_1 - \xi_2}{\sigma^2} \cdot f(-\xi_1 + b^*) \cdot f(-\xi_2 - b^*) \right)}{\left(f(-\xi_1 + b^*) + f(-\xi_2 - b^*) \right)^2} < 0 \; for \;\; \Xi > 0. \tag{40}$$

Under a floor system, demand-side effects are less of a concern for the central bank. On the one hand, potential borrowers' transaction costs bring the interbank rate even closer to the central bank's target. And on the other hand, borrowing transaction costs only involve a relatively small drop in the deficit banks' precautionary liquidity demand (due to the relatively low interest sensitivity of demand) such that the demand for interbank liquidity (at low interbank rates close to i^{target}) and therewith the equilibrium interbank rate will be relatively insensitive to transaction costs, too. As the interbank rate is inherently robust to demand-side frictions under a floor system, volatility control would require the central bank to implement a corridor system that makes the interbank rate robust to supply-side frictions. As captured by Property 10, this is achieved by implementing a relatively wide interest corridor.[31]

7.2.2 Distribution of Interbank Rates and Model Simulations

The implications of the comparative static results for the sources of interbank rate volatility and the distribution of interbank rates under a floor system are the same as under a symmetric corridor system in the presence of transaction costs. With transaction cost heterogeneity in the two dimensions cross-section and time in addition to the aggregate

[31] The ultimate reason for this stabilizing effect is that effective marginal surplus costs and thus opportunity costs of liquidity banks hold in excess of their expected liquidity needs increase in the corridor width. Consequently, targeting large quantities of precautionary liquidity, and thus the outside option of using the deposit facility, become relatively unattractive for surplus banks (i.e., for potential lenders). A wide corridor stabilizes potential lenders' precautionary liquidity demand, therewith interbank liquidity supply and ultimately the interbank rate.

liquidity shock representing the two sources of interbank rate volatility, the distribution of a time series of interbank rates consistent with the model is determined by the distribution of the time series of potential lenders' and borrowers' transaction costs and by the distribution of aggregate liquidity shocks. But Property 10 yields the following:

Implication 8 (Dispersion of Interbank Rates and Corridor Width Effect): *A high dispersion of a time series of potential lenders' transaction costs leads to a relatively pronounced increase in the dispersion of a time series of interbank rates. A high dispersion of a time series of potential borrowers' transaction costs only leads to a relatively small increase in the dispersion of a time series of interbank rates. Thus, demand-side effects are less of a concern. With regard to the control of friction-induced interbank rate volatility the main implication is thus that the dispersion of a time series of interbank rates that is explained by the dispersion of a time series of potential lenders' transaction costs can be systematically lowered by increasing the width of the interest corridor, as illustrated by figure 5(b).*[32]

7.3 Implications for a Ceiling System

The analysis made in the previous section leads to some direct implications for a ceiling system. This is an asymmetric corridor scheme where the targeted interbank rate corresponds to the central banks' lending rate. The central bank can implement a ceiling system by making sure that the banking sector's aggregate liquidity position stays significantly negative. In turn, the aggregate liquidity deficit drives up the interbank rate. Similar to a floor system, this scheme is robust against aggregate liquidity shocks even for a relatively wide interest corridor. However, under a ceiling system, potential interbank borrowers have the more attractive outside option to using the interbank market in the presence of market frictions. Thus, as the pass-through rate of potential borrowers' transaction costs on the interbank rate is larger than that of potential interbank lenders, supply-side effects are less relevant under this scheme. So, if the main source of volatility lies in borrowing transaction costs, friction-induced interbank rate volatility can be controlled systematically by increasing the corridor width of a ceiling system.

[32] At that, a possibly desirable property for the central bank is that, in the presence of transaction costs, widening the interest corridor leads to an increase in the interbank market transaction volume (formally stated by $\frac{\partial b^*}{\partial w} = \frac{\gamma_1+\gamma_2}{2w^2\left(f(-T_1^*)+f(-T_2^*)\right)} \geq 0$). For a discussion of trade-offs a central bank faces when choosing the width of an interest corridor see, for instance, Bindseil and Jabłecki (2011).

8 Concluding Remarks

Interbank market frictions can lead to higher interbank rate volatility. Bank- and time-specific transaction costs can cause fluctuations in interbank liquidity demand and supply that will be transmitted into interbank rate volatility. New banking regulations that pose additional financial burdens on interbank market participants might have such a volatility effect (Bindseil, 2016; Committee on the Global Financial System and Markets Committee, 2015; Jackson and Noss, 2015). Thus, at some point, central banks could actually be confronted with interbank rate volatility that stems from market frictions.

This paper points out that the control of interbank rate volatility which has its origin in market frictions basically is subject to different rules than the control of volatility that stems from aggregate liquidity shocks to the banking sector. Generally, a central bank's options to control volatility that stems from frictions are to switch from a symmetric corridor system to an asymmetric corridor system (floor or ceiling system) and to increase the width of the asymmetric corridor. Under a symmetric corridor system, the corridor width cannot be used systematically at all to control friction-induced volatility. Under a floor system, the interbank rate is inherently robust to demand-side frictions, under a ceiling system it is robust to supply-side frictions. Consequently, if under a symmetric corridor system demand-side (supply-side) frictions are the dominant source of interbank rate volatility, the central bank should switch to a floor system (ceiling system) to reduce the volatility. If a floor system (ceiling system) is already implemented, still occurring friction-induced volatility will then be the result of supply-side (demand-side) frictions. The control of this volatility requires the central bank to widen the interest corridor – which is the inversion of the traditional principle.

Bibliography

Ashcraft, A., J. McAndrews, and D. Skeie (2011): "Precautionary Reserves and the Interbank Market," *Journal of Money, Credit, and Banking*, 43(s2), 311–348.

Baltensperger, E. (1980): "Alternative Approaches to the Theory of the Banking Firm," *Journal of Monetary Economics*, 6(1), 1–37.

Bech, M. L., and T. Keister (2017): "Liquidity Regulation and the Implementation of Monetary Policy," *Journal of Monetary Economics*, 92, 64–77.

Bech, M. L., and E. Klee (2012): "Frictions in the Interbank Market and the Demand for Reserves: Lessons from the Financial Crisis," unpublished manuscript, November 2012, available at https://pdfs.semanticscholar.org/2c9a/41b00f49d1aed9998fc319ccfc861e1e782b.pdf.

Bech, M. L., and C. Monnet (2013): "The Impact of Unconventional Monetary Policy on the Overnight Interbank Market," in *Liquidity and Funding Markets*, ed. by A. Heath, M. Lilley, and M. Manning, RBA Annual Conference Volume (Discontinued), pp. 147–177. Reserve Bank of Australia.

Berentsen, A., and C. Monnet (2008): "Monetary Policy in a Channel System," *Journal of Monetary Economics*, 55(6), 1067–1080.

Bindseil, U. (2016): "Evaluating Monetary Policy Operational Frameworks," Paper for presentation at the Jackson Hole Economic Policy Symposium 2016, August 2016, Federal Reserve Bank of Kansas City, available at https://www.kansascityfed.org/~/media/files/publicat/sympos/2016/econsymposium-bindseil-paper2.pdf?la=en.

Bindseil, U., G. Camba-Mendez, A. Hirsch, and B. Weller (2006): "Excess Reserves and the Implementation of Monetary Policy of the ECB," *Journal of Policy Modeling*, 28(5), 491–510.

Bindseil, U., and J. Jabłecki (2011): "The Optimal Width of the Central Bank Standing Facilities Corridor and Banks' Day-to-Day Liquidity Management," Working Paper Series No. 1350, European Central Bank.

Bucher, M., A. Hauck, and U. Neyer (2020): "Interbank Market Friction-Induced Holdings of Precautionary Liquidity: Implications for Bank Loan Supply and Monetary Policy Implementation," *Economic Theory*, 70(1), 165–222.

Cassola, N., and M. Huetl (2010): "The Euro Overnight Interbank Market and ECB's Liquidity Management Policy During Tranquil and Turbulent Times," Working Paper Series No. 1247, European Central Bank.

Clouse, J. A., and J. P. Dow (1999): "Fixed Costs and the Behaviour of the Federal Funds Rate," *Journal of Banking and Finance*, 23(7), 1015–1029.

Committee on the Global Financial System and Markets Committee (2015): "Regulatory Change and Monetary Policy," CGFS Papers No. 54, Bank for International Settlements, available at https://www.bis.org/publ/cgfs54.htm.

European Central Bank – Statistical Data Warehouse (2019): "EONIA," [Data file], retrieved from https://sdw.ecb.europa.eu/quickview.do?SERIES_KEY=198.EON.D.EONIA_TO.RATE&resetBtn=+Reset+Settings&start=&end=&trans=N, 21 October 2019.

Furfine, C. H. (2000): "Interbank Payments and the Daily Federal Funds Rate," *Journal of Monetary Economics*, 46(2), 535–553.

Gaspar, V., G. Pérez Quirós, and H. Rodríguez Mendizábal (2008): "Interest Rate Dispersion and Volatility in the Market for Daily Funds," *European Economic Review*, 52(3), 413–440.

Jackson, C., and J. Noss (2015): "A Heterogeneous Agent Model for Assessing the Effects of Capital Regulation on the Interbank Money Market under a Corridor System," Staff Working Paper No. 548, Bank of England.

Pérez Quirós, G., and H. Rodríguez Mendizábal (2012): "Asymmetric Standing Facilities: An Unexploited Monetary Policy Tool," *IMF Economic Review*, 60(1), 43–74.

Poole, W. (1968): "Commercial Bank Reserve Management in a Stochastic Model: Implications for Monetary Policy," *The Journal of Finance*, 23(5), 769–791.

Potter, S. (2016): "Implementing Monetary Policy Post-Crisis: What Have We Learned? What Do We Need to Know?," Remarks at a workshop organized by Columbia University SIPA and the Federal Reserve Bank of New York, New York City, 4 May 2016, available at https://www.newyorkfed.org/newsevents/speeches/2016/pot160504 and https://www.bis.org/review/r160510a.htm.

The MathWorks, Inc. (2016): "MATLAB, Version R2016a," Natick, Massachusetts.

United States – Federal Open Market Committee (2008): "Memo: Interest on Reserves: A Preliminary Analysis of Basic Options," in United States – Federal Open Market Committee. "Meeting, April 29-30, 2008," Federal Open Market Committee Meeting Minutes, Transcripts, and Other Documents (April 29-30, 2008): 1–53, available at https://fraser.stlouisfed.org/title/677/item/23274/content/pdf/fomc20080401memo02/toc/517175.

Vari, M. (2014): "Implementing Monetary Policy in a Fragmented Monetary Union," Document de Travail No. 529, Banque de France.

Whitesell, W. (2006): "Interest Rate Corridors and Reserves," *Journal of Monetary Economics*, 53(6), 1177–1195.

Woodford, M. (2001): "Monetary Policy in the Information Economy," NBER Working Paper No. 8674, National Bureau of Economic Research, Published in: *Proceedings – Economic Policy Symposium – Jackson Hole*, Federal Reserve Bank of Kansas City, 2001, pp. 297–370.

Optimal Timing of Calling In Large-Denomination Banknotes under Natural Rate Uncertainty

Thomas Link

Abstract

The elimination of large-denomination banknotes is one of several options to relax the effective-lower-bound constraint on nominal interest rates. We explore timing issues associated with the calling-in of large notes from a central banker's perspective and employ an optimal stopping model to show how the volatility and the expected path of the natural rate of interest determine an optimal timing strategy. Our model shows that such a strategy can involve a wait-and-see component analogously to an optimal exercise rule for a perpetual American option. In practice, a wait-and-see component might induce a central banker not to call in large notes until the natural rate has fallen to an exceptionally low level.

JEL classification: E42, E58

Keywords: cashless economy, phase-out of paper currency, wait-and-see policy, option value

Contents

Contributions, Acknowledgments, and Published Paper Versions

The paper “Optimal Timing of Calling In Large-Denomination Banknotes under Natural Rate Uncertainty” is a single author paper by Thomas Link. I am grateful to Ulrike Neyer for the numerous discussions and many precious remarks. I also thank Jennifer Rontganger for proofreading the paper. Furthermore, I thank seminar and conference participants at the Heinrich Heine University Düsseldorf, the University of Gießen, Sciences Po Aix (Aix-en-Provence), and Leipzig University for the fruitful discussions and many helpful comments. I am especially grateful to Michalis Macrominas for a fruitful discussion on real options and games after the “33rd International Symposium on Money, Banking and Finance” at the University of Auvergne in Clermont-Ferrand in 2016. The impetus for this paper grew out of this discussion.

An earlier version of the paper “Optimal Timing of Calling In Large-Denomination Banknotes under Natural Rate Uncertainty” has been published as:

Link, T. (2019): “Optimal Timing of Calling In Large-Denomination Banknotes under Natural Rate Uncertainty,” DICE Discussion Paper, No. 327, Düsseldorf Institute for Competition Economics (DICE), Düsseldorf, available at `http://www.dice.hhu.de/fileadmin/redaktion/Fakultaeten/Wirtschaftswissenschaftliche_Fakultaet/DICE/Discussion_Paper/327_Link.pdf`.

A very early draft from which the papers “Optimal Timing of Calling In Large-Denomination Banknotes under Natural Rate Uncertainty” and “International Coordination and Optimal Timing of Calling In Large-Denomination Banknotes in a Two-Player Game” (included as third essay in this thesis) emerged has been made available online for presentation at the “35th International Symposium on Money, Banking and Finance”, organized by the European Research Group (GdRE) on Money, Banking and Finance, and held at Sciences Po Aix (Aix-en-Provence) in June 2018. This early draft is available as unpublished manuscript at `https://gdre-scpo-aix.sciencesconf.org/195408/document`.

List of Figures

List of Tables

1 Introduction

It is well established that cash implies an "effective lower bound (ELB)" on nominal interest rates and thus a constraint for monetary policy.[1] Proposals to relax the ELB-constraint range from abolishing cash straightaway (for instance Buiter, 2009), on the one hand, to implementing measures that reduce the attractiveness of cash as an outside option in times of negative nominal interest rates on the other hand. Concrete measures of this kind include issuing banknotes with a "magnetic strip" that can be used to enforce holders of such banknotes to pay a carry tax on currency (Goodfriend, 2000, p. 1016) or "phasing out" large-denomination banknotes (Rogoff, 2017a, p. 57).[2] While Goodfriend's approach relies on technical ways to enforce higher costs of carry for cash artificially, Rogoff's approach aims at directly increasing the costs associated with cash hoardings like transportation, storage, and insurance costs.[3] In this paper, we take up Rogoff's proposal and consider the implied decision problem of a central banker that has the power to implement it. Our starting point is the observation that Rogoff's approach to lower the ELB is easy to implement, practical (in contrast to the other two approaches mentioned), scalable, and, most notably, that it is in fact feasible within the mandate and set of instruments of at least one major monetary authority. Evidence has recently been provided by the European Central Bank that stopped the issuance of 500-euro notes in April 2019.[4]

A central bank like the European Central Bank (ECB) with a clear mandate and the primary objective of maintaining price stability should face a relatively narrow and well-defined problem in deciding whether or not to phase out large notes – especially since it is not the complete "abolition" of cash that is at stake, but just an adjustment in the

[1] For a discussion on a "zero bound" or a "zero lower bound" see, for instance, Goodfriend (2000, p. 1007), Buiter (2009, p. 214), or Rogoff (2017a, p. 47). For the rationale of why the "effective" lower bound on policy rates, in fact, lies below zero see, for instance, Goodfriend (2000, footnote 3), Buiter and Rahbari (2015), or Rogoff (2017a, p. 59). However, it has also been pointed out that cash is not the only reason that the ELB exists: We refer to Rogoff (2017a, p. 61) and (as also cited therein) to McAndrews (2015) for a discussion of various other frictions that had to be tackled in order for central banks to be able to effectively implement negative rates.

[2] See also Buiter and Panigirtzoglou (2003) for a detailed discussion with a comment on the feasibility and a theoretical analysis of the "carry tax" approach. Buiter (2007a) discusses an alternative to that approach which, in principle, involves the introduction of an exchange rate between cash and central bank reserves. A discussion of the approach to "phase out" large-denomination banknotes is also presented in Rogoff (2016) and Rogoff (2017b). For a survey of several approaches to relax the ELB-constraint and a discussion of feasibility issues see, for instance, Agarwal and Kimball (2019).

[3] See Rogoff (2017a, p. 59).

[4] Agarwal and Kimball (2019, p. 44) suggest that the decision to curtail the denominational structure of banknotes has increased the scope for the ECB to lower its policy rates – simply because the physical effort of storing multiples of €500 in cash and thus the hurdle to take a flight into euro-denominated cash are higher now (which is the exact same rationale put forward by Rogoff, 2017a, p. 59).

denominational structure of banknotes.[5] No weight should be assigned to arguments that stand against completely abolishing cash, such as the loss of privacy or of the possibility to make payments independent of information technology and access to the internet (for a central bank like the ECB, such arguments should in any case be subordinate to monetary policy goals).[6] However, even for a "cash-averse" central banker, there is a major reason to keep issuing large banknotes: As Rogoff (2015, pp. 450–452) suggests, it is natural to assume that phasing out large notes will reduce cash demand and thus the seignorage revenue a central bank makes by issuing cash.[7] Rogoff (2015, p. 452) states that the loss of seignorage profits is in so far problematic for a central bank as its capacity to finance itself and to thus shield its operational independence is put at stake.[8] The decision to phase out large banknotes thus involves a major dilemma for a central bank when it must trade off the benefits for monetary policy from a relaxed ELB-constraint against the loss of seignorage revenues. In this paper, a central bank's decision problem is, in principle, reduced to this dilemma because we believe that other arguments for or against the issuance of large notes, for instance, of 200- and 100-euro notes, should be less relevant for a central bank with a clear mandate like the ECB.

It is obvious that the intensity of this dilemma is state-dependent and uncertain in many dimensions with the net benefits from phasing out large banknotes dependent on the likeliness, frequency, and scale of ELB-episodes in the future as well as on the amount of forgone seignorage revenues. We focus on one key determinant and start with the hypothesis that both factors, i.e., the probability and costs of ELB-episodes in the future as well as seignorage losses, are a monotone function of the natural rate of interest. Model-based simulation results that point toward the assumption that a lower natural rate level involves a higher probability of ELB-episodes are provided by Kiley and Roberts (2017) and Chung, Gagnon, Nakata, Paustian, Schlusche, Trevino, Vilán, and Zheng (2019). These studies assess the likelihood of ELB-episodes in the United States for different states of the world and different interest rate levels and find a significant risk in some

[5] According to Mersch (2018), the ECB, particularly, the Governing Council of the ECB, can in fact adjust the denominational structure of banknotes. Moreover, as Mersch (2018) points out, it is *only* the Governing Council of the ECB that may adjust the denominational structure of euro banknotes.

[6] For detailed discussions of arguments in favor of and against the issuance of cash, respectively large-denomination banknotes, see, for instance, Rogoff (1998), Rogoff (2015), Rogoff (2017a) or Krüger and Seitz (2018).

[7] We use the term "seignorage" for central bank revenue from issuing cash, noting that there are other measures of seignorage (see, for instance, Buiter, 2007b).

[8] See also Rogoff (2016, chapter 6), Buiter (2009, p. 224), Thiele, Niepelt, Krüger, Seitz, Halver, and Michler (2015, p. 10), or Krüger and Seitz (2017, chapter 4.1).

scenarios that the ELB-constraint will become binding again in the future.[9] In light of Holston, Laubach, and Williams (2017) who observe a significant fall of natural rates in the United States, Canada, the euro area, and the United Kingdom during the last three decades, it is thus natural to assume that the probability of ELB-episodes in those other regions, ceteris paribus, is also now significantly higher than it was three decades ago.[10] Altogether, with seignorage revenues typically decreasing in the interest rate level (we comment on this relationship in section 2), we build our analysis on the assumption that the net benefits from phasing out large notes are higher the lower the natural rate of interest is.

Rogoff (2016, chapter 7) discusses how a "phase out" of large-denomination banknotes could be implemented. In principle, the implementation schemes he considers range, on the one hand, from removing the legal tender status of certain banknotes without delay or at relatively short notice (similar to the calling-in of 500- and 1,000-rupee notes in India in 2016) to, on the other hand, a "soft" implementation version where certain banknotes are gradually removed from circulation over time by simply stopping their issuance while keeping their legal-tender status (which is the ECB's approach to phasing out the 500-euro note). In this paper, a "tough" scheme is considered where the central bank stops the issuance of a large-denomination banknote and immediately removes its status as legal tender. We refer to this move as the "calling-in" of the large-denomination banknote.

Our goal is to explore optimal timing issues from a central banker's perspective under three assumptions: 1) The net benefits from calling in large notes are uncertain and a function of the (stochastic) natural rate of interest. 2) The calling-in move is irreversible (for instance, because the reputational costs of reversing it are prohibitively high). 3) The move can be timed freely. With these three features, the central banker's problem of finding the optimal timing to make the calling-in move has a structure that is equivalent to the optimal exercise problem of a perpetual American option or to a firm's optimal

[9]Kiley and Roberts (2017) analyze the risk of ELB-episodes for different "steady-state nominal interest rates". Since they assume a 2% inflation target (see ibid., p. 337), the different steady-state nominal interest rates they consider are driven by different "equilibrium real interest rates" (i.e., natural rates of interest). Chung, Gagnon, Nakata, Paustian, Schlusche, Trevino, Vilán, and Zheng (2019) present a related study. They show that the probability that the United States will experience an ELB-episode in the future increases with a decreasing "neutral level of the real federal funds rate in the longer run" (see ibid., pp. 7–8). See also Yates (2004) for a review of earlier studies on the risk of ELB-episodes.

[10]See Rogoff (2017a, pp. 49–51) for a more detailed discussion on the relationship between low real interest rate levels, low monetary policy rates, and the risk of future ELB-episodes.

timing problem for an irreversible investment under uncertainty (a "real option").[11] As pointed out by Dixit and Pindyck (1994, pp. 6–7, 153), a key feature of corresponding decisions with an option element is the "wait-and-see" component of an optimal timing strategy. That is, under certain circumstances, optimality does not require a decision-maker to exercise a real option – analogously to a financial option – until the expected net benefits from making a move are significantly greater than zero. With regard to a central banker's decision problem this implies that even if the expected net benefits from calling in large banknotes today are greater than zero, under certain circumstances, there can be a simple reason to postpone such a calling-in move to a future date. We use a stylized optimal stopping model to show how the optimal timing of calling in large-denomination banknotes depends on the volatility and the expected path of the natural rate of interest. In the idea of concentrating on the option structure of a real policy problem, our paper is most closely related to Alvarez and Dixit (2014) who explore the euro area's "real option" of abandoning the common currency.[12]

Section 2 formalizes a central banker's decision problem and introduces an optimal stopping model of calling in large-denomination banknotes. The model is solved in section 3.1 for the deterministic and in section 3.2 for the stochastic case. Section 4 sheds light on the determinants of optimal policy and the "value" of the central banker's option to make a calling-in move. Section 5 presents some numerical examples to assess a central banker's wait-and-see behavior in different states of the world. Section 6 concludes.

2 An Optimal Stopping Model of Calling In Large-Denomination Banknotes

A central banker with power over the legal tender in a closed economy has the option to make a change in the denominational structure of banknotes. The central banker initially issues banknotes in two denominations, "*small*" and "*large*," and has the authority to call in the large denomination by stopping its issuance and removing its status as legal tender straightaway.[13] The reputational costs of reversing such a calling-in move are prohibitively

[11] See Dixit and Pindyck (1994, pp. 3–25) for a discussion on the analogy between financial and real options and their characterizing features.

[12] In solving our model and in technical regards, we primarily draw on the dynamic programming methods and solution concepts described by Dixit and Pindyck (1994).

[13] See also Rognlie (2016) who uses a model with two cash denominations in the context of the elimination of large denominations to lower the ELB-constraint, too. However, Rognlie (2016) considers the elimination of large denominations in the context of a New Keynesian framework and analyzes household utility under

high, thus the change in the denominational structure is irreversible. The move is one-shot and can be made at any time. Time is continuous and the time horizon is $[0, \infty)$.

Cash demand and banknotes in circulation are not explicitly modeled. We just assume that *small* and *large* banknotes are only imperfect substitutes, such that cash demand depends on the denominational structure and is larger the higher the value of the largest denomination available is (for instance, because there is a specific demand for large notes as a store of value, as Fischer, Köhler, and Seitz (2004) describe). Calling in the large notes reduces cash demand and ultimately cash in circulation. For the central banker in particular, two consequences of such a calling-in move are relevant: The benefit from relaxing the ELB-constraint and a cost in the form of lost seignorage revenues.

Our way of capturing the costs and benefits from calling in large banknotes that will ultimately enter the central banker's objective function is extremely stylized: We follow the approach that Alvarez and Dixit (2014) use to formalize a currency union's decision problem of choosing the optimal timing to break up the union (they consider a potential break-up of the euro area). In principle and to put it simply, Alvarez and Dixit (2014) capture the currency union's costs and benefits from having the common currency by an exogenous process of flow utilities that is independent of the union's timing and of other future variables.[14] Although extremely stylized, this approach is convenient as it allows for a clear analysis of the effects of uncertainty over future states of the world on the decision-makers' actions and wait-and-see behavior. In this manner, we capture the central banker's net benefits from calling in large notes by the flows of utility u_t in period t that are received once the calling-in move has been made. With U_t denoting the central banker's overall period-t utility we can write

$$U_t = \begin{cases} 0 & \text{for } t \in [0, T), \\ u_t & \text{for } t \in [T, \infty), \end{cases} \tag{1}$$

where $T \in [0, \infty)$ denotes the point in time when the calling-in move is made.

Since we want to describe a central banker whose main benefit from calling in large notes is a welfare gain from the relaxation of the ELB-constraint on monetary policy rates

optimal monetary policy dependent on the existence of small and large cash denominations (see Rognlie, 2016, pp. 41–42). In contrast to our paper, Rognlie (2016) does not discuss issues regarding the optimal timing of eliminating large denominations.

[14] Actually, Alvarez and Dixit (2014) start with the modeling of flow benefits a member of the currency union has from belonging to the union, see Alvarez and Dixit (2014, p. 80, equation (3)).

and whose main cost is the loss of seignorage revenues, we can state the central banker's net period utility from calling in large notes as

$$u_t = g - r_t - \omega, \tag{2}$$

where $g \in \mathbb{R}_{>0}$ and $\omega \in \mathbb{R}_{>0}$ are known and constant parameters and r_t denotes the natural rate of interest in period t. The natural rate is governed by an Ornstein-Uhlenbeck process (OU process) with

$$dr_t = \theta(r^{ss} - r_t)dt + \sigma dB_t, \ t \geq 0, \tag{3}$$

where B_t is Brownian motion (with $B_0 = 0$), $\theta \in \mathbb{R}_{>0}$ is the speed of reversion to a long-run mean or steady-state level $r^{ss} \in \mathbb{R}$, and $\sigma \in \mathbb{R}_{\geq 0}$ is a volatility parameter. This specific process is simple enough to allow for an analytical solution of the central banker's decision problem but it is also sophisticated enough to describe a variety of plausible empirical scenarios (we discuss different scenarios in section 3.1).[15] The two constants ω and g are level parameters that capture the costs and benefits from calling in large notes that do not depend on the state of the world, respectively on the natural rate of interest.

Letting the natural rate enter the utility function with a negative sign reflects two assumptions. The first assumption is that the potential benefits from relaxing the ELB-constraint on monetary policy rates by calling in large banknotes are larger the lower the natural rate is. To accept this assumption it helps to consider the economics within a basic New Keynesian model as, for instance, presented in Galí (2015, chapter 3). The deviation of actual output and inflation from their natural and/or efficient levels (i.e., output and inflation gaps) in such a framework increases in the difference between the actual and the natural real rate of interest (i.e., in the real rate gap) with a positive real rate gap being associated with negative inflation and output gaps.[16] In turn, the real rate gap during ELB-episodes is larger the lower the natural rate of interest is during these periods.[17] The reason is that the central bank is unable to lower the policy rate during ELB-episodes to reduce the actual real rate of interest down to a desirable level. All in all, the welfare

[15] Of course, the denominational structure itself and a lowered ELB could in turn influence the structure of the economy and in particular the natural rate, but we shall ignore such and other interdependencies and assume that the natural rate follows an exogenous process.

[16] See Galí (2015, p. 63, equations (22) and (23)) and also Holston, Laubach, and Williams (2017, p. 560) for a short note on this relationship.

[17] See, for instance, Galí (2015, chapter 5.4) for monetary policy under an ELB-constraint.

losses due to the inflation and output gaps during an ELB-episode and thus the benefits from relaxing the ELB-constraint are larger the lower the natural rate is.

The second assumption that is reflected in the negative sign of r_t in the central banker's period utility is that the loss in seignorage revenues in the form of central bank profits from issuing cash is larger the higher the natural rate is. A real world example can illustrate this relationship:[18] The European Central Bank's interest income from banknotes in circulation as stated in its profit and loss account is computed per convention. In principle, it is the interest the ECB earns on its share of euro banknotes in circulation, applying the rule that this share is always 8% and using a rate of return that is just the ECB's main refinancing rate (MRO rate). Consequently, the ECB's interest income from banknotes in circulation is increasing in the MRO rate and in particular it is zero at all when the MRO rate is zero (which has been the case in recent years).[19] So, returning to our model framework, the negative dependence of the central banker's utility from calling in large notes on the natural rate can be thought of as describing a world where, on the one hand, the interest income from banknotes in circulation increases in the policy rate which in turn is an increasing function of the natural rate of interest – and on the other hand, a world where cash demand and thus currency in circulation is smaller when only small denominations are issued.[20]

Let us now consider the central banker's timing problem. We take a $t = 0$-perspective and assume that only the current level of the natural rate $r_0 = r \in \mathbb{R}$ is known such that the decision to call in large notes must be made under uncertainty over the future path of the natural rate and thus under uncertainty over the future net benefits from making a calling-in move. At this point we refer to the real options literature and in particular to Dixit and Pindyck (1994, pp. 3–25) who point out the analogy between financial options and real options, i.e., opportunities to make real investments that can be timed freely, that are irreversible, and that are made under uncertainty over future states of the world.

[18] We are grateful to Franz Seitz who pointed out this example after a seminar talk in Leipzig. See also Krüger and Seitz (2017, chapter 4.5).

[19] See European Central Bank (2019, p. A4) for these accounting rules. The ECB's share is 8% irrespective of its true value so that 92% of the euro banknote issuance are allocated to euro area national central banks that actually issue banknotes. See also European Central Bank (2019, p. A24) for the position "interest income arising from the allocation of euro banknotes within the Eurosystem" which was zero in 2017 and 2018. In contrast, for example, with the relatively high interest rates (compared to current levels) that prevailed ten years earlier, in 2008, this position amounted to over 2.2 billion euros (see European Central Bank, 2009, p. 218).

[20] See, for instance, Rogoff (2016, chapter 6) for this last point in the context of a central bank's seignorage revenues.

In principle, our central banker's calling-in move can be regarded as an investment that shares the three exact same characteristic features: the move is irreversible, can be timed freely, and is made under uncertainty. So, since the central banker's timing problem has the same structure as the problem of pricing an American option or as a firm's problem of finding the optimal timing to make an investment, we use dynamic programming as described in Dixit and Pindyck (1994, pp. 93–132, 135–174) to solve this problem. In doing so, we follow Alvarez and Dixit (2014) who apply dynamic programming to value a currency union's "real option" to break up the union.

Accordingly, we solve the central banker's problem of when to optimally call in large notes by finding the "value" $V(r)$ of her option to make this calling-in move depending on the period-0 natural rate $r_0 = r$. We use the term "calling-in option" from now on and measure "value" in terms of utility such that the value of the calling-in option is the expected present value of the stream of flow utilities from calling-in large notes provided that the calling-in move will be timed optimally.[21] Determining the value function yields an optimal timing strategy for the "exercise" of the calling-in option in the form of the rule to call in large notes as soon as the natural rate hits or falls below a certain threshold $\underline{r}$. We define the value of the calling-in option, given that the central banker follows this rule, and given that the period-0 level of the natural rate is $r \geq \underline{r}$, as supremum of the expected present value of the stream of flow utilities the central banker receives after having made the calling-in move at time T. With $\delta \in \mathbb{R}_{>0}$ denoting the rate at which the central banker discounts future utility, the value of the calling-in option is thus

$$V(r) = \sup_{T \geq 0} \mathbb{E}\left[\int_T^\infty (g - r_t - \omega) \cdot \exp(-\delta t) dt \;\middle|\; r_0 = r\right], \tag{4}$$

where the supremum is taken over all timings $T \geq 0$ to make the calling-in move. Note that each timing is a stopping time, i.e., a random variable that is defined by a timing strategy in the form of a rule to make the move once the natural rate hits or falls below a certain threshold. Finding the optimal threshold $\underline{r}$ and the value V of the calling-in option is an optimal stopping problem. We solve this problem in the next section.

[21]See Dixit and Pindyck (1994, pp. 99–101) for a discussion of the basic role of value functions in a dynamic programming context.

3 Optimal Policy and Option Value

3.1 Optimal Timing and Option Value under Perfect Foresight

Before we solve the model for the case where $\sigma > 0$ in section 3.2, we consider a world with perfect foresight and thus without uncertainty over the future path of the natural rate of interest. So, for the remainder of this section, we assume $\sigma = 0$. In solving a deterministic version of the model first, we choose the same order of analysis as Dixit and Pindyck (1994, pp. 136–147) (for a generic timing problem) in order to provide some intuition on how the results of the model are driven by the non-stochastic variables, and in particular, by the anticipated path of the natural rate. For that purpose, we analyze the central banker's timing strategy in six scenarios with different paths of the natural rate.

The central banker's timing problem under perfect foresight is relatively simple and the solution is obtained as follows: Consider first the path of the natural rate. In general, with the natural rate behaving as described by equation (3), r_t conditional on $r_0 = r \in \mathbb{R}$ is Gaussian with $\mathbb{E}[r_t] = r \cdot \exp(-\theta t) + r^{ss} \cdot (1 - \exp(-\theta t))$ and $\mathrm{Var}(r_t) = \frac{\sigma^2}{2\theta} \cdot (1 - \exp(-2\theta t))$.[22] So, the expected natural rate is a monotone function of time. For $\sigma = 0$ the actual natural rate in period t will be equal to the expected natural rate in period t where, with $r_0 = r$,

$$r_t = r \cdot \exp(-\theta t) + r^{ss} \cdot (1 - \exp(-\theta t)) \,. \tag{5}$$

Therewith, we can compute the expected (period-0) present value of the stream of flow utilities $u_t = g - r_t - \omega$ the central banker will receive after having called in the large note, given that $r_0 = r$ and given that the calling-in move will be made at time $T \geq 0$. For $\sigma = 0$, this expected value equals the actually realized value $F(r, T)$ with[23]

$$\begin{aligned} F(r,T) &= \int_T^{\infty} (g - r_t - \omega) \cdot \exp(-\delta t) dt \qquad (6) \\ &= \frac{1}{\delta} \cdot (g - r^{ss} - \omega) \cdot \exp(-\delta T) \;-\; \frac{1}{\delta + \theta} \cdot (r - r^{ss}) \cdot \exp(-(\delta + \theta)T). \end{aligned}$$

This is the central banker's objective function under perfect foresight. She maximizes (6) simply by choosing an optimal point in time $T^* \in [0, \infty) \cup \{\infty\}$ to make the calling-in move

[22]See, for instance, Maller, Müller, and Szimayer (2009, p. 423) and Dixit and Pindyck (1994, pp. 74–78).

[23]Here, a great technical advantage of the OU process for our purposes becomes apparent: the integral in (6) is easy to solve with r_t as defined in (5).

($T^* = \infty$ means that the move will never be made). So, the optimal stopping problem (4) degenerates to

$$V(r) = \sup_{T \geq 0} F(r, T), \tag{7}$$

where $T \in [0, \infty)$ is a deterministic variable and $T^* \in [0, \infty) \cup \{\infty\}$ is thus already known in period $t = 0$ (in the general case for $\sigma > 0$, T is a random variable). Note, that as the central banker can simply choose never to exercise the calling-in option at all, its value V in terms of future utility must be bounded from below by zero.[24]

The optimal timing strategy to make the move in period T^* can also be formulated as a decision rule: do not make the calling-in move as long as the natural rate is above a certain threshold $\underline{r}$ and make the move as soon as the natural rate hits or has fallen below this optimal threshold. In the absence of uncertainty over the natural rate, the optimal threshold $\underline{r}$ can easily be computed by evaluating (5) at $t = T^*$ to obtain $\underline{r} = r_{T^*}$.

Therewith, the value V of the calling-in option can also be expressed as the present value of the "*exercise payoff*" $\mathcal{V}$ with

$$V(r) = \mathcal{V}(\underline{r}) \cdot \exp(-\delta T^*), \tag{8}$$

where the exercise payoff $\mathcal{V}$ at a given natural rate r is defined as the expected present value of the stream of flow utilities from calling in the large note given that the calling-in move is made at this given natural rate r with[25]

$$\mathcal{V}(r) := \mathbb{E}[F(r, T = 0)] = \mathbb{E}\left[\int_0^\infty (g - r_t - \omega) \cdot \exp(-\delta t) dt \mid r_0 = r\right] \tag{9}$$

$$= \frac{1}{\delta} \cdot (g - r^{ss} - \omega) - \frac{1}{\delta + \theta} \cdot (r - r^{ss}). \tag{10}$$

It is the specific path of the natural rate that determines whether (7) has an interior or a corner solution. In the following, we solve the model for different scenarios and show how T^* and $\underline{r}$ depend on the anticipated path of the natural rate. Recall, we have assumed that the natural rate is governed by a mean-reverting process. In the absence of stochastic

[24] Note, that the value of the calling-in option cannot simply be stated as maximum of F over T since such a maximum does not necessarily exist (think of F converging to zero from below for $T \to \infty$). So, we use the supremum in the formulation of the deterministic timing problem, too.

[25] We use the term "exercise payoff" but mean the same concept that Dixit and Pindyck (1994, p. 99) define as "termination payoff." For analogous timing problems see, for instance, the generic problems in Dixit and Pindyck (1994, chapter 5) or in Chevalier-Roignant and Trigeorgis (2011, chapter 9).

movements (if $\sigma = 0$), the path of the natural rate is a deterministic and monotone function of time, and whether this function is decreasing or increasing in time depends on whether the natural rate reverts to its steady state level r^{ss} from above or from below. Therewith, the central banker's period utility from calling in large notes $u_t = g - r_t - \omega$ (as a linear function of the natural rate) features mean-reversion as well with the path of utility being inversely related to the path of the natural rate (the two constants g and ω are just level parameters). The long-run steady state level of period utility is thus just $u^{ss} = g - r^{ss} - \omega$. So, the path of period utility is monotonically increasing in time if the natural rate reverts to its steady state from above ($r_0 > r^{ss}$) and monotonically decreasing if the natural rate reverts to its steady state from below ($r_0 < r^{ss}$). In the context of the central banker's decision problem, these two cases describe states of the world where the ELB-constraint becomes more, respectively less, relevant as time evolves.

Now, for the following analysis, we define two regions, $\mathcal{A}$ and $\mathcal{B}$, of the parameter space where $\mathcal{A} : u^{ss} > 0$ and $\mathcal{B} : u^{ss} \leq 0$. A positive steady state of utility means that in the long run the benefits from calling in large-denomination banknotes will be greater than the costs, a negative steady state means that the costs will exceed the benefits in the long run. In each of the two regions, we consider three scenarios I, II, and III with an initially *"high,"* an initially *"near-steady-state,"* and an initially *"low"* natural rate reverting to its steady state, respectively.

Scenario $\mathcal{A}.I$ ($r_0 = r > g - \omega > r^{ss}$) This scenario describes a world where in the long run, the benefits from calling in large notes are greater than the costs. With our interpretation of the costs and benefits this means that, in the long run, the central banker's benefits from relaxing the ELB-constraint exceed the costs in the form of lost seignorage revenues. But with a relatively high natural rate in period $t = 0$, this scenario also describes a world where ELB-issues are at first irrelevant and only gain importance as time evolves and the natural rate decreases to a relatively low steady state which involves a relatively high "risk" of policy rates hitting the ELB. Formally, this is reflected in the period-0 flow utility $u_0 = g - r_0 - \omega$ which is negative for $r_0 > g - \omega$ but increases as time evolves. What we want to show is that the central banker will wait to make the calling-in move until the natural rate has fallen to a sufficiently low level – although moving earlier would already yield a positive exercise payoff.

Now, if the steady state of utility u^{ss} is strictly positive, we can derive the optimal timing T^* of the calling-in move as well as the level $\underline{r}$ of the natural rate at which the option is optimally exercised simply by maximizing F over $T \in [0, \infty)$. In this case,

$$V(r) = \sup_{T \geq 0} F(r, T) = \max_{T \geq 0} F(r, T). \tag{11}$$

The first-order condition for an interior maximum is implied by

$$\frac{\partial F(r,T)}{\partial T} = -(g - r^{ss} - \omega) \cdot \exp(-\delta T) + (r - r^{ss}) \cdot \exp(-(\delta + \theta)T), \tag{12}$$

and reads

$$(g - r^{ss} - \omega) \cdot \exp(-\delta T) = (r - r^{ss}) \cdot \exp(-(\delta + \theta)T). \tag{13}$$

This yields the unique interior solution[26]

$$T^* = \frac{1}{\theta} \cdot \ln \left(\frac{r - r^{ss}}{g - r^{ss} - \omega} \right) \tag{14}$$

which implies

$$\underline{r} = r_{t=T^*} = g - \omega. \tag{15}$$

This $g - \omega$ threshold is critical: If the natural rate is below the threshold $g - \omega$ it is so low that the central banker's period utility u_t is positive (with $\underline{r} = g - \omega$ it is clear that $u = g - r' - \omega > 0$ for all $r' < \underline{r}$). With our interpretation of the central banker's costs and benefits we can reformulate this statement: A natural rate below the threshold $g - \omega$ is so low that ELB-issues outweigh seignorage losses. This justifies the decision rule that is implied by $\underline{r}$, which is to make the calling-in move as soon as the period utility u_t becomes non-negative (recall that without stochastic fluctuations, the period utility u_t will steadily approach its steady state u^{ss} which in this scenario is positive, so once it has become non-negative, the period utility will stay positive forever, given that $\sigma = 0$).

If the central banker times the calling-in move optimally, respectively follows the decision rule to make the move once the natural rate hits $\underline{r}$, she will receive only positive

[26] With $\frac{\partial^2 F(r,T)}{\partial T^2} = \delta(g - r^{ss} - \omega) \cdot \exp(-\delta T) - (\delta + \theta) \cdot (r - r^{ss}) \cdot \exp(-(\delta + \theta)T)$ it easily checked that $\partial^2 F(r, T^*)/\partial T^2 < 0$ for $r > g - \omega > r^{ss}$ and that the global maximum of F on $[0, \infty)$ is in fact in $T^* > 0$.

flows of utility. Consequently, the present value of these flows, i.e., the value V of the calling-in option, will be strictly positive.[27] So, the central banker could still receive a positive exercise payoff if she moved "somewhat" earlier before the natural rate hits $\underline{r}$.

All in all, this scenario describes a situation where deferring the calling-in move is rational although the payoff from making the move "somewhat" earlier would already be greater than zero. The reason for this deferral is the anticipated decline of an initially high natural rate to a relatively low steady state in the future – which describes a world where ELB-issues are initially irrelevant but are only gaining importance over time. As making the calling-in move at a relatively high natural rate level initially would lead to negative period utilities, waiting until the period utility becomes greater than zero increases the central banker's overall payoff.

The central banker's tendency to defer the exercise of the calling-in option, that is, to wait due to the anticipated reversion of the natural rate to its steady state is reflected in the length of the interval between the optimal threshold $\underline{r}$ and the "*break-even threshold*" $\hat{r}$ defined as the natural rate below which the central banker would make the calling-in move in a situation where she would have to decide between making the move now or never. The break-even threshold is implicitly defined by $F(r_0 = \hat{r}, T = 0) = 0$ which yields

$$\hat{r} = g - \omega + \frac{\theta}{\delta} \cdot (g - r^{ss} - \omega) = g - \omega + \frac{\theta}{\delta} u^{ss}. \tag{16}$$

The representation of the break-even threshold $\hat{r}$ in (16) illustrates that the larger the benefits from calling in large notes are in the long run, i.e., the larger u^{ss} is, the earlier the central banker could make the move without incurring a negative exercise payoff.

Thus, if $r_0 = r \in (\underline{r}, \hat{r})$, the period-0-value of the payoff from exercising the calling-in option at $T = 0$ is already strictly greater than zero, but optimality requires the central bank to defer the calling-in move to T^* until also the period flow utility u_t exceeds zero (recall that $u_{t=T^*} = g - r_{t=T^*} - \omega = 0$, with $r_{t=T^*} = \underline{r} = g - \omega$). So, the move is deferred to avoid negative streams of period utilities.

Also the next two scenarios describe a world where the long-run benefits from calling in large notes are greater than the costs ($u^{ss} > 0$). But with $r_0 < g - \omega$ in both scenarios,

[27] This can easily be checked by considering $V(r) = F(r, T^*) = \mathcal{V}(\underline{r}) \cdot \exp(-\delta T^*) = \left(\frac{1}{\delta}(g - r^{ss} - \omega) - \frac{1}{\delta+\theta}(g - r^{ss} - \omega)\right) \cdot \exp(-\delta T^*) = \frac{\theta}{\delta^2+\delta\theta} \cdot u^{ss} \cdot \exp(-\delta T^*)$ which is strictly greater than zero if $u^{ss} > 0$.

the natural rate is and will remain so low that the short-run benefits from calling in large notes are also so large that optimality requires the central banker to make the calling-in move without delay in period $t = 0$. So, the next two scenarios describe an economic environment where ELB-issues will be relevant from the outset and forever. Trivially, the decision rule to make the move as soon as the natural rate hits or has fallen below the threshold $g - \omega$ also applies in the next two scenarios.

Scenario $\mathcal{A}.II$ $(g - \omega \geq r_0 = r > r^{ss})$ The optimal timing of the calling-in move in this scenario is $T^* = 0$. This is a corner solution of $\max_{T \geq 0} F(r, T)$ with $F(r, T = 0) > 0$ and ${}^{\partial F(r,T)}/_{\partial T} < 0 \; \forall \; T \in (0, \infty)$. The reason that the move should be made in period $t = 0$ is that period utility u_t is positive from the outset (since $r_0 < g - \omega$).

Scenario $\mathcal{A}.III$ $(g - \omega > r^{ss} \geq r_0 = r)$ As in scenario $\mathcal{A}.II$, the optimal timing of the calling-in move is $T^* = 0$. Again, this is a corner solution of $\max_{T \geq 0} F(r, T)$ with $F(r, T = 0) > 0$ and ${}^{\partial F(r,T)}/_{\partial T} < 0 \; \forall \; T \in [0, \infty)$. The difference to the previous two scenarios is that the natural rate reverts to its steady state from below. One could interpret this scenario as describing a world during or after a financial or an economic crisis where a relatively low natural rate produces a severe and pronounced ELB-episode. As time evolves, this severe episode will find an end, but the relatively low steady state of the natural rate implies that ELB-issues will remain relevant forever.

Scenario $\mathcal{B}.I$ $(r_0 = r > r^{ss} > g - \omega)$ With a high natural rate r_0 and a relatively high steady-state level r^{ss}, this scenario describes a world in which the ELB-constraint is and will be of no importance such that relaxing it would be useless. Neither the short-run benefits nor the long-run benefits from making the calling-in move will exceed the costs. With $r_0 = r > r^{ss} > g - \omega$, the period flow utility u_t from calling-in large notes will never be positive so that $F(r, T) < 0 \; \forall \; T \geq 0$ and thus $V(r) = 0$. The calling-in option will never be exercised.

Scenario $\mathcal{B}.II$ $(r^{ss} \geq r_0 = r > g - \omega)$ As in scenario $\mathcal{B}.I$, the natural rate will always remain so high that the period utility u_t from calling in large notes will always be negative. Hence, $F(r, T) < 0 \; \forall \; T \geq 0$ and $V(r) = 0$. The calling-in move will never be made.

Scenario $\mathcal{B}.III$ ($r^{ss} > g - \omega \geq r_0 = r > -\infty$) This scenario describes a world where the benefits from making the calling-in move exceed the costs in the short run ($r_0 < g - \omega$ such that $u_0 > 0$) but not in the long run ($u^{ss} < 0$). For instance, large-scale financial or economic crises could feature such exceptionally low natural rates. With monetary policy rates that have reached the ELB in such a scenario, calling in large notes could be a rational move even if this decision entailed long-run losses (as long as future losses are discounted – which we assume by setting $\delta > 0$). The condition for making the move is that the short-run benefits are sufficiently large, which is only the case for exceptionally low natural rates such that the short-run period utility from calling in large notes is significantly greater than zero. The condition for a significantly positive utility u_t is that the natural rate is significantly below the threshold $g - \omega$. We can state this more precisely by considering the break-even level $\hat{r}$ again as defined in (16) as $\hat{r} = g - \omega + \frac{\theta}{\delta} u^{ss}$. Recalling that $u^{ss} < 0$ in this scenario, it is clear that the break-even threshold $\hat{r}$ is lower, the smaller the steady state u^{ss} is. Since the natural rate increases as time evolves, optimality requires the central banker to make the move without delay in period $t = 0$ if $r_0 \leq \hat{r}$. Thus, the rule of whether or when to make the calling-in move in a world where this move implies long-run losses ($u^{ss} < 0$) is implied by the optimal threshold $\underline{r} = \hat{r}$: move immediately if $r_0 \leq \hat{r}$, and do not move if $r_0 > \hat{r}$.[28]

3.2 Optimal Exercise Rule and Option Value under Uncertainty

We now solve the central banker's optimal stopping problem (4) in a world without perfect foresight where the path of the natural rate of interest is uncertain, i.e., where $\sigma > 0$. The solution approach we use is taken from Dixit and Pindyck (1994) which is our main reference in technical regards (as far as possible, we use the [shorthand] notation proposed therein).[29] In addition, we refer to Øksendal (2013) for some basic methods of Itô calculus

[28]We assume implicitly that the calling-in option is also exercised if $r_0 = \underline{r} = \hat{r}$. Moreover, we have not considered the case $r^{ss} = g - \omega$ so far which implies $u^{ss} = 0$. With (16) it becomes clear that in this case $\hat{r} = g - \omega = r^{ss}$ so that the option is exercised never if $r_0 = r > r^{ss}$ and exercised at $t = 0$ if $r_0 = r < r^{ss}$. For the case $r_0 = r = r^{ss} = g - \omega$ implying $F(r, T) = 0 \; \forall \; T \in [0, \infty)$ we break ties in favor of the calling-in option being exercised at $t = 0$.

[29]For instance, Dixit and Pindyck (1994) describe how to solve the optimal stopping problem of a firm that has the option to make a real investment under uncertainty over the future value of that investment. Inter alia, they show how to use dynamic programming to solve the firm's problem when the evolution of the project value is described by geometric Brownian motion (ibid., pp. 140-147) or by a mean-reverting process that has an absorbing state (ibid., pp. 161-167) (note, the process defined by (3) that we use to describe the path of the natural rate has no absorbing state).

that are used here and Lebedev (1965) for the differential equation and the solution we obtain.

So, following Dixit and Pindyck (1994) in this technical regard, we use dynamic programming based on Bellman's principle of optimality to solve the optimal stopping problem (4).[30] Accordingly, the value function $V(r)$ and the optimal threshold $\underline{r}$ can be obtained by solving the Bellman equation

$$\delta V dt = \mathbb{E}[dV] \tag{17}$$

for V where (17) holds for all levels of the natural rate r that are so high that optimality requires the central banker to keep on issuing large banknotes – i.e., for all $r \in [\underline{r}, \infty)$.[31]

Finding V and $\underline{r}$ is a free boundary problem.[32] We solve this problem by using the solution approach described by Dixit and Pindyck (1994, pp. 95–114, 130–132) and in particular by ibid. (pp. 140–147). Accordingly, we start by using the Itô formula to write the Bellman equation (17) as the homogeneous ordinary differential equation

$$\frac{1}{2}\sigma^2 V'' + \theta(r^{ss} - r)V' - \delta V = 0 \tag{18}$$

(see appendix A for the detailed derivation).[33]

By introducing economically meaningful boundary conditions we can solve equation (18) for V and obtain $\underline{r}$.[34] We assume that the solution of (18) must satisfy two left boundary conditions, a monotonicity condition, and a non-negativity constraint. The two left boundary conditions we apply are standard in the literature where they are often referred to as "value-matching condition" and "smooth-pasting condition": Referring to Dixit and Pindyck (1994, pp. 109, 130–132, 141) and Peskir and Shiryaev (2006, chapters 8,

[30] See Dixit and Pindyck (1994, p. 100) for a discussion of Bellman's principle of optimality in the context of the valuation of investment projects.

[31] For a discussion of related Bellman equations that solve optimal stopping problems in continuous time with an infinite time horizon see Dixit and Pindyck (1994, pp. 101-114) and in particular ibid. (p. 140, equation (7)) where a Bellman equation is discussed that is equivalent to the Bellman equation we have. For a discussion of why a value function in an infinite-time-horizon setting does not explicitly depend on time see, for instance, Dixit and Pindyck (1994, p. 107). For a Bellman equation in the context of a monetary union's problem of when to optimally break up the union see Alvarez and Dixit (2014, p. 81, equation (9)).

[32] See Dixit and Pindyck (1994, p. 109).

[33] Although the context of these papers is far away from our subject, we refer to Parlour and Walden (2009, p. 13, equation (14)), Garlappi and Yan (2011, p. 819, equation (A2)), and Suzuki (2016, p. 39, equation (42)) who obtain equivalent/similar differential equations in their respective valuation problems with state variables that also follow an Ornstein-Uhlenbeck process, respectively.

[34] See Dixit and Pindyck (1994, p. 109) for a short note on the rationales of respective boundary conditions in optimal stopping problems in an economic context.

9) for a further discussion of the concepts, respectively the rationales, of these conditions, we use the value-matching condition

$$V(\underline{r}) \stackrel{!}{=} \mathcal{V}(\underline{r}) \tag{19}$$

that requires the value of the calling-in option V to equal the option's exercise payoff $\mathcal{V}$ in the moment the option is exercised, i.e., when $r = \underline{r}$, and the smooth-pasting condition

$$V'(\underline{r}) \stackrel{!}{=} \frac{\partial \mathcal{V}(\underline{r})}{\partial \underline{r}} \tag{20}$$

that requires the value function V at $\underline{r}$ to have the same slope as the exercise payoff function $\mathcal{V}$ at $\underline{r}$.[35]

In addition to these two left boundary conditions, we introduce a monotonicity condition arguing that this is a natural assumption with respect to the calling-in option valued at a higher natural rate of interest: We require that

$$V'(r) < 0 \; \forall \; r \;\in\; (\underline{r}, \infty) \tag{21}$$

and thus capture the intuition that the expected present value of the exercise payoff (and therewith the value of the calling-in option) should be smaller the longer it will presumably take until the natural rate hits the optimal exercise threshold $\underline{r}$.[36]

The value function must also satisfy the non-negativity constraint

$$V \geq 0 \tag{22}$$

which, trivially, just reflects that the central banker has the option to issue large notes forever. Now, it is straightforward to use the monotonicity condition together with the non-negativity constraint and the two left boundary conditions to obtain a particular solution of (18). We have placed the single steps in the appendix and summarize the results in the next proposition:

[35] For an application of the value-matching and smooth-pasting condition in the context of a currency union's optimal stopping problem of when to break up the union see Alvarez and Dixit (2014, p. 81).

[36] To see this, consider equation (8) stating that $V(r) = \mathcal{V}(\underline{r}) \cdot \exp(-\delta T^*)$ for the case of $\sigma = 0$ and recall that T^* increases in r.

Proposition 1. *A particular solution of the Bellman equation* (18) *that solves the central banker's optimal stopping problem* (4) *subject to* (19), (20), (21), *and* (22) *is given by*

$$V(r) = c_1 \cdot H_{-\frac{\delta}{\theta}}\left(\frac{\sqrt{\theta}}{\sigma} \cdot (r - r^{ss})\right), \tag{23}$$

with

$$c_1 \;=\; \frac{1}{\delta+\theta} \cdot \frac{\sqrt{\theta}\cdot\sigma}{2\delta} \cdot \frac{1}{H_{-1-\frac{\delta}{\theta}}\left(\frac{\sqrt{\theta}}{\sigma}\cdot(\underline{r} - r^{ss})\right)}, \tag{24}$$

and with $\underline{r}$ *being implicitly defined by*

$$\frac{1}{\delta+\theta} \cdot \frac{\sqrt{\theta}\cdot\sigma}{2\delta} \cdot \frac{H_{-\frac{\delta}{\theta}}\left(\frac{\sqrt{\theta}}{\sigma}\cdot(\underline{r} - r^{ss})\right)}{H_{-1-\frac{\delta}{\theta}}\left(\frac{\sqrt{\theta}}{\sigma}\cdot(\underline{r} - r^{ss})\right)} = \frac{1}{\delta}(g - r^{ss} - \omega) - \frac{1}{\delta+\theta}(\underline{r} - r^{ss}), \tag{25}$$

where $H_\nu(z)$ *denotes a Hermite function (as defined, for instance, in Lebedev, 1965, p. 285). Thereby, the value function is defined piecewise: The value function* $V(r)$ *is given by* (23) *for all* $r \in [\underline{r}, \infty)$. *For all* $r < \underline{r}$, *immediate exercise of the calling-in option is optimal such that the value function for all* $r < \underline{r}$ *is just defined as* $V(r) = \mathcal{V}(r)$.

Proof. See appendix (note that since our main intention is to show that the central banker's problem of finding the optimal timing of calling in large banknotes has a structure that is equivalent to the structure of an option valuation problem, we just prove the existence and not the uniqueness of our solution). □

In the next section, we discuss the determinants of the optimal exercise threshold and the value of the calling-in option in detail. In section 5, in order to obtain illustrative results, we solve equation (25) numerically for specific parameter values and use the resulting approximations of $\underline{r}$ to compute and analyze the value V of the calling-in option in different scenarios.

4 Determinants of Optimal Policy and Option Value

4.1 Option Value and Measures of the Central Banker's Tendency to Wait and See

In the following, we first make some preliminary remarks on the value of the calling-in option and on different measures of the central banker's tendency to wait and see. Subsequently, in sections 4.2 to 4.5, we discuss the determinants of the break-even threshold $\hat{r}$, the optimal exercise threshold $\underline{r}$, and the option value.

The value of the calling-in option consists of two components. Keeping the terminology commonly used for financial options, we refer to these two components as "intrinsic" and "time value". The intrinsic value (IV) is the positive part of the payoff from immediately making the calling-in move, that is, $IV(r) = \max\{\mathcal{V}(r), 0\}$. The time value ($TV$), is defined as $TV(r) = V(r) - IV(r)$.[37] The time value is a measure of the central banker's tendency to wait and see instead of calling in large banknotes at the first opportunity where a non-negative exercise payoff could be realized.

Two factors can add time value to the calling-in option: The first one is related to the expected path of the natural rate. If the natural rate is expected to decrease as time evolves, the central banker will expect the benefits from calling in large notes to increase over time. If, additionally, the long-run benefits from calling in large notes are positive, i.e., if $u^{ss} > 0$, there can be a reason to delay the calling-in move until the short-run benefits from removing large notes are sufficiently large – even if the immediate exercise of the calling-in option yields a positive payoff (as illustrated in scenario $\mathcal{A}.I$ in section 3.1). The second factor that adds time value to the calling-in option, of course, is uncertainty over the future path of the natural rate. This uncertainty is reflected in $\mathrm{Var}(r_t)$ for $t > 0$ and thus depends on the volatility parameter σ and on the speed of mean-reversion θ (recall, as noted in section 3.1, that $\mathrm{Var}(r_t) = \frac{\sigma^2}{2\theta} \cdot (1 - \exp(-2\theta t))$).

The time value reflects the central banker's tendency to wait and see and thus crucially determines the optimal timing of the calling-in move. Below, we evaluate this tendency to wait and see in different scenarios in terms of the option's time value and with the following other measures:[38] The tendency to wait and see due to an expected decline of the natural rate and thus due to increasing expected benefits from calling in large

[37] For a short definition of these concepts see, for instance, Hull (2019, p. 284).
[38] See Alvarez and Dixit (2014, pp. 85–86) for a discussion of different measures of option value.

notes is reflected in the difference between the break-even threshold $\hat{r}$ and the optimal threshold under perfect foresight $\underline{r}|_{\sigma=0}$. The tendency to wait due to uncertainty over the future path of the natural rate is reflected in the difference between the optimal threshold under perfect foresight and the actually optimal threshold $\underline{r}$. And obviously, the "overall" tendency to wait and see is reflected in the length of the interval $[\underline{r}, \hat{r}]$. So, in the following, we analyze the variables (in that order) $\hat{r}$, $\underline{r}|_{\sigma=0}$, $\underline{r}$, and TV with respect to their dependency on σ, θ, and u^{ss}. Note, that we use the variance of r_t (conditional on $r_0 = r$) as a measure of uncertainty over the natural rate in period t and use the terms "long-term uncertainty" in this respect for "large" t and "short-term uncertainty" for "small" t. With this classification, uncertainty (over both short and long horizons) is increasing in the volatility parameter σ and decreasing the speed of mean-reversion θ.[39]

4.2 Determinants of the Break-Even Threshold

Uncertainty The way in which we specify our model implies that the break-even threshold $\hat{r} = g - \omega + \frac{\theta}{\delta} \cdot (g - r^{ss} - \omega) = g - \omega + \frac{\theta}{\delta} u^{ss}$ (as stated in (16)) does not depend on the volatility parameter σ and is thus independent of uncertainty over the future path of the natural rate. The advantage of this specification is that the effect of σ on the optimal exercise threshold $\underline{r}$ is – as intuition suggests – monotonic (we discuss this effect in section 4.4).[40] To see that $\hat{r}$ is independent of σ, recall that $\hat{r}$ is determined by the expected path of period utility which, in turn, is a linear function of the expected path of the natural rate. There is no link between $\hat{r}$ and σ because the two expected paths are independent of the volatility parameter σ.

Steady-State Utility In a world where the central banker would lack the flexibility to time the calling-in move freely being in a situation where she would have to choose between making the move immediately or issuing large notes forever, a relatively high steady state of utility (i.e., the prospect of large eternal long-run benefits from calling in large notes) could make her willing to call in large notes even if she had to accept losses in the short run. Thus, under special circumstances, a central banker who would have to make the decision

[39] While it is obvious that $\frac{\partial \mathrm{Var}(r_t)}{\partial \sigma} > 0$, it is not immediately clear why the second statement that $\frac{\partial \mathrm{Var}(r_t)}{\partial \theta} < 0$ holds. To see why it holds, consider $\frac{\partial \mathrm{Var}(r_t)}{\partial \theta} = \frac{\sigma^2}{2\theta^2} \cdot (\exp(-2t\theta) \cdot (2t\theta + 1) - 1)$. To show that this expression is less than zero, it is sufficient to show that $\exp(-2t\theta) \cdot (2t\theta + 1) < 1$. Taking the $ln(\cdot)$, rearranging and using the substitution $x = 2t\theta$ the last inequality can be transformed to $ln(x + 1) < x$ which, in turn, can easily be proved.

[40] For instance, a non-monotonic effect of volatility on a respective optimal threshold is described in Alvarez and Dixit (2014).

whether to call in large notes now or never while expecting large long-run benefits from calling in large notes would be willing to make the move already and even at a relatively high natural rate which would involve short-run losses. Thus, the break-even threshold $\hat{r}$ (being the relevant threshold for a central banker in a hypothetical now-or-never situation) is higher the higher the steady-state utility is. Formally, this relationship is captured by (16) which illustrates that a positive steady-state utility induces a markup over the level $g - \omega$ which is the threshold below which natural rates result in a positive period utility.[41]

Speed of Mean-Reversion In the following, we consider a world where the central banker lacks the flexibility to time the calling-in move freely. Concretely, we assume that the central banker can choose between making the calling-in move immediately or issuing large notes forever. We consider two scenarios in this "now-or-never" world to illustrate that a higher speed of mean-reversion either increases or decreases $\hat{r}$, dependent on whether utility has a positive or a negative steady state (which, formally, is immediately clear with $\hat{r} = (1 + \frac{\theta}{\delta}) \cdot u^{ss} + r^{ss}$). Since $\hat{r}$ does not depend on σ (see above), we let, without loss of generality, $\sigma = 0$ in the following two scenarios and assume perfect foresight with respect to the future path of the natural rate of interest. The scenarios differ in the steady state of utility with the first scenario characterized by $u^{ss} < 0$ and the second by $u^{ss} > 0$. The present in both scenarios is $t = 0$ with the natural rate at this point in time just having hit the break-even level, i.e., $r_0 = r = \hat{r}$.

So, consider the first scenario where period utility has a negative steady state while $r_0 = r = \hat{r}$ which implies that the natural rate reverts to its steady state from below. With respect to the anticipated path of period utility, the future from a $t = 0$ perspective can be partitioned into two consecutive phases: the first phase is one of relatively low natural rates (since $\hat{r} < r^{ss}$) and thus a phase where the calling-in of large banknotes is beneficial to the central banker (period utility is positive during this phase because the ELB-constraint is relevant). The second phase features relatively high natural rates such that (due to the irrelevance of the ELB-constraint) having stopped issuing large notes now involves losses for the central banker (due to forgone seignorage revenues), i.e., period utility is negative during the second phase. It is clear that making the calling-in move is only optimal if the short-run benefits from calling in large notes outweigh the long-run

[41] The break-even threshold can also be written as a weighted sum of steady-state utility and the steady-state natural rate which acts as a level parameter such that $\hat{r} = (1 + \frac{\theta}{\delta}) \cdot u^{ss} + r^{ss}$. This makes clear that the exercise of the calling-in option in a now-or-never situation in a $u^{ss} < 0$-scenario can only be optimal at natural rates that are strictly below their steady-state level.

losses. That is, the move is only optimal if the first phase with positive period utilities is sufficiently long and/or if the period utilities during this phase are sufficiently high. However, the first phase is shorter the higher the speed of mean-reversion θ is. So, the higher θ, the higher are the short-run benefits the central banker requires to make the calling-in move and thus the lower the break-even threshold $\hat{r}$ (see the red lines in figure 1(a) for an illustration).

Consider now the second scenario with period utility having a positive steady state. Again, the time after the calling-in move is made can also be partitioned into two consecutive phases, a phase of negative period utilities followed by a phase of positive period utilities (recall that $\hat{r} > g - \omega > r^{ss}$ implying that the natural rate reverts to its steady state from above in that case). Now, with the prospect of eternal long-run benefits from calling in large notes (that will be received once the natural rate has fallen below the level $g - \omega$, i.e., when ELB-issues outweigh seignorage losses), the central banker will accept making short-run losses, that is, she will accept negative period utilities during the first phase of relatively high natural rates. The cumulative short-run losses are smaller the shorter the first phase is, that is, the higher the speed of mean-reversion is. Conversely, the shorter the first phase is, the higher the period losses the central banker accepts during the first phase. Thus, the higher θ, the higher the break-even threshold $\hat{r}$ (see the red lines in figure 1(b) for an illustration).

4.3 Determinants of the Optimal Threshold under Perfect Foresight

In this section, we analyze the optimal threshold $\underline{r}$ in the absence of uncertainty over the natural rate in order to focus on the non-stochastic determinants of the central banker's optimal timing and a potential tendency to wait and see. So, let $\sigma = 0$ in the following.

Steady-State Utility First, we discuss the relationship between the optimal threshold given $\sigma = 0$ and the steady-state utility and show why it depends on the sign of u^{ss} in the following way (see also the illustrative scenarios in section 3.1):

$$\underline{r}\ |_{\sigma=0} \ = \begin{cases} g - \omega + \frac{\theta}{\delta} \cdot u^{ss} \ (= \ \hat{r}), & \text{if } u^{ss} < 0 \\ g - \omega, & \text{if } u^{ss} \geq 0. \end{cases} \tag{26}$$

In principle, optimal timing under perfect foresight does only depend on the anticipated path of period utility and thus on the anticipated path of the natural rate. With regard to the dependence of the optimal threshold given $\sigma = 0$ on the steady state of period utility, there are basically two scenarios of interest: one where the natural rate is initially below, and one where it is initially above its steady state:

Consider first the scenario where the natural rate is initially below its steady state, i.e., where $r_0 = r < r^{ss}$, and where period utility thus decreases in time. In this case, the calling-in move is either made immediately or never – there is nothing to wait for (as discussed in section 3.1). Clearly, it is optimal to never make the move if $r > \hat{r}$ which in a world where the natural rate is below its steady state can only occur if utility has a negative steady state.[42] However, it is optimal to make the move immediately if $r \leq \hat{r}$. If r_t is increasing and u_t thus decreasing in time waiting would involve the loss of potential benefits. In this case, the present value of the stream of period utilities given that the option is exercised in some future period t, $F(r_0 = r, T = t)$, is a decreasing function of time. Thus, if the natural rate is below its steady state, it is optimal to exercise the calling-in option immediately if $r \leq \hat{r}$ and thus if the exercise payoff $\mathcal{V}(r)$ is non-negative – regardless of whether the utility has a positive or a negative steady state.

A calling-in move at a natural rate above its steady state will generally only be made if $u^{ss} > 0$. In this case, the break-even threshold $\hat{r}$ is relatively high which reflects that the central banker would in principle accept short-run losses, i.e., negative period utilities, during some first phase and make the calling-in move at any $r_0 \leq \hat{r}$ if she lacked the flexibility to time the move freely.[43] If the flexibility to choose the optimal timing exists, those potentially acceptable losses can be avoided by deferring the calling-in move until the natural rate is sufficiently low such that positive period utilities will be realized. This is the case once the natural rate has fallen below the threshold $g - \omega$. So, if utility has a positive steady state, $\underline{r}|_{\sigma=0} = g - \omega$.

Speed of Mean-Reversion Equation (26) shows that the optimal threshold given $\sigma = 0$ only depends on the speed of mean-reversion θ if utility has a negative steady state such that $\underline{r}|_{\sigma=0} = \hat{r}$. However, as argued above, the threshold $\hat{r}$ is lower the faster the natural rate reverts to its steady state and thus the shorter the first phase is of the positive period utilities the central banker receives after the calling-in move. So, in order to accept such a

[42]Recall that if utility has a positive steady state $r > \hat{r} = (1 + \frac{\theta}{\delta})u^{ss} + r^{ss}$ would require that $r > r^{ss}$.
[43]Recall that if $r > r^{ss}$ period utility will increase as time evolves.

shorter initial phase the central banker requires higher period utilities during that phase and thus a lower natural rate at the exercise of the calling-in option. Hence, as the origins of the blue lines in figure 1(a) for $u^{ss} < 0$ show, the optimal threshold is decreasing in θ.

The origins of the blue lines in figure 1(b) shows that the optimal threshold is independent of θ if the utility has a positive steady state. If $u^{ss} > 0$, a central banker with the ability to choose the timing of the calling-in move freely will wait until the natural rate hits the level $g - \omega$ below which natural rates imply positive period utilities.[44] Obviously, $g - \omega$ is independent of θ.

4.4 Determinants of the Optimal Threshold under Uncertainty

In the following, we remove the previous section's restriction and assume that $\sigma > 0$ to discuss the effects of uncertainty over the natural rate on the optimal threshold $\underline{r}$. Essentially, uncertainty over the future path of the natural rate and thus about the future utility of making the calling-in move adds value to the flexibility to time the move freely and thus, in general, increases the central banker's tendency to wait and see. Intuition suggests that a higher tendency to wait and see from uncertainty due to a volatile natural rate will be reflected in a larger difference between the exercise threshold that would be optimal under perfect foresight and the generally optimal threshold. That is, intuition suggests that the length of the interval $[\underline{r}, \underline{r}|_{\sigma=0}]$ will be increasing in σ. We confirm this intuition for specific sets of parameter values numerically in section 5. In the following, we refer to the effect of uncertainty on the length of the interval above as "variance effect" (the variance, in turn, is increasing in σ and decreasing in θ).

However, the relationship between the *absolute level* of the optimal threshold $\underline{r}$ and uncertainty measured in terms of $\mathrm{Var}(r_t)$ (for some fixed t), can be non-monotonic if utility has a negative steady state. The blue lines in figure 1(a) show this property and it is the speed of mean-reversion θ that accounts for this non-monotonicity. The reason is that θ affects two variables: the optimal threshold under perfect foresight ($\underline{r}|_{\sigma=0,u^{ss}<0} = \hat{r}$) and the tendency to wait and see. So, in addition to a variance effect, θ has a level effect through its impact on $\hat{r}$ which in turn determines the optimal threshold under perfect foresight if utility has a negative steady state. Since these two effects are of opposite signs, the overall effect of θ on $\underline{r}$ depends on the relative size of the level effect compared

[44] Recall that $g - \omega > r^{ss}$ if $u^{ss} > 0$. So, expected period utility is monotonically increasing in time for natural rates above their steady state.

with the variance effect. With a parameter specification as in figure 1(a), the level effect will dominate for small σ whereas the variance effect will dominate if σ is large. If utility has a positive steady state, there will be no level effect, as shown by the blue lines in figure 1(b) where the optimal threshold is always increasing in θ for all $\sigma > 0$.

As opposed to this ambiguous relationship between the speed of mean-reversion and the optimal threshold, we show numerically for specific parameter constellations in the next section that a higher volatility parameter σ will reduce the level of $\underline{r}$ in these scenarios, regardless of whether θ is small or large.[45] The reason is that the exercise payoff the central banker requires in order to be willing to make the calling-in move under uncertainty is higher, the higher the extent of uncertainty is.[46] While the extent of uncertainty, in turn, is increasing in σ, the exercise payoff is larger, the lower the natural rate is at which large notes will be called in (and thus the more relevant ELB-issues are compared to forgone seignorage revenues at a low natural-rate level).

4.5 Determinants of the Time Value of the Calling-In Option

In the following, we shed light on the time value of the calling-in option in different scenarios and its dependence on the parameters of the stochastic process that governs the natural rate: the volatility parameter σ and the speed of mean reversion θ. From a policy point of view, the question is simply: When is the time value large and thus the central banker's tendency to wait and see strong? Our statements in the following hold for the numerical examples we provide in section 5 but intuition suggests that they can be generalized to arbitrary parameter constellations. To obtain comparable results, we consider the value of the calling-in option measured at $r = \hat{r}$ in the different scenarios. At this point, the option has no intrinsic but just time value ($\mathcal{V}(\hat{r}) = 0$ implies that $V(\hat{r}) = TV(\hat{r})$).

Short-Term Natural Rate Volatility In the next section, we present some numerical illustrations for different scenarios to show that a higher volatility parameter σ adds time value to the calling-in option (see the blue lines in figure 2). The reason for this positive relationship between σ and TV is that a higher σ increases the probability that the natural

[45] Alvarez and Dixit (2014) describe a non-monotonic relationship between a volatility parameter and a respective optimal threshold in a currency union's optimal stopping problem of breaking-up the union.

[46] Obviously, the requirement of a higher exercise payoff with increasing uncertainty is a main characteristic feature of equivalent financial/real option exercise problems (see, for instance, Dixit and Pindyck (1994, p. 153) in this regard).

rate will reach "exceptionally" low levels in the future. At such "exceptionally" low natural rate levels, ELB-issues will be highly relevant such that the central banker's benefits from calling in large banknotes will be "exceptionally" large (for instance, because a very low natural rate level can imply that unconstrained optimal monetary policy rates lie far below the ELB). Exceptionally large benefits, in turn, are reflected by a relatively high exercise payoff $\mathcal{V}(\underline{r})$ and thus by a relatively high time value the calling-in option has at natural rates that lie above the optimal threshold $\underline{r}$.

Speed of Mean Reversion The speed of mean reversion θ affects the calling-in option's time value through two channels: through a mean-reversion channel and through an uncertainty channel. The uncertainty channel describes the impact θ has on the TV through its effect on the variance of the natural rate in future periods and thus on uncertainty over the future path of the natural rate. For $\sigma > 0$, the variance of r_t is a decreasing function of θ (as already noted in section 4.1). Analogously to the effect of σ on the TV, a lower speed of mean reversion θ and thus higher uncertainty over the future path of the natural rate adds time value to the calling-in option. However, the speed of mean-reversion also has a negative effect on the time value through the mean-reversion channel such that the overall effect of θ on the TV depends on whether it is the uncertainty or the mean-reversion effect that dominates.

In what follows, we want to isolate the mean-reversion channel and point out how exactly θ affects the TV at $r = \hat{r}$ through this channel. For this purpose we assume, for the moment, perfect foresight and let $\sigma = 0$: Whether the calling-in option has time value due to mean-reversion or not in a $\sigma = 0$-setting depends on whether the natural rate reverts to its steady state from above or from below. As argued above, if $r_0 = r < r^{ss}$, there will be nothing to wait for the central banker. In this case, the path of period utility is decreasing in time and thus optimality requires the central banker to make the calling-in move either immediately (if $r \leq \hat{r}$) or never (otherwise). If we consider $r = \hat{r}$, the existence of time value at this point will just depend on whether $\hat{r} \lesseqgtr r^{ss}$. With $\hat{r} = g - \omega + \frac{\theta}{\delta}(g - r^{ss} - \omega) = g - \omega + \frac{\theta}{\delta}u^{ss} = (1 + \frac{\theta}{\delta})u^{ss} + r^{ss}$ showing that $\hat{r}$ is greater (less) than r^{ss} if utility has a positive (negative) steady state, it is clear that the calling-in option will have time value at $r = \hat{r}$ due to mean reversion only if $u^{ss} > 0$ (where $\underline{r}|_{u^{ss}>0} < \hat{r}|_{u^{ss}>0}$). Moreover, and obviously, if there is time value at $\hat{r}$ it will be increasing in the steady-state utility.

Since the mean-reversion channel will be relevant only if utility has a positive steady state, as argued above, it is sufficient to consider a perfect-foresight scenario where $u^{ss} > 0$: In such a scenario, the calling-in option's time value at $\hat{r}$ is increasing in θ. Two effects of a large θ contribute to a larger time value: Firstly, a higher speed of mean reversion shortens the time it takes for the natural rate to decline from $\hat{r}$ to $\underline{r}$ (although $\hat{r}$ increases in θ).[47] And secondly, a large θ increases the payoff from exercising the calling-in option at $\underline{r}$.[48] With $V(\hat{r}) = F(\hat{r}, T^*) = \mathcal{V}(\underline{r}) \cdot \exp(-\delta T^*) = TV(\hat{r})$ it is clear that a larger θ (in a $\sigma = 0$-scenario) increases the calling-in option's time value at $\hat{r}$ through its effects on the optimal exercise time and the exercise payoff.

To conclude, if utility has a positive steady state, the sign of the overall effect of θ on the TV, in general, will depend on the extent of the short-term volatility as captured by σ. In the next section, we show numerically for specific parameter values that while the effect of θ through the mean-reversion channel will dominate for small values of σ, the effect of θ through the uncertainty channel will dominate for large values of σ. The blue lines in figure 2(b) illustrate for a specific set of parameter values in a world where utility has a positive steady state, i.e., where $u^{ss} > 0$, that the time value of the calling-in option is increasing in θ for small σ while it is decreasing in θ for large σ. In contrast, if utility has a negative steady state which implies that only the uncertainty channel will be effective, the overall effect of an increase in θ will be to decrease the calling-in option's time value, as illustrated by the blue lines in figure 2(a).

5 Numerical Illustrations

In what follows, we present and discuss several numerical examples for the central banker's wait-and-see tendency in different scenarios, i.e., for specific parameter values. Our claim is that the length of the wait-and-see region $[\underline{r}, \hat{r}]$ together with the time value of the calling-in option at some $r \in [\underline{r}, \hat{r}]$ in the single scenarios can be used to assess the relative

[47] To see this, consider $T^* = \frac{1}{\theta} \ln\left(\frac{r_0 - r^{ss}}{g - r^{ss} - \omega}\right)$ which for $r_0 = \hat{r} = (1+\frac{\theta}{\delta})u^{ss} + r^{ss}$ becomes $T^* = \frac{1}{\theta} \cdot \ln\left(\frac{\delta+\theta}{\delta}\right)$. It is easy to show that $\frac{\partial T^*}{\partial \theta} = \frac{1}{\theta(\delta+\theta)} - \frac{\ln\left(\frac{\delta+\theta}{\delta}\right)}{\theta^2} < 0$ by making the substitution $x = \frac{\delta+\theta}{\delta}$ and then using the mean value theorem to show that $1 - \frac{1}{x} < \ln(x)$ for all $x > 1$.

[48] To see this, note that at the point in time when the calling-in option is optimally exercised (which for $\sigma = 0$ is reached when $\underline{r} = g - \omega$), the natural rate rate will be still above its steady state. So, period utility from that point on is reverting to its steady state from below. The faster this reversion, the larger the present value of the stream of future period utilities. Hence, the exercise payoff at $\underline{r}$ is increasing in θ. Formally, this is obvious with $\mathcal{V}(\underline{r}) = F(r = g - \omega, T = 0) = \frac{1}{\delta}(g - r^{ss} - \omega) - \frac{1}{\delta+\theta}(g - r^{ss} - \omega) = \frac{\theta}{\delta(\delta+\theta)} u^{ss}$ and $\frac{\partial \mathcal{V}(\underline{r})}{\partial \theta} = \frac{1}{(\delta+\theta)^2} u^{ss} > 0$.

likelihoods that wait-and-see behavior might actually be observed in corresponding scenarios in practice. The results in figures 1 to 3, and tables 1 and 2 have been obtained with the computer algebra system *Mathematica* (see Wolfram Research, Inc., 2015) by using the function "`FindRoot`" to solve equation (25) numerically for $\underline{r}$ and then using (24) to evaluate $V(r)$ as defined in (23). Figure 1 shows how the break-even threshold $\hat{r}$ and the optimal threshold $\underline{r}$ depend on the steady-state utility u^{ss}, on the speed of mean-reversion θ, and on the volatility parameter σ. Figure 2 shows how the value V of the calling-in option measured at $r = \hat{r}$ (compared to the exercise payoff $\mathcal{V}$) depends on u^{ss}, θ, and σ. Figure 3 shows how the exercise payoff of the calling-in option evaluated at $r = \underline{r}$ where $\mathcal{V}(\underline{r}) = V(\underline{r})$ depends on these parameters. For the sake of clarity, tables 1 and 2 show some values of $V(\hat{r})$ and $\mathcal{V}(\underline{r})$ for selected parameter constellations.

Scenarios with Negative Steady-State Utility For the regions of the parameter space specified below we show numerically that the length of the interval $[\underline{r}, \hat{r}]$, the time value of the calling-in option at $\hat{r}$, as well as the exercise payoff at $\underline{r}$ will increase in σ and decrease in θ if utility has a negative steady state (see also the argumentation in section 4). So there will be little room for wait-and-see behavior if uncertainty over the future path of the natural rate is relatively low. This implies that the natural rate of interest does not have to be at levels that are far below the break-even threshold to induce the central banker to make the calling-in move. In the numerical examples given in tables 1 and 2 such scenarios (i.e., scenarios with relatively low uncertainty over the future path of the natural rate in a world where calling in large notes involves long-run losses for the central banker) are described for $u^{ss} < 0$, $\sigma \in \{0.01,\ 0.5\}$, and $\theta \in \{0.5,\ 1\}$. While table 1 shows that the time value of the calling-in option measured at the break-even threshold will be relatively small in the aforementioned scenarios, table 2 shows that the net benefits the central banker will require to make the calling-in move in these scenarios are close to or just slightly above zero. Thus, in corresponding scenarios in practice, the move could be made as soon as the natural rate is so low that eliminating large banknotes will have net benefits that are just slightly above zero. A severe ELB-episode during or in the aftermath of a large-scale financial or economic crisis could be such a situation. Calling in large banknotes would be rational in such a scenario if the short-run benefits exceeded the long-run losses and waiting to make the move implied losing short-run benefits.

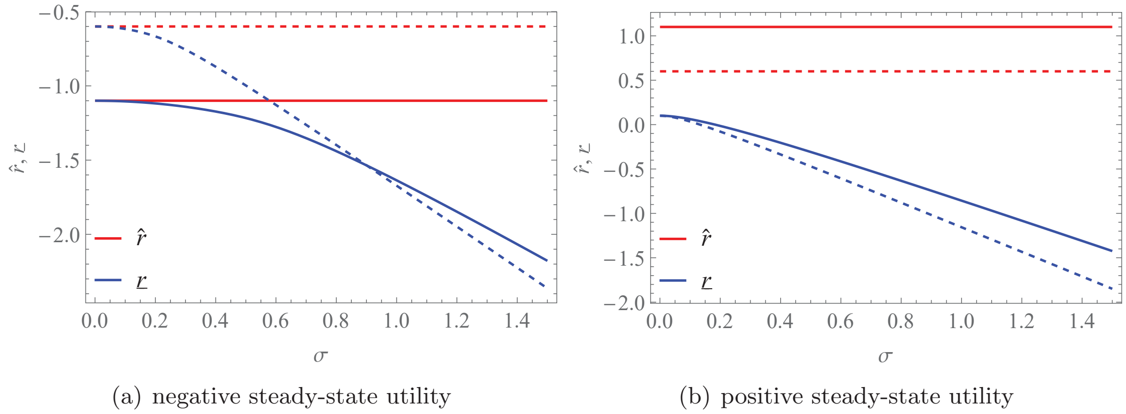


Figure 1: Break-even threshold $\hat{r}$ (red lines) and optimal threshold $\underline{r}$ (blue lines) for high (solid lines) and low (dashed lines) speed of mean reversion in dependence on the volatility parameter σ (parameter values (a): $r^{ss} = 0,\ g - \omega = -0.1,\ \delta = 0.1$; parameter values (b): $r^{ss} = 0,\ g - \omega = 0.1,\ \delta = 0.1$; solid lines: $\theta = 1$, dashed lines: $\theta = 0.5$).

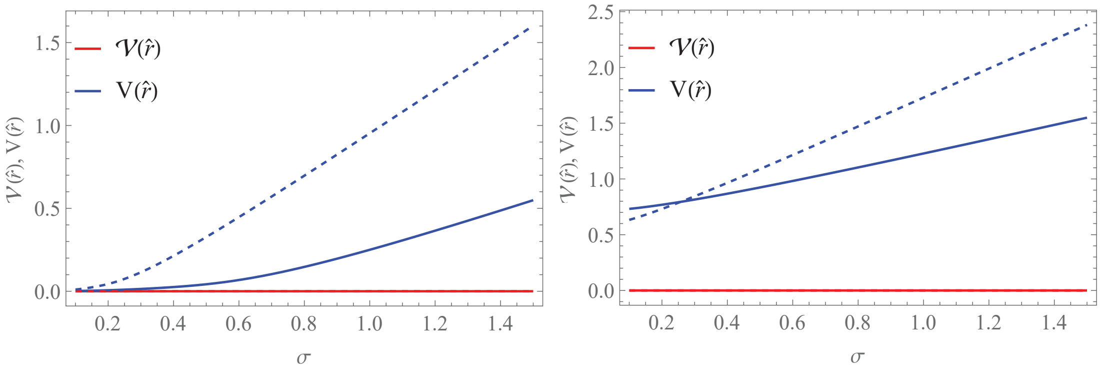
Time value of calling-in option at break-even threshold $V(\hat{r}) = TV(\hat{r})$						
	$\sigma = 0.01$		$\sigma = 0.5$		$\sigma = 1$	
u^{ss}	$\theta = 0.5$	$\theta = 1$	$\theta = 0.5$	$\theta = 1$	$\theta = 0.5$	$\theta = 1$
-0.2	≈ 0	≈ 0	0.148	0.019	0.654	0.084
-0.1	≈ 0	≈ 0	0.327	0.042	0.953	0.25
0	0.013	0.007	0.658	0.335	1.317	0.67
0.1	0.583	0.715	1.089	0.924	1.73	1.228
0.2	1.165	1.431	1.573	1.582	2.178	1.847

Table 1: Numerical solutions of the calling-in option's value at $r = \hat{r}$ for different values of the steady-state utility u^{ss}, the volatility parameter σ, and the speed of mean-reversion θ with $r^{ss} = 0$ and $\delta = 0.1$. All values are rounded to three decimal places.

Exercise payoff at optimal threshold $\mathcal{V}(\underline{r})$							
		$\sigma = 0.01$		$\sigma = 0.5$		$\sigma = 1$	
u^{ss}	$\mathcal{U}$	$\theta = 0.5$	$\theta = 1$	$\theta = 0.5$	$\theta = 1$	$\theta = 0.5$	$\theta = 1$
-0.2	-2	≈ 0	≈ 0	0.356	0.052	1.324	0.212
-0.1	-1	≈ 0	≈ 0	0.662	0.106	1.787	0.487
0	0	0.023	0.011	1.164	0.53	2.328	1.061
0.1	1	0.835	0.91	1.784	1.28	2.926	1.778
0.2	2	1.668	1.818	2.469	2.107	3.567	2.56

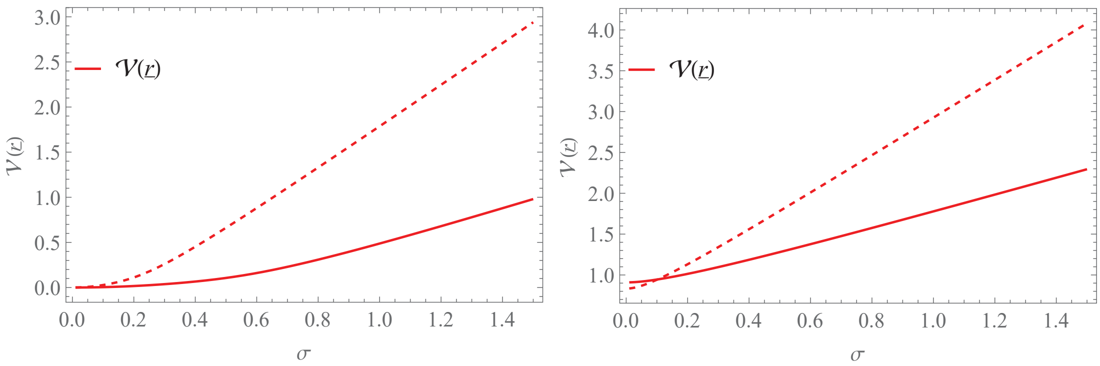
Table 2: Numerical solutions of the exercise payoff at $r = \underline{r}$ for different values of the steady-state utility u^{ss} (with $\mathcal{U}$ being defined as $\mathcal{U} = \int_0^\infty u^{ss} \cdot \exp(-\delta t) dt = \frac{1}{\delta} u^{ss}$), the volatility parameter σ, and the speed of mean-reversion θ with $r^{ss} = 0$ and $\delta = 0.1$. All values are rounded to three decimal places.

Scenarios with Positive Steady-State Utility In section 4, we discuss the non-monotonic relationship between uncertainty over the future path of the natural rate of



(a) negative steady-state utility (b) positive steady-state utility

Figure 2: Exercise payoff $\mathcal{V}(\hat{r})$ (red lines) and option value $V(\hat{r})$ (blue lines) for high (solid lines) and low (dashed lines) speed of mean reversion in dependence on the volatility parameter σ (parameter values (a): $r^{ss} = 0,\ g - \omega = -0.1,\ \delta = 0.1$; parameter values (b): $r^{ss} = 0,\ g - \omega = 0.1,\ \delta = 0.1$; solid lines: $\theta = 1$, dashed lines: $\theta = 0.5$).



(a) negative steady-state utility (b) positive steady-state utility

Figure 3: Exercise payoff $\mathcal{V}(\underline{r})$ for high (solid lines) and low (dashed lines) speed of mean reversion in dependence on the volatility parameter σ (parameter values (a): $r^{ss} = 0,\ g - \omega = -0.1,\ \delta = 0.1$; parameter values (b): $r^{ss} = 0,\ g - \omega = 0.1,\ \delta = 0.1$; solid lines: $\theta = 1$, dashed lines: $\theta = 0.5$).

interest (as reflected in $\mathrm{Var}(r_t)$) and the time value of the calling-in option in a world where the utility from calling in large notes has a positive steady state. We also describe the behavior of the break-even threshold and the optimal threshold which will both increase in the speed of mean-reversion θ if $u^{ss} > 0$ – this means that the overall effect of a change in θ on the length of the wait-and-see region $[\underline{r},\ \hat{r}]$ depends on which bound of the interval reacts more sensitively to a change in θ. Taken all together, these properties imply that under special circumstances there will be more room for wait-and-see behavior if uncertainty over the natural rate is low (and not high, as intuition suggests). The numerical examples given in figure 1(b) and tables 1 and 2 for $u^{ss} > 0$ and $\sigma = 0.01$ show this case: the length of $[\underline{r},\ \hat{r}]$, the time value at $\hat{r}$ and the exercise payoff at $\underline{r}$ are all greater the lower the uncertainty over the natural rate is (recall that $\text{Var}(r_t)|_{\sigma=0.01,\ \theta=0.5} > \text{Var}(r_t)|_{\sigma=0.01,\ \theta=1}$).

But beyond that, the numerical examples primarily illustrate that the room for wait-and-see behavior will in general be relatively large if utility has a positive steady state. Table 2 shows for $u^{ss} > 0$ that the net benefit the central banker will require to make the calling-in move under relatively high levels of uncertainty ($\sigma \geq 0.5$) is significantly greater than zero and even a multiple of the present value of the infinite stream of steady-state utility, $\mathcal{U}$, defined as $\mathcal{U} = \int_0^\infty u^{ss} \cdot \exp(-\delta t)dt$. For $u^{ss} = 0.1$, $\sigma = 1$, and $\theta = 0.5$ the exercise payoff at the optimal threshold $\underline{r}$ is almost three times larger than the present value $\mathcal{U}$ of the infinite stream of steady-state utility. In practice, central banks could keep issuing large banknotes for a long time in a corresponding scenario even though it may already be beneficial to eliminate them immediately.

6 Conclusion

The goal of this paper is to stimulate research on optimal timing issues associated with any plans to phase out large-denomination banknotes. This research is essential because the debate on such plans is incomplete if it is only concerned with the costs and benefits while any timing aspects and a potential wait-and-see component of an optimal timing strategy in an uncertain economic environment are ignored. We condense the stochastic state of the economy into the natural rate and employ an optimal stopping model as a framework to explore such timing issues and to rationalize a central banker's wait-and-see tendency. The purpose of this approach is not to determine the exact empirical magnitudes of our results but to make clear that the stochastic properties and the expected path of the natural rate can be used as a first rough indicator of wait-and-see behavior in practice. In concrete terms, this means that the volatility and the expected path of the natural rate can be used to gauge whether the issuance of large notes could (in a positive dimension) or should (in a normative dimension) continue for years or decades even if the expected net benefits from calling them in right now were already greater than zero.

We use several numerical examples to illustrate states of the world where an optimal timing strategy to call in large notes involves or, on the other hand, does not involve a wait-and-see component. On the basis of these theoretical examples, we can state some conjectures and formulate three hypotheses regarding the existence of a wait-and-see component in practice: 1) We expect that a wait-and-see component exists in a world where the long-run net benefits from calling in large banknotes are positive rather than in a world where lost seignorage revenues outweigh ELB-issues in the long run; 2) that it exists in a world where the expected path of the natural rate is decreasing rather than in a world where ELB-issues will become less and less relevant as time evolves; 3) and that it exists in a world where the natural rate is highly volatile rather than in a world where there is little uncertainty over the occurrence and duration of ELB-episodes in the near future. For corresponding scenarios, the net benefits a central banker requires to be willing to call in large notes can by far be greater than zero. This can imply that the option to call in large banknotes will not be exercised until the natural rate has fallen to an exceptionally low level. Such situations could emerge in the course of pronounced recessions or in the aftermath of large-scale economic crises where monetary policy remains stuck at the ELB.

On the basis of our simple model, a guesstimate of the ECB's stance on plans to stop the issuance of the 200- or the 100-euro note is not difficult to divine: It is natural to assume that the ECB – if ever – will only make such a move during a severe ELB-episode. Of course, this speculation is highly sensitive to the assumptions we make. Any serious forecast in this regard will require at least a large-scale macro model as a framework to analyze timing issues associated with calling-in plans. Our model is highly stylized and has left out a number of factors that could change our results. For instance, the specification with the mean-reverting Ornstein-Uhlenbeck process we chose can easily be extended by assuming that the natural rate follows a jump diffusion. This could account for the hypothesis that large-scale economic or financial crises can lead to exceptional drops in the natural rate in their immediate aftermath (as indicated by the findings of Holston, Laubach, and Williams (2017) for the global financial crisis). We have also ignored that there is far more than one dimension of uncertainty. It is not only the future path of the natural rate that is unknown, but also the natural rate itself and its historic path. The reason is that the natural rate must be estimated and cannot be measured directly.[49] Moreover, there is also much uncertainty regarding the costs and benefits from calling in large notes. It could be hard to quantify them precisely. Another crucial assumption we make is that the central banker has full flexibility to time a calling-in move freely. A central banker approaching the end of her term in office (or a central bank's decision-making body shortly before its members change) could be driven by a precautionary motive from the

[49] See, for instance, Weber, Lemke, and Worms (2008, section 5).

fear that her successor will have a different objective function (another issue is a potential intervention of the government which could be driven by its own objective function and could try to change the denominational structure of banknotes by law). A number of further research questions will arise from allowing for the possibility of making sequential calling-in moves (e.g., to stop the issuance of the 100-euro note at a later date than the issuance of the 200-euro note) as well as from considering strategic interactions between central bankers in a multi-country setting.

Appendix

A Detailed Solution of the Central Banker's Optimal Stopping Problem

A.1 Derivation of the Bellman Equation written as Ordinary Differential Equation

We show how to derive the Bellman ordinary differential equation (18) starting from equation (17). In technical regards, we refer to Dixit and Pindyck (1994, pp. 140–141) where the single steps we have to take are pointed out (note, that the differential equation in Dixit and Pindyck (1994, p. 140, equation (8)) results from geometric Brownian motion and thus, obviously, differs from (18)).

For the derivation of (18) that follows we use that

$$(dr_t)^2 = (dr_t) \cdot (dr_t) \tag{27}$$

$$= \theta^2(r^{ss} - r_t)^2(dt)^2 + 2 \cdot \theta(r^{ss} - r_t)dt \cdot \sigma dB_t + \sigma^2(dB_t)^2 \tag{28}$$

$$= \sigma^2 dt, \tag{29}$$

which is obtained by using the rules $dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0$ and $dB_t \cdot dB_t = dt$ (recall that dr_t describes the dynamics of the mean-reverting Ornstein-Uhlenbeck process that governs the natural rate of interest with $dr_t = \theta(r^{ss} - r_t)dt + \sigma dB_t$ where B_t is Brownian motion and $\theta \in \mathbb{R}_{>0}$, $r^{ss} \in \mathbb{R}$, and $\sigma \in \mathbb{R}_{>0}$ are known constants).[50]

[50] For the rules used to compute $(dr_t)^2$ see, for instance, Øksendal (2013, p. 45, equation (4.1.8)).

Now, let $V : \mathbb{R} \to \mathbb{R}$ be a twice-differentiable function for the value of the calling-in option dependent on the natural rate of interest r. Applying the Itô formula, using (3) and (29), we obtain

$$dV = V' dr_t + \frac{1}{2} V''(dr_t)^2 \tag{30}$$

$$= V' \cdot (\theta(r^{ss} - r_t)dt + \sigma dB_t) + \frac{1}{2} V'' \cdot \sigma^2 dt \tag{31}$$

with

$$\mathbb{E}[dV | r_t = r] = \theta(r^{ss} - r) \cdot V' dt + \frac{1}{2}\sigma^2 \cdot V'' dt \tag{32}$$

which is implied by $\mathbb{E}[dB_t] = 0$.[51] The Bellman ordinary differential equation (18) can now be obtained by using (32) to replace the right-hand side of (17).

A.2 Proof of Proposition 1

In section A.3 of this appendix, we point out *how* the solution of (18) can be obtained. In this section, we just prove proposition 1 and show, first, that (23) solves the Bellman ODE (18), second, that this solution satisfies the monotonicity condition and the non-negativity constraint, third, how to obtain c_1 in (24), and fourth, how to obtain the implicit definition of $\underline{r}$ in (25).

The first part of the proof is thus to show that

$$\frac{1}{2}\sigma^2 V'' + \theta(r^{ss} - r)V' - \delta V = 0 \tag{33}$$

for

$$V(r) = c_1 \cdot H_{-\frac{\delta}{\theta}} \left(\frac{\sqrt{\theta}}{\sigma} \cdot (r - r^{ss}) \right) \tag{34}$$

where $H_\nu(z)$ denotes a Hermite function (see, for instance, Lebedev, 1965, p. 285).

[51] Note that the Itô formula with the notation used here is given in Dixit and Pindyck (1994, pp. 79–81, 140). A discussion of the Itô formula in greater depth and in a general context is given, for instance, in Øksendal (2013, p. 44).

We use the following representations of the first and second derivative of the Hermite function as given in Lebedev (1965, p. 289) with

$$H'_\nu(z) \coloneqq \frac{\partial H_\nu(z)}{\partial z} = 2\nu H_{\nu-1}(z), \tag{35}$$

$$H''_\nu(z) \coloneqq \frac{\partial^2 H_\nu(z)}{\partial z^2} = 2\nu H'_{\nu-1}(z), \tag{36}$$

to obtain

$$V'(r) \coloneqq \frac{\partial V(r)}{\partial r} = c_1 \cdot 2 \cdot \left(-\frac{\delta}{\theta}\right) \cdot H_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma} \cdot (r - r^{ss})\right) \cdot \frac{\sqrt{\theta}}{\sigma} \tag{37}$$

$$= -2c_1 \frac{\delta}{\sigma\sqrt{\theta}} H_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma} \cdot (r - r^{ss})\right) \tag{38}$$

and

$$V''(r) \coloneqq \frac{\partial^2 V(r)}{\partial r^2} = -2c_1 \frac{\delta}{\sigma\sqrt{\theta}} \cdot H'_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma} \cdot (r - r^{ss})\right) \cdot \frac{\sqrt{\theta}}{\sigma} \tag{39}$$

$$= -2c_1 \frac{\delta}{\sigma^2} H'_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma} \cdot (r - r^{ss})\right). \tag{40}$$

Now, we can use (34), (38), and (40) to reformulate the left-hand side of the Bellman ODE (33) and then show that this expression is in fact zero by making the following transformations:

$$\frac{1}{2}\sigma^2 \cdot (-2)c_1 \frac{\delta}{\sigma^2} H'_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right) + \theta(r^{ss} - r) \cdot (-2)c_1 \frac{\delta}{\sigma\sqrt{\theta}} H_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right)$$
$$- \delta c_1 H_{-\frac{\delta}{\theta}}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right) = 0 \quad \Big| \cdot \frac{1}{c_1} \tag{41}$$

$$\Leftrightarrow -\delta H'_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right) - 2\theta(r^{ss} - r)\frac{\delta}{\sigma\sqrt{\theta}} H_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right)$$
$$- \delta H_{-\frac{\delta}{\theta}}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right) = 0 \quad \Big| \cdot \frac{2}{\theta} \tag{42}$$

$$\Leftrightarrow -2\frac{\delta}{\theta} H'_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right) = 2\frac{\sqrt{\theta}}{\sigma}(r^{ss} - r) \cdot 2\frac{\delta}{\theta} H_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right)$$
$$+ 2\frac{\delta}{\theta} H_{-\frac{\delta}{\theta}}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right) \quad \Big| \text{substituting } \nu \coloneqq -\frac{\delta}{\theta} \text{ and } z \coloneqq \frac{\sqrt{\theta}}{\sigma}(r - r^{ss}) \tag{43}$$

$$\Leftrightarrow 2\nu H'_{\nu-1}(z) = 2zH'_\nu(z) - 2\nu H_\nu(z) \tag{44}$$

where the last transformation is obtained by applying (35) to the first term on the right-hand side of (43). As shown in Lebedev (1965, p. 289, equation (10.4.5)), equation (44) is true – thus (41) is true, too. This proves that (23) is a solution of the Bellman ODE (18).

Let us now show that (23) satisfies the monotonicity condition (21) and the non-negativity constraint (22). To see that the non-negativity constraint is satisfied, consider the integral representation of the Hermite function for Re $\nu < 0$ as given in Lebedev (1965, p. 290, equation (10.5.2)) with

$$H_\nu(z) = \frac{1}{\Gamma(-\nu)} \int_0^\infty \exp(-t^2 - 2tz) t^{-\nu-1} dt, \tag{45}$$

where $\Gamma(\cdot)$ is the Gamma function as defined in Lebedev (1965, p. 1). With $\nu = -\frac{\delta}{\theta} \in \mathbb{R}_{<0}$ and therefore $\Gamma(-\nu) \in \mathbb{R}_{>0}$ it becomes immediately clear that $H_\nu(z) > 0$ for all $z \in \mathbb{R}$, and thus that (23) is non-negative for all $c_1 \in \mathbb{R}_{>0}$ and for all $\frac{\delta}{\theta} > 0$. That c_1 as defined in (24) is in fact greater than zero is easy to see with the same argument.

To see that the monotonicity condition is satisfied, consider the representation of the first derivative of the Hermite function as given in Lebedev (1965, p. 289) with $H'_\nu(z) = 2\nu H_{\nu-1}(z)$. For $\nu - 1 < 0$, it can be shown with the same argumentation as above (where we showed that the non-negativity constraint is satisfied) that $H_{\nu-1}(z) > 0$ for all $z \in \mathbb{R}$. With (38) stating that $V'(r) = -2c_1 \frac{\delta}{\sigma\sqrt{\theta}} H_{-\frac{\delta}{\theta}-1}\left(\frac{\sqrt{\theta}}{\sigma} \cdot (r - r^{ss})\right)$ it is immediately clear that $V'(r) < 0$ for all $r \in (\underline{r}, \infty)$ (recall and see above that $c_1 > 0$).

To see how c_1 as defined in (24) is obtained, consider the smooth-pasting condition (20) which requires that $V'(\underline{r}) \stackrel{!}{=} \frac{\partial \mathcal{V}(\underline{r})}{\partial \underline{r}}$. Using (38) and

$$\frac{\partial \mathcal{V}(\underline{r})}{\partial \underline{r}} = -\frac{1}{\delta + \theta}, \tag{46}$$

(which is obtained by differentiating $\mathcal{V}$ as defined in (10)), the smooth-pasting condition can be rearranged to obtain c_1 as in (24).

To see how the implicit definition of $\underline{r}$ in (25) is obtained, consider the value-matching condition (19) requiring that $V(\underline{r}) \stackrel{!}{=} \mathcal{V}(\underline{r})$. The left-hand side of equation (25) is directly obtained by using (23) for V with c_1 in the representation given by (24). The right-hand side of equation (25) is just the exercise payoff as defined in (10) evaluated at $\underline{r}$. □

A.3 Solution of the Bellman Ordinary Differential Equation

Let us now outline how we obtained a solution of the central banker's optimal stopping problem and thus of the free boundary problem of solving the Bellman ODE (18) and finding $\underline{r}$. As a first step, we used the computer algebra system *Mathematica* (see Wolfram Research, Inc., 2015) to obtain a solution of (18) by using the *Mathematica*-function "`DSolve`". *Mathematica* returned two linearly independent solutions – one of them being (23), the other one being a Kummer confluent hypergeometric function. After having proved that (23) solves (18) as outlined in section A.2 of this appendix, and since we do not need to prove uniqueness of our solution, we chose (23) to solve the central banker's optimal stopping problem.

To see that this solution is in fact correct, we can use the argument $\frac{\sqrt{\theta}}{\sigma} \cdot (r - r^{ss})$ of the Hermite function returned by Mathematica in order to transform (18) into a canonical form and then look up in Lebedev (1965) for the solution of this differential equation.[52] To transform (18), we use the technique outlined by Dixit and Pindyck (1994, p. 163) and accordingly introduce

$$z(r) = \frac{\sqrt{\theta}}{\sigma}(r - r^{ss}) \text{ with } z'(r) = \frac{\sqrt{\theta}}{\sigma}, \tag{47}$$

and use a function $w(z)$ to substitute

$$V(r) = w(z) \text{ with } V'(r) = w'(z(r)) \cdot z'(r) = \frac{\sqrt{\theta}}{\sigma} w'(z) \text{ and } V''(r) = \frac{\theta}{\sigma^2} w''(z). \tag{48}$$

Therewith, and with $r = \frac{\sigma}{\sqrt{\theta}} z + r^{ss}$, equation (18) can be transformed into

$$\frac{1}{2}\sigma^2 \frac{\theta}{\sigma^2} w''(z) + \theta \left(r^{ss} - \left(\frac{\sigma}{\sqrt{\theta}} z + r^{ss}\right) \right) \frac{\sqrt{\theta}}{\sigma} w'(z) - \delta w(z) = 0, \tag{49}$$

which can be simplified to

$$w'' - 2zw' + 2\nu w = 0, \tag{50}$$

where $\nu := -\frac{\delta}{\theta} < 0$. This differential equation, its solutions, and the Hermite function are discussed in great detail, for instance, in Lebedev (1965, pp. 283-299).

[52] See, for instance, Suzuki (2016, p. 35, equation (2)) for an equivalent substitution.

The general solution of (50) as stated by Lebedev (1965, p. 286, equation (10.2.17)) reads

$$w = MH_{\nu}(z) + N\exp(z^2)H_{-\nu-1}(\mathbb{i}z), \tag{51}$$

where M, N are constants, $\mathbb{i}^2 = -1$, and $H_{\nu}(z)$ is a Hermite function (the definition of a Hermite function is given by Lebedev (1965, p. 285)).

Re-substituting with $w = V$, $z = \frac{\sqrt{\theta}}{\sigma}(r - r^{ss})$ and $\nu = -\frac{\delta}{\theta}$, the general solution of the Bellman equation (18) can be written as

$$\begin{aligned} V(r) = {} & c_1 H_{-\frac{\delta}{\theta}}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right) \\ & + c_2 \exp\left(\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right)^2\right) H_{\frac{\delta}{\theta}-1}\left(\mathbb{i}\left(\frac{\sqrt{\theta}}{\sigma}(r - r^{ss})\right)\right), \end{aligned} \tag{52}$$

where c_1, c_2 are constants. As argued above, we now let $c_2 = 0$.[53] This yields a particular solution that satisfies the monotonicity condition (21) and the non-negativity constraint (22). As also shown above, (19) and (20) can now be used to determine c_1 and $\underline{r}$.

Bibliography

Agarwal, R., and M. S. Kimball (2019): "Enabling Deep Negative Rates to Fight Recessions: A Guide," IMF Working Paper 19/84, International Monetary Fund.

Alvarez, F., and A. Dixit (2014): "A Real Options Perspective on the Future of the Euro," *Journal of Monetary Economics*, 61, 78–109.

Buiter, W. H. (2007a): "Is Numérairology the Future of Monetary Economics?," *Open Economies Review*, 18(2), 127–156.

——— (2007b): "Seigniorage," *Economics: The Open-Access, Open-Assessment E-Journal*, 1(2007-10), 1–49.

——— (2009): "Negative Nominal Interest Rates: Three Ways to Overcome the Zero Lower Bound," *The North American Journal of Economics and Finance*, 20(3), 213 – 238.

Buiter, W. H., and N. Panigirtzoglou (2003): "Overcoming the Zero Bound on Nominal Interest Rates With Negative Interest on Currency: Gesell's Solution," *The Economic Journal*, 113(490), 723–746.

[53] See also, for instance, Parlour and Walden (2009, pp. 14-15), Garlappi and Yan (2011, p. 819), or Suzuki (2016, p. 39) who have equivalent/similar differential equations and obtain equivalent/similar solutions.

Buiter, W. H., and E. Rahbari (2015): "High Time to Get Low: Getting Rid of the Lower Bound on Nominal Interest Rates," Global Economics View, Citi Research, Citigroup Global Markets Inc., 9 April 2015, available at `https://willembuiter.com/ELB.pdf`.

Chevalier-Roignant, B., and L. Trigeorgis (2011): *Competitive Strategy: Options and Games*. MIT Press, Cambridge, Massachusetts and London, England.

Chung, H., E. Gagnon, T. Nakata, M. Paustian, B. Schlusche, J. Trevino, D. Vilán, and W. Zheng (2019): "Monetary Policy Options at the Effective Lower Bound: Assessing the Federal Reserve's Current Policy Toolkit," Finance and Economics Discussion Series 2019-003. Washington: Board of Governors of the Federal Reserve System, available at `https://doi.org/10.17016/FEDS.2019.003`.

Dixit, A. K., and R. S. Pindyck (1994): *Investment under Uncertainty*. Princeton University Press, Princeton, New Jersey.

European Central Bank (2009): "ECB Annual Report 2008," European Central Bank, Frankfurt am Main, available at `https://www.ecb.europa.eu/pub/pdf/annrep/ar2008en.pdf?0da1f91236eeb11eed9f832a4a874095`.

——— (2019): "Annual Accounts of the ECB 2018," European Central Bank, Frankfurt am Main, available at `https://www.ecb.europa.eu/pub/annual/annual-accounts` and included in "ECB Annual Report 2018", European Central Bank, pp. A1–A66.

Fischer, B., P. Köhler, and F. Seitz (2004): "The Demand for Euro Area Currencies: Past, Present and Future," Working Paper Series No. 330, European Central Bank.

Galí, J. (2015): *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and its Applications*. Princeton University Press, Princeton and Oxford, second edn.

Garlappi, L., and H. Yan (2011): "Financial Distress and the Cross-Section of Equity Returns," *The Journal of Finance*, 66(3), 789–822.

Goodfriend, M. (2000): "Overcoming the Zero Bound on Interest Rate Policy," *Journal of Money, Credit and Banking*, 32(4), 1007–1035.

Holston, K., T. Laubach, and J. C. Williams (2017): "Measuring the Natural Rate of Interest: International Trends and Determinants," *Journal of International Economics*, 108, S59–S75.

Hull, J. C. (2019): *Optionen, Futures und andere Derivate*. Pearson Deutschland GmbH, Hallbergmoos, tenth edn.

Kiley, M. T., and J. M. Roberts (2017): "Monetary Policy in a Low Interest Rate World," *Brookings Papers on Economic Activity*, Spring 2017, 317–396.

Krüger, M., and F. Seitz (2017): *The Benefits of Cash: Costs and Benefits of Cash and Cashless Payment Instruments (Module 2)*. Fritz Knapp Verlag, Frankfurt am Main.

——— (2018): "Pros and Cons of Cash: The State of the Debate," *Credit and Capital Markets*, 51(1), 15–40.

Lebedev, N. N. (1965): *Special Functions and their Applications*. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Maller, R. A., G. Müller, and A. Szimayer (2009): "Ornstein–Uhlenbeck Processes and Extensions," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch, pp. 421–437. Springer, Berlin Heidelberg.

McAndrews, J. (2015): "Negative Nominal Central Bank Policy Rates: Where is the Lower Bound?," Remarks at the University of Wisconsin, May 8, 2015; available at: `https://www.newyorkfed.org/newsevents/speeches/2015/mca150508.html`.

Mersch, Y. (2018): "The Role of Euro Banknotes as Legal Tender," Speech at the 4th Bargeldsymposium of the Deutsche Bundesbank, Frankfurt am Main, 14 February 2018; available at: `https://www.ecb.europa.eu/press/key/date/2018/html/ecb.sp180214.en.html`.

Øksendal, B. (2013): *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin Heidelberg, sixth edn., sixth corrected printing.

Parlour, C. A., and J. Walden (2009): "Competition and Asset Prices," unpublished manuscript, February 2009, available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1344876`.

Peskir, G., and A. Shiryaev (2006): *Optimal Stopping and Free-Boundary Problems*, Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, Switzerland.

Rognlie, M. (2016): "What Lower Bound? Monetary Policy with Negative Interest Rates," unpublished manuscript, July 2016, available at `http://mattrognlie.com/negative_rates.pdf`.

Rogoff, K. (1998): "Blessing or Curse? Foreign and Underground Demand for Euro Notes," *Economic Policy*, 13(26), 262–303.

——— (2015): "Costs and Benefits to Phasing out Paper Currency," *NBER Macroeconomics Annual*, 29, 445–456.

——— (2016): *The Curse of Cash*. Princeton University Press, Princeton and Oxford.

——— (2017a): "Dealing with Monetary Paralysis at the Zero Bound," *Journal of Economic Perspectives*, 31(3), 47–66.

——— (2017b): "Monetary Policy in a Low Interest Rate World," *Journal of Policy Modeling*, 39(4), 673–679.

Suzuki, K. (2016): "Optimal Switching Strategy of a Mean-Reverting Asset over Multiple Regimes," *Automatica*, 67, 33–45.

Thiele, C.-L., D. Niepelt, M. Krüger, F. Seitz, R. Halver, and A. F. Michler (2015): "Diskussion um das Bargeld: Hätte eine Abschaffung von Banknoten und Münzen wirklich Vorteile?," *ifo Schnelldienst*, 68(13), 3–18.

Weber, A. A., W. Lemke, and A. Worms (2008): "How Useful is the Concept of the Natural Real Rate of Interest for Monetary Policy?," *Cambridge Journal of Economics*, 32(1), 49–63.

Wolfram Research, Inc. (2015): "Wolfram Mathematica, Version 10.1," Champaign, Illinois.

Yates, T. (2004): "Monetary Policy and the Zero Bound to Interest Rates: A Review," *Journal of Economic Surveys*, 18(3), 427–481.

International Coordination and Optimal Timing of Calling In Large-Denomination Banknotes in a Two-Player Game

Thomas Link

Abstract

This paper proposes a two-country version of an optimal timing model of calling in large-denomination banknotes. Our goal is to explore whether central banks should coordinate the elimination of large notes internationally. We find that coordination can prevent central bankers from inefficient timing decisions. The inefficiencies in our model have their root in central bankers' expectations of extra seignorage gains or losses that arise when timing strategies diverge. The gains from coordination depend on the substitutability of banknotes of different denominations and currencies. The substitutability also determines whether central bankers face a prisoner's dilemma or a coordination problem. Under certain circumstances, optimality requires two fully symmetric central bankers to call in large notes sequentially.

JEL classification: E42, E58

Keywords: cashless economy, phase-out of paper currency, timing game, international monetary policy coordination

Contents

Contributions, Acknowledgments, and Published Paper Versions

The paper “International Coordination and Optimal Timing of Calling In Large-Denomination Banknotes in a Two-Player Game” is a single author paper by Thomas Link. I am grateful to Ulrike Neyer and Hans-Theo Normann for the discussions and precious remarks. I also thank Jennifer Rontganger for proofreading the paper. Furthermore, I thank seminar and conference participants at the Heinrich Heine University Düsseldorf, the University of Gießen, Sciences Po Aix (Aix-en-Provence), and Leipzig University for the fruitful discussions and many helpful comments.

A very early draft from which the papers “International Coordination and Optimal Timing of Calling In Large-Denomination Banknotes in a Two-Player Game” and “Optimal Timing of Calling In Large-Denomination Banknotes under Natural Rate Uncertainty” (included as second essay in this thesis) emerged has been made available online for presentation at the “35th International Symposium on Money, Banking and Finance”, organized by the European Research Group (GdRE) on Money, Banking and Finance, and held at Sciences Po Aix (Aix-en-Provence) in June 2018. This early draft is available as unpublished manuscript at `https://gdre-scpo-aix.sciencesconf.org/195408/document`.

List of Figures

List of Tables

1 Introduction

Several options to relax the effective lower bound (ELB) constraint on monetary policy are available (see Goodfriend, 2000; Buiter, 2009; Rogoff, 2017; Agarwal and Kimball, 2019). Calling in or gradually "phasing out" large-denomination banknotes is one of them (see Rogoff, 2017, p. 57). In a recent paper, we took the perspective of a central banker and showed how the volatility and the expected path of the natural rate of interest determine the optimal timing of calling in large-denomination banknotes (see Link, 2019). There, our starting point was that the central banker's net benefits from calling in large notes depend on the stochastic state of the world and in particular on the natural rate of interest. This state dependency, in turn, creates a need to time a calling-in move optimally. Now, we use a two-country version of the optimal timing model proposed in Link (2019) with two central bankers to explore whether the elimination of large-denomination banknotes should be coordinated internationally. In doing so, we address a question that was recently raised by Rogoff (2016, chapter 13).

We argue in Link (2019) that a central banker's problem of deciding when (if ever) to call in large notes involves only one major trade-off in a one-country setting: On the one hand, calling in large notes is beneficial because relaxing the ELB-constraint increases the ability to reach monetary policy objectives. On the other hand, as suggested by Rogoff (2015, pp. 450–452), cash demand is lower in a world without large banknotes. As a consequence of a lower cash demand, and thus of less cash in circulation, the seignorage profits a central bank makes by issuing cash are smaller.[1] Losing seignorage profits, in turn, can weaken a central bank's ability to stay independent from external financing and thus pose a threat to central bank independence.[2] So, a central banker must trade off the benefits from calling in large notes that come from a lower ELB-constraint against the costs through forgone seignorage revenues. We outline in Link (2019) why the net benefits from calling in large notes thus negatively depend on the natural rate of interest.

In addition to this trade-off that drives a central banker's behavior in the one-country setting of Link (2019), we consider international spillovers in a two-country setting in the present paper. Our starting point is the assumption that spillovers will occur if one central banker calls in large notes while the other central banker keeps on issuing large notes. In

[1] As in Link (2019), we use a narrow definition of the term "seignorage" and mean central bank revenues from issuing cash. For a discussion of various measures of seignorage see, for instance, Buiter (2007).

[2] See Rogoff (2015, p. 452), Rogoff (2016, chapter 6), Buiter (2009, p. 224), Thiele, Niepelt, Krüger, Seitz, Halver, and Michler (2015, p. 10), or Krüger and Seitz (2017, chapter 4.1).

particular, the central banker that keeps on issuing large notes will benefit from an increase in the demand for "her" currency and make higher seignorage profits. Rogoff (2016, chapter 13) has pointed out these thoughts. This assumption is natural because there is a specific demand for large banknotes as a store of value (see, for instance, Fischer, Köhler, and Seitz, 2004; Feige, 2012; Bartzsch, Seitz, and Setzer, 2015) and from the underground economy (see, for instance, Rogoff, 1998, 2016, chapter 13). If a near substitute for a discontinued large banknote exists in a different currency, hoarders or criminals will have the option to use a substitute after one central banker makes a unilateral calling-in move. This argument is especially strong for major reserve currencies like the euro or the U.S. dollar for which there is also high foreign demand.[3] With regard to the demand for large-denomination euro and U.S. dollar banknotes from the global shadow and underground economy, Rogoff (2016, chapter 13) predicts that such substitution effects would possibly occur.

We take Rogoff's presumption that the unilateral elimination of large notes will shift seignorage revenues between central banks and incorporate these kinds of spillovers into a two-country version of the optimal timing model of Link (2019). We explore the timing game between the two central bankers and show that inefficient timing decisions when central bankers do not cooperate create a need for policy coordination. Inefficient timing decisions can result from the fear, respectively from the prospect, of extra seignorage losses, respectively gains, when large notes of different currencies are called in sequentially. The point we make is that it is the substitutability of banknotes of different currencies and of different denominations that determines the points in time and the sequence of moves central bankers choose to call in large notes. We argue that what matters in this respect

[3] Especially for the euro and the U.S. dollar there is a large foreign demand not only from the underground economy but also as a store of value. According to the European Central Bank (2019), a large amount of euro currency is held outside the euro area. Precise figures are unknown but a rough indicator for the currency in circulation outside the euro area can be obtained by summing up the observed net shipments of banknotes to the rest of the world since the introduction of the euro. Concretely, the accumulated net "shipments of euro banknotes to destinations outside the euro area" amounted to approximately 170 billion euros as of February 2019 (see European Central Bank, 2019, p. 34). However, this rough indicator is rather a lower bound, as, for instance, outlined by Calza and Zaghini (2016, pp. 234–235). The true, unobserved amount of euro banknotes held outside the euro area should be significantly higher. That a large share of all euro banknotes in circulation is held abroad is also supported by Bartzsch, Rösl, and Seitz (2013) who estimate that a share of 45% of the euro banknotes issued by Germany circulates outside the euro area. The foreign demand for currency denominated in U.S. dollars is even higher. Even conservative estimates point toward a share of 30 to 37% of all U.S. currency that is held abroad (see Feige, 2012). Because the estimation of the stock of currency circulating abroad is a non-trivial task, former estimates of the share of U.S. currency holdings abroad even range up to 70% (see Feige (2012, p. 245) who presents an overview of several former studies that estimated the share of U.S. currency that is – or was at the time the respective study was conducted – held abroad).

is not only the substitutability of large notes of different currencies but also whether there is a substitute in a smaller denomination for a discontinued large note available to private hoarders or criminals. To illustrate this point, we discuss two scenarios: one in which central bankers face a prisoner's dilemma with one simultaneous move Nash equilibrium (in pure strategies); and one scenario in which they face a coordination problem with two sequential move equilibria. The sequential move equilibria exist even though we consider two fully symmetric central bankers.[4] The asymmetry that lies in the sequential move equilibria emerges in the scenario where large and small banknotes are not substitutable at all from the perspective of private hoarders or criminals. Taken all together, our analysis illustrates that central bankers can avoid inefficient timing decisions by cooperating and coordinating their calling-in moves.[5]

The next section introduces a two-country version of the optimal timing model proposed in Link (2019) with international spillovers. The model is solved for a spillover-free benchmark scenario in section 3 and for the full specification in section 4. Section 4.1 contains some remarks on the solution approach we choose. Section 4.2 analyzes the optimal behavior of central bankers in a setting with an appointed leader and an appointed follower. Section 4.3 analyzes the timing game between two fully symmetric central bankers. Policy implications are discussed in section 5. Section 6 contains some concluding remarks.

[4]In posing the research question of which conditions would imply simultaneous, respectively sequential, move equilibria, we were mainly inspired, in addition to Dixit and Pindyck (1994, pp. 309–315) and Fudenberg and Tirole (1985), by two further papers in an industrial organization context, namely Reinganum (1981) and Weeds (2002). Reinganum (1981) considers a timing game between two firms which have the option to adopt a new technology. She describes a state of the world where the game between two identical firms has two sequential move Nash equilibria (see Reinganum, 1981, pp. 398–399). Weeds (2002) considers a game between two firms which must choose the timing of making an investment in a research project under uncertainty, respectively. Weeds (2002) also explores whether the timing game between the firms has simultaneous or sequential move equilibria. Crucially, she describes a state of the world where strategic considerations in the absence of cooperation can be a reason for a firm to delay an investment (a major focus of Weeds (2002) lies on *"strategic delay"*).

[5]Our contribution thus tackles an issue that is outside the traditional monetary policy coordination literature. The typical focus in the monetary policy coordination literature lies on strategies for the instruments that are used to conduct rule-based monetary policy, e.g., a policy rate (see, for instance, Pappa, 2004, p. 764). In principle, the question that is addressed in this literature is whether the central banks' individual policy rules should incorporate international spillovers. It has been pointed out that, at least until the last decade, there was little need to coordinate rule-based monetary policy internationally (see, for instance, Taylor (2013, pp. 4–6) and the comments on Taylor's paper by Rogoff (2013); see also Engel (2016, p. 22)). However, some factors can make room for gains from the international coordination of monetary policy: For a short summary of such factors with remarks on the respective literature that addresses these factors see Liu and Pappa (2008, p. 2086). For a survey of the monetary policy coordination literature see Engel (2016).

2 A Two-Country Optimal Timing Model of Calling In Large-Denomination Banknotes

We build on a deterministic version of the optimal stopping model of calling in large-denomination banknotes proposed in Link (2019) and extend it to a two-country setting: There are two countries, *home* (H) and *foreign* (F), with two symmetric central bankers (one in the *home* and one in the *foreign* country). The two countries are symmetric as well with the exception of the currencies they issue. Legal tender in the *home* country are *"small"*- and *"large"*-denomination *home* currency banknotes, in the *foreign* country *"small"*- and *"large"*-denomination *foreign* currency banknotes.[6] The central bankers are in charge of "their" currencies' denominational structures and both have the option to call in their large-denomination banknotes. To call in large notes in the definition that we use here and in Link (2019) means to stop the issuance and to remove the legal tender status of these notes with immediate effect. As in Link (2019), a calling-in move is one-shot and irreversible (due to the high reputational costs of reintroducing a large denomination after having previously removed it). The central bankers can time their moves freely and call in their large notes independently from each other. Time is continuous and the points in time when the calling-in moves are made are denoted by $T^H \in [0, \infty)$ for the *home* and $T^F \in [0, \infty)$ for the *foreign* central banker's timing. Future values are discounted at the rate $\delta \in \mathbb{R}_{>0}$.

We use the same highly stylized approach as in Link (2019) to capture a central banker's costs and benefits associated with the calling-in of large notes. The spillover effects in the present two-country setting are introduced in this stylized way as well. Accordingly, we capture the net benefits, including the international spillovers central banker $i \in \{H, F\}$ has from calling in large notes by flow utilities U_t^i that are received in period $t \in [0, \infty)$. If

[6] In the context of the idea to relax the ELB by removing large banknotes, a similar formalization of a currency that is issued in two denominations is proposed in Rognlie (2016, pp. 41–42). However, Rognlie (2016) analyzes optimal monetary policy in a one-country New Keynesian framework. In contrast to our subject, Rognlie (2016) does not focus on the optimal timing of removing the large denomination. Instead, he considers household utility under optimal monetary policy dependent on the denominational structure of cash (see Rognlie, 2016, p. 42).

central banker i makes her calling-in move strictly before central banker $j \in \{H, F\}$ with $j \neq i$, her utility function is defined by $U_t^i = U_t^{first}$ with

$$U_t^{first} = \begin{cases} 0 & \text{for } t < T^{first}, \\ u_t + e^{\text{first}} & \text{for } t \in [T^{first}, T^{second}), \\ u_t & \text{for } t \geq T^{second}, \end{cases} \tag{1}$$

where T^{first}, respectively T^{second}, denotes the point in time when the first, respectively second, mover calls in large notes. u_t is an exogenous process that captures the beneficial or adverse effects of a central banker's *own* actions without regard to the behavior of her counterpart. If the two central bankers' behavior diverges, i.e., if they move sequentially, there will be an external effect e^{first} on the first and e^{second} on the second mover's utility. These externalities capture the spillovers from unilateral calling-in moves. We skip the details on u_t and the externalities for the moment and discuss them further below. If central banker i moves strictly after j, her utility function is defined by $U_t^i = U_t^{second}$ with

$$U_t^{second} = \begin{cases} 0 & \text{for } t < T^{first}, \\ e^{\text{second}} & \text{for } t \in [T^{first}, T^{second}), \\ u_t & \text{for } t \geq T^{second}. \end{cases} \tag{2}$$

And if central banker i moves at the same time as j does with T^{sim} denoting this point in time, her utility function is defined by $U_t^i = U_t^{sim}$ with

$$U_t^{sim} = \begin{cases} 0 & \text{for } t < T^{sim}, \\ u_t & \text{for } t \geq T^{sim}. \end{cases} \tag{3}$$

Note, that in this case there are no spillovers such that U_t^{sim} corresponds to the central banker's overall period-t utility function in the one-country setting in Link (2019).

Let us now specify the sub-utility function u_t that describes the effects of a central banker's own actions. This component of overall utility U_t^i describes the net benefits from calling in large notes, though excluding the effects of international spillovers. We provided the foundations of u_t in Link (2019) and introduced it as an exogenous process of flow utilities a central banker receives once she has made her calling-in move (in this regard, we follow Alvarez and Dixit (2014) who, in simplified terms, use exogenous flow utilities to describe a currency union's net benefits from having a common currency).[7] Accordingly,

$$u_t = g - r_t - \omega, \tag{4}$$

where $g \in \mathbb{R}_{>0}$ and $\omega \in \mathbb{R}_{>0}$ are known constants and r_t is an exogenous process which represents the (*home* and *foreign*) natural rate of interest in period t. In our symmetric setting we do not distinguish between the *home* and *foreign* natural rate of interest, they are identical for all $t \geq 0$. In contrast to Link (2019), we always assume perfect foresight over the future path of the natural rate of interest and define the natural rate as a (deterministic) function with

$$r_t = r \cdot \exp(-\theta t) + r^{ss} \cdot (1 - \exp(-\theta t)) \text{ for } t \geq 0, \tag{5}$$

where $r_0 = r \in \mathbb{R}$ is known, $r^{ss} \in \mathbb{R}$ is the known long-run steady-state level of the natural rate, and $\theta \in \mathbb{R}_{>0}$ captures the speed with which the natural rate reverts to its steady state. With this definition, we might as well say that the natural rate is governed by an Ornstein-Uhlenbeck process (OU process) that has no stochastic component.[8] We choose this specification of the natural rate to consider different states of the world with different paths of the natural rate. In particular, our focus is on such states where a central banker's optimal timing strategy to call in large notes involves a wait-and-see component.

Equation (4) that defines the sub-utility u_t states that the net benefits from calling in large notes are a negative function of the natural rate of interest. On the one hand, this is consistent with the assumption that the benefits from relaxing the ELB-constraint are larger the lower the natural rate is. On the other hand, (4) is consistent with the assumption that the costs of calling in large notes in the form of lost seignorage profits are increasing in the natural rate. We refer to our discussion in Link (2019, section 2) for the reasoning of these properties. Ultimately, it is the state-dependency of u_t on a mean-reverting natural rate that creates the need for a central banker to time her actions

[7] See Link (2019, section 2).

[8] For $t \geq 0$ and $r_0 = r$, r_t as defined in (5) describes the expected path of an Ornstein-Uhlenbeck process with $dr_t = \theta(r^{ss} - r_t)dt + \sigma dB_t$ where B_t is Brownian motion starting at $B_0 = 0$ and $\sigma \geq 0$ is a volatility parameter (see, for instance, Maller, Müller, and Szimayer (2009, p. 423) or Dixit and Pindyck (1994, pp. 74–78)). This stochastic specification is used in Link (2019).

optimally. The benefits and costs of calling in large notes that do not depend on the level of the natural rate are captured by the constants g and ω.

We believe our way of introducing an external effect e^{first} on the first and e^{second} on the second mover's utility is a convenient way to capture the seignorage losses of a first, and gains of a second, mover after a unilateral calling-in move. For the main part of our analysis we abstract from modeling cash in circulation and thus seignorage revenues explicitly. However, in appendix A, we set up a stylized network model of banknote demand and shifts of seignorage profits that result from calling in large notes. This model illustrates that all we need to describe seignorage effects within our framework is given by the externalities e^{first} and e^{second}. In particular, it illustrates the extent to which currency substitution can involve a disadvantage for a first and an advantage for a second mover. The crucial point is that the magnitude of the dis-/advantage depends on the degree of substitutability of banknotes of different denominations and currencies from the perspective of private hoarders or criminals. We argue in the appendix that the ratio of e^{first} to e^{second} can be related to this degree of substitutability. In this respect, we focus on two scenarios in the analysis that follows. We interpret the constellation $e^{second} > 0$ with $e^{first} = -e^{second}$ as describing a scenario where *home* and *foreign* as well as *small* and *large* banknotes are substitutable. And we interpret the constellation $e^{second} > 0$ with $e^{first} = 0$ as describing a state of the world where *small* and *large* banknotes are not substitutable at all while *home* and *foreign* banknotes of the same denomination are substitutes. Eventually, the central bankers' timing decisions and strategic considerations are crucially determined by these externalities.[9]

[9] Whether, respectively why, there will be shifts in banknote demand and in seignorage revenues in the real world is an empirical question which we leave unanswered. We just base our argumentation on a *possible* reason for such shifts. When considering the theoretical case that small and large notes can be "perfect substitutes", there are two issues that require a short remark (in the following, the term "perfect substitutes" shall be used for different banknote denominations if the total demand for all banknotes of these denominations in terms of value does not depend on the denominational structure [in this case, we might as well say that a discontinued denomination can be *perfectly replaced* by other notes] – accordingly, if a large note is called in while a "perfect substitute" for that note is available, the demand for the discontinued large note will *fully* shift to the remaining denominations; to be clear, this does not mean that banknote demanders consider different notes as being "perfect substitutes" in the proper meaning of the word while all denominations are still available): Firstly, if small and large notes were perfect substitutes from the perspective of *all* agents in the economy, calling in large notes would have no effect on the ELB. So, to be effective in terms of lowering the ELB, a calling-in move would require that there is at least one other type of agent in the economy having a demand for banknotes but having different capabilities to replace large by small notes. It is natural to assume that such a difference exists, for instance, between wholesale and private demanders of cash, like banks and private households (for a note with regard to the effect the elimination of large notes would have on *"ordinary retail transactions"*, respectively on the cash hoarding costs of *"large-scale financial institutions"*, see Rogoff, 2017, pp. 59–60). So, for instance, while a private household might be relatively indifferent between keeping, say, 10,000 euros in 50 200-euro notes or in 200 50-euro notes in a safe deposit box that has a volume large enough to store a multiple of that

The goal of a central banker $i \in \{H, F\}$ is to maximize the present value of the stream of period flow utilities U_t^i from $t = 0$ on by choosing an optimal timing strategy to call in her large-denomination banknotes. Since the stream of utilities depends not only on the central banker's own but also on her counterpart's actions, an optimal strategy is the equilibrium outcome of a timing game between the two central bankers. Before we explore this strategic setting, we solve a central banker's timing problem in the absence of international spillovers in the next section to provide a benchmark for the analysis that follows.

3 Optimal Timing in the Absence of International Spillovers

For the moment and for the remainder of this section, we ignore any international spillovers from the calling-in of large-denomination banknotes and assume that $e^{first} = e^{second} = 0$. The forces that drive a central banker's optimal timing strategy are thus free from strategic considerations so that we have a setting that is a useful benchmark for the analysis in the following sections. Essentially, the *home*, respectively *foreign*, central banker's problem is thus reduced to the deterministic case of the timing problem in the one-country setting that is analyzed in detail in Link (2019). This section reviews the solution of this problem as described in Link (2019) and briefly summarizes the main results there.

In the absence of international spillovers, the period flow utility of central banker $i \in \{H, F\}$ is

$$U_t^i = \begin{cases} 0 & \text{for } t < T^i, \\ u_t^i & \text{for } t \geq T^i, \end{cases} \tag{6}$$

where $u_t^i = g - r_t - \omega$ and T^i denotes the point in time i makes her calling-in move. Since central bankers are symmetric, we drop all superscripts in the remainder of this section

amount, a bank might consider the costs associated with the hoarding of several billion euros in cash to be prohibitively high if only small banknotes were available (as noted above, for a discussion of hoarding costs of *"large-scale financial institutions"* in a world where only small banknotes exist see Rogoff, 2017, pp. 59–60). The second issue in a world where discontinued large notes can be perfectly replaced by small notes that requires a remark is related to a central banker's seignorage revenues: Our argumentation in the appendix for the case that large and small notes are substitutable is based on the assumption that cash demanders have a preference to split cash holdings equally in terms of value across the denominations that are available to them. Crucially, the denominations that are available include banknotes of *both* currencies. So, the *home* central banker's calling-in move will imply a shift in banknote demand from large *home* to small *foreign* notes which, in turn, implies an additional loss of seignorage revenues for the *home* central banker.

for ease of notation. The present value of the stream of flow utilities from $t = 0$ on is $\int_0^\infty U_t \cdot \exp(-\delta t)dt$ (recall that a central banker discounts future utility with rate δ). A central banker aims to maximize this present value by timing her calling-in move optimally. With equation (5) for the path of the natural rate and with $r_0 = r$, her objective function F, like in Link (2019), is

$$\begin{aligned} F(r,T) &= \int_T^\infty (g - r_t - \omega) \cdot \exp(-\delta t)dt \\ &= \frac{1}{\delta} \cdot (g - r^{ss} - \omega) \cdot \exp(-\delta T) \; - \; \frac{1}{\delta + \theta} \cdot (r - r^{ss}) \cdot \exp(-(\delta + \theta)T), \end{aligned} \tag{7}$$

with the associated maximization problem

$$\max_{T \geq 0} F(r,T). \tag{8}$$

The optimal timing is denoted by T^* with $T^* \in [0, \infty) \cup \{\infty\}$ where $T^* = \infty$ means that the calling-in move is never made. Our main point in the deterministic case in Link (2019) was that any optimal timing strategy and in particular whether T^* is an interior or a corner solution of the central banker's decision problem is determined by the specific path of the natural rate and thus by the specific path of period utility u_t (note that the process u_t features mean-reversion, too, since it is only determined by the mean-reverting process r_t and the two constants g and ω). Since we rule out any stochastic movements of the natural rate, its future path is a monotone function of time that reverts to its steady state r^{ss} from either above or below. The same holds for the path of utility u_t. As a consequence, and with u_t being defined as in (4), there is a critical threshold such that making a calling-in move at a natural rate below (above) that threshold yields a positive (negative) payoff with the "payoff" being defined as the present value of the stream of flow utilities from calling in large notes. As in Link (2019), we refer to this critical threshold as "break-even threshold" $\hat{r}$ and define it as the natural rate level at which an instantaneous calling-in move has a payoff of exactly zero. $\hat{r}$ is implicitly defined by $F(r = \hat{r}, T = 0) = 0$ with

$$\hat{r} = g - \omega + \frac{\theta}{\delta} u^{ss}. \tag{9}$$

We show in Link (2019) that there is a set of paths of the natural rate for which optimality requires a central banker to "wait" and choose a strictly positive T^* even if the natural rate has already fallen below the break-even threshold $\hat{r}$, i.e., even if making the calling-in move "somewhat" earlier than at T^* would already lead to a positive payoff. This set of paths is a subset of all paths of such period utility processes u_t that have a positive long-run steady state $u^{ss} > 0$ where $u^{ss} := g - r^{ss} - \omega$. To see this, we consider two scenarios in the following – one where the long-run benefits from calling in large notes are positive ($u^{ss} > 0$) and one where a central banker, in the long run, incurs losses from calling in large notes ($u^{ss} < 0$).

Scenario I ($u^{ss} > 0$): Let us first consider a world where $u^{ss} > 0$ and where the calling-in of large-denomination banknotes is thus a beneficial strategy in the long run. The first-order condition for an interior maximum of (8) is

$$\frac{\partial F(r,T)}{\partial T} = -(g - r^{ss} - \omega) \cdot \exp(-\delta T) + (r - r^{ss}) \cdot \exp(-(\delta + \theta)T) \stackrel{!}{=} 0, \tag{10}$$

which implies that

$$T^* = \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega}\right), \tag{11}$$

where $T^* > 0$ for all $r > g - \omega$ given that $u^{ss} = g - r^{ss} - \omega > 0$. We show in Link (2019) that the second-order condition can be used to show that (11) maximizes (8) in the specified region of the parameter space. In addition, equation (7) can be used to show that this maximum is strictly greater than zero. By plugging in T^* into (5) it is easy to see that the natural rate at time $T = T^*$ is $r_{T^*} = g - \omega$. So, the timing strategy to make the calling-in move at T^* can also be formulated as a simple rule to make the move once the natural rate has hit or fallen below a specific threshold. Like in Link (2019), we refer to this level as the "optimal threshold" and denote it by $\underline{r}$. In case the utility has a positive steady state the optimal threshold to make the calling-in move is thus $\underline{r} = g - \omega$.[10]

Comparing $\underline{r}$ with the break-even threshold $\hat{r}$ illustrates the range for the natural rate where optimality requires the central banker to wait, although an instantaneous calling-in move would have a positive payoff. This natural rate range where waiting is optimal is $(\underline{r}, \hat{r}]$ which is obviously broadening in u^{ss}. The interpretation of these properties is straightforward and starts with the path of period utility conditional on $u^{ss} > 0 \Leftrightarrow$

[10]Note that (8) has a corner solution at $T = 0$ for all $r < g - \omega$ in case $u^{ss} > 0$. This can easily be seen by recalling the constraint $T \geq 0$ and considering the logarithmic term in (11). For a discussion of such scenarios see Link (2019).

$g - \omega > r^{ss}$. The future path of a natural rate that is initially above its steady state, i.e., $r_0 = r > r^{ss}$, is monotonically decreasing. In turn, the path of period utility is monotonically increasing in that case with $u_t < 0$ for all $r_t > g - \omega$ and with $u_t > 0$ for all $r_t < g - \omega$. So, a central banker with the flexibility to time her calling-in move freely chooses to wait until the period utility from calling in large notes becomes greater than zero – which is the case at the moment r_t hits the optimal threshold $\underline{r}$. As a consequence, a strictly positive payoff in terms of cumulative utility can be realized. In contrast, a calling-in move that is made at a natural rate above $\underline{r}$ involves an initial phase with negative period utilities that partially or fully offset the cumulative benefits of the second phase with positive period utilities once the natural rate has fallen to sufficiently low levels. These cumulative benefits of the second phase are fully offset if the move is made at $r = \hat{r}$ and partially offset if it is made at $r \in (\underline{r}, \hat{r})$.

Scenario II ($u^{ss} < 0$): Let us now consider a central banker's optimal behavior in case the utility has a negative steady state, i.e. if $u^{ss} \leq 0$. We point out in Link (2019) that there is no interior maximum of (8) in this case. Dependent on the constellation of parameters, the move is optimally made either never or immediately at $T = 0$. The sole decision criterion is whether the present value of the stream of flow utilities from making the calling-in move immediately is positive – or, put another way, whether the natural rate is below the break-even threshold $\hat{r}$. So, if the utility has a negative steady state, the optimal threshold $\underline{r}$ corresponds to the break-even threshold $\hat{r} = g - \omega + \frac{\theta}{\delta} u^{ss}$. In such a scenario with $r_0 < \hat{r} < u^{ss}$ where the natural rate is monotonically increasing toward its steady state and where period utility is thus monotonically decreasing in time, any deferral to make the calling-in move later than at $T = 0$ would involve a loss of benefits and is thus not an optimal timing strategy.

To sum up, the optimal behavior of a central banker in the absence of international spillovers can be formulated as a simple decision rule that is dependent on the steady state of period utility. The rule is to make the move if the natural rate hits or is below $\underline{r}$ with

$$\underline{r} = \begin{cases} \hat{r} = g - \omega + \frac{\theta}{\delta} \cdot u^{ss}, & \text{if } u^{ss} \leq 0 \\ g - \omega, & \text{if } u^{ss} > 0. \end{cases} \tag{12}$$

4 Optimal Timing in the Presence of International Spillovers

4.1 Solution Approach and States of the World

In this section, we explore the set of equilibria of the timing game between two fully symmetric central bankers in the presence of international spillovers from unilateral calling-in moves. In particular, our goal is to show that the substitutability of banknotes of different denominations and currencies determines the number of Nash equilibria and – crucially – whether central bankers' equilibrium strategies are to make their calling-in moves simultaneously or sequentially even though both central bankers are fully symmetric. As a preliminary step before we consider the game between two symmetric central bankers, we remove symmetry and appoint one central banker as leader, the other as follower. The leader can freely choose the timing to call in large notes, the follower must wait for the leader to make the first move to call in her large notes. Since we restrict our analysis to states of the world where, at some point, central bankers will make their calling-in moves, we call the leader "first mover" and the follower "second mover." Before symmetry is restored in section 4.3, section 4.2 analyzes the central bankers' timing decisions in the case of an appointed first and second mover. We start by analyzing the second mover's optimal timing.[11]

We restrict our analysis for the rest of this paper to the interesting case where, in the short run, that is, in period $t = 0$ and for some time thereafter, seignorage losses after a calling-in move outweigh the benefits from a relaxed ELB-constraint such that the central bankers' optimal timing strategy has a wait-and-see component. From a $t = 0$-perspective this means that no central banker would make her calling-in move immediately at $t = 0$. The two conditions that are necessary to guarantee such an outcome formally are that the net benefits from calling in large notes are negative in the short run ($r_0 = r > g - \omega$) but positive in the long run ($u^{ss} > 0$) which means that, eventually, ELB-issues will outweigh seignorage considerations. In addition to these assumptions, we only consider states of the world that have the following characteristics: (1) The external effect of the first mover's

[11] An earlier version of our model accounted for uncertainty over the future path of the natural rate such that the game between the two central bankers would become a stochastic game. To solve this earlier version of our model, we made use of the solution techniques for stochastic games described by Dixit and Pindyck (1994). In particular, we took the idea to start with a leader-follower setting and to analyze the follower's decision problem first from Dixit and Pindyck (1994, pp. 309–315).

unilateral calling-in move on the second mover's utility is always greater than zero, i.e., $e^{second} > 0$. This describes a second mover's free ride in the form of additional seignorage revenues that result from currency substitution if she keeps on issuing large notes while her counterpart has removed them. (2) The external effect on the first mover's utility of the second mover's decision not to call in large notes together with the first mover is always less than or equal to zero, i.e., $e^{first} \leq 0$. So, there is no specific advantage for the first mover from being able to move first. (3) We also require that $e^{first} \in [-e^{second}, 0]$ which just describes that the first mover's extra seignorage losses until also the second mover has called in large notes are not larger (in absolute value) than the second mover's additional seignorage revenues. (4) To rule out that the second mover has an incentive to free ride forever, we let $e^{second} < u^{ss}$.

4.2 Timing in a Setting with an Appointed First and Second Mover

4.2.1 Optimal Timing of the Second Mover

Let us now assume that one central banker has the flexibility to time her calling-in move freely at $T^{first} \geq 0$ with the sole restriction of having to move first. The other central banker is forced to wait until her counterpart has moved but is then able to time her move freely at $T^{second} \geq T^{first}$ (this is a strong assumption since it means that the second mover can immediately observe the actions of her counterpart and move without delay at T^{first}). We take a $t = 0$-perspective and consider the second mover's timing problem first.

The second mover's objective function, given that the first central banker moves at T^{first} and given that the natural rate at time $t = 0$ is $r_0 = r$, is

$$\begin{aligned} F^{second}(r, T^{first}, T^{second}) &= \int_{T^{first}}^{T^{second}} e^{second} \cdot \exp(-\delta t) dt \\ &+ \int_{T^{second}}^{\infty} (g - r_t - \omega) \cdot \exp(-\delta t) dt \\ &= \frac{1}{\delta} e^{second} \cdot \Big(\exp(-\delta T^{first}) - \exp(-\delta T^{second}) \Big) \\ &+ \frac{1}{\delta}(g - r^{ss} - \omega) \cdot \exp(-\delta T^{second}) \\ &- \frac{1}{\delta + \theta}(r - r^{ss}) \cdot \exp(-(\delta + \theta) T^{second}). \end{aligned} \tag{13}$$

In deciding on the optimal timing of her calling-in move, the second-moving central banker must take into account that the additional seignorage revenues (captured by the external effect $e^{second} > 0$) she receives after the first mover has made a calling-in move are lost at the instant she makes her own move. We have formalized this property in equation (2). The additional seignorage revenues during the period from T^{first} to T^{second} create an opportunity cost of calling in large notes. This opportunity cost makes the second mover reluctant to make her own move. Comparing (13) with the objective function (7) in the spillover-free benchmark illustrates this property. In the presence of spillovers, i.e., if $e^{second} \neq 0$, an additional term enters the objective function. This term, $\int_{T^{first}}^{T^{second}} e^{second} \cdot \exp(-\delta t)dt$, that captures the present value of the stream of flow externalities from T^{first} until T^{second} is smaller the sooner the second-moving central banker makes her calling-in move. The second mover thus faces a trade-off between receiving an externality in the form of additional seignorage revenues or a utility u_t from making her own calling-in move.

The assumptions we have made in section 4.1 and the state of the world we have defined for the rest of the paper guarantee that this trade-off is resolved in favor of a calling-in move at some finite T^{second^*} so that the second mover's optimization problem

$$\max_{T^{second} \geq T^{first}} F^{second}(r, T^{first}, T^{second}) \tag{14}$$

has an interior maximum. The first-order condition for an interior solution of (14) is

$$(g - r^{ss} - \omega - e^{second}) \cdot \exp(-\delta T^{second}) \;=\; (r - r^{ss}) \cdot \exp(-(\delta + \theta)T^{second}) \tag{15}$$

and yields a unique interior maximum at

$$T^{second^*} \;=\; \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega - e^{second}}\right), \tag{16}$$

given that the first central banker has already moved at some $T^{first} \leq \frac{1}{\theta} \cdot \ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega-e^{second}}\right)$, otherwise $T^{second^*} = T^{first}$.[12]

Comparing the optimal timing T^{second^*} as stated in (16) with the optimal timing in the spillover-free benchmark stated in equation (11) shows that the additional seignorage revenues (described by e^{second}) make the second mover more reluctant to make her own

[12] It is easily checked that the second-order condition for an interior maximum at T^{second^*} is satisfied since $\frac{\partial^2 F^{second}}{(\partial T^{second})^2} = \delta(g - r^{ss} - \omega - e^{second}) \exp(-\delta T^{second}) - (\delta+\theta)(r - r^{ss}) \exp(-(\delta+\theta)T^{second})$ evaluated at T^{second^*} as stated in (16) is less than zero if $u^{ss} > 0$, $e^{second} \in [0, u^{ss})$, and $r > g - \omega > r^{ss}$. The uniqueness of the maximum is implied by the monotonicity of the path of the natural rate of interest and the fact that at some point period utility u_t stays greater than e^{second}.

move with $T^{second^*} > T^*$ and with T^{second^*} increasing in e^{second}. A crucial feature is that T^{second^*} does not depend on the first mover's timing if it is an interior solution of (14). It is only determined by u_t and e^{second} and marks the period where the flow of period utilities from making a calling-in move starts to exceed the flow of externalities. At T^{second^*}, the natural rate starting at $r_0 = r$ has fallen to the level $r_{T^{second^*}} = g - \omega - e^{second}$ so that $u_{T^{second^*}} = e^{second}$ and that receiving a flow of utilities u_t becomes an alternative that is more attractive than receiving a flow of externalities e^{second} for all $t > T^{second^*}$. Beyond this point, the benefits from calling in large notes and relaxing the ELB-constraint exceed the costs in the form of forgone seignorage revenues when the central banker stops issuing large notes. It is clear that the optimal threshold for the natural rate to make the calling-in move as a second mover is thus

$$\underline{r}^{second} = g - \omega - e^{second}, \tag{17}$$

which is lower than the optimal threshold $\underline{r} = g - \omega$ in the spillover-free benchmark.

4.2.2 Optimal Timing of the First Mover

The first mover's objective function given that the second moving central banker makes the calling-in move at $T^{second} \geq T^{first}$ is

$$\begin{aligned} F^{first}(r, T^{first}, T^{second}) &= \int_{T^{first}}^{T^{second}} e^{first} \cdot \exp(-\delta t)dt \\ &+ \int_{T^{first}}^{\infty} (g - r_t - \omega) \cdot \exp(-\delta t)dt \\ &= \frac{1}{\delta} e^{first} \cdot \Big(\exp(-\delta T^{first}) - \exp(-\delta T^{second})\Big) \\ &+ \frac{1}{\delta}(g - r^{ss} - \omega) \cdot \exp(-\delta T^{first}) \\ &- \frac{1}{\delta + \theta}(r - r^{ss}) \cdot \exp(-(\delta + \theta)T^{first}). \end{aligned} \tag{18}$$

We have already specified that $e^{first} \in [-e^{second}, 0]$ (see section 4.1). e^{first} being strictly smaller than zero describes states of the world where a first mover loses more seignorage revenues if both central bankers move sequentially compared to the case where both move simultaneously. We illustrate such a situation in appendix A and argue that the root of a first mover's disadvantage is the substitutability of small and large banknotes of different currencies. If small and large notes are substitutable, a unilateral calling-in move at T^{first} can imply that a share of the demand for the first mover's large notes shifts to the second mover's small notes. The first term in the objective function (18) describes the first mover's loss of utility from additional seignorage losses if T^{first} and T^{second} diverge.

The first mover can anticipate the second mover's optimal timing T^{second^*}, conditional on her own actions, and solve the optimization problem

$$\max_{T^{first}\geq 0} F^{first}(r, T^{first}, T^{second^*}) \tag{19}$$

by choosing an optimal timing T^{first^*}. The assumptions made in section 4.1 imply that T^{first^*} is an interior solution of (19).[13] The first-order condition for an interior solution of (19) is

$$(g - r^{ss} - \omega + e^{first}) \cdot \exp(-\delta T^{first}) \;=\; (r - r^{ss}) \cdot \exp(-(\delta + \theta)T^{first}), \tag{20}$$

and yields

$$T^{first^*} \;=\; \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega + e^{first}}\right). \tag{21}$$

The additional seignorage losses the first central banker incurs as a first mover (captured by $e^{first} \leq 0$) introduce an additional cost of calling in large banknotes. This cost induces the first mover to defer her calling-in move and choose a point in time that lies beyond the optimal timing in the spillover-free benchmark (i.e., $T^{first^*} \geq T^*$).[14] This optimal timing strategy, in turn, is reflected in the natural rate threshold where the calling-in move is optimally made. The optimal threshold for the first mover is

$$\underline{r}^{first} = g - \omega + e^{first}, \tag{22}$$

and is thus lower than the optimal threshold in the spillover-free benchmark. Once the natural rate has fallen to that level, the first-moving central banker's period utility from making the move, u_t, exceeds the period disadvantage from being a first mover, i.e., $u_t = g - r_t - \omega > |e^{first}|$ for all $r_t < g - \omega + e^{first}$. This is the point in time when the

[13] $e^{first} \in [-e^{second}, 0]$ also implies that $T^{first^*} \leq T^{second^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega - e^{second}}\right)$.

[14] It is easily checked that the second-order condition for an interior maximum at T^{first^*} is satisfied since $\frac{\partial^2 F^{first}}{(\partial T^{first})^2} = \delta(g - r^{ss} - \omega + e^{first})\exp(-\delta T^{first}) - (\delta + \theta)(r - r^{ss})\exp(-(\delta + \theta)T^{first})$ evaluated at T^{first^*} as stated in (21) is less than zero if $u^{ss} > 0$, $e^{first} \in (-u^{ss}, 0]$, and $r > g - \omega > r^{ss}$.

benefits from calling in large notes and relaxing the ELB-constraint exceed the costs in the form of forgone seignorage revenues.

4.2.3 Non-cooperative Outcome

In the following, we explore the outcome of the timing game between the appointed first- and second-moving central banker if they do not cooperate. We consider two different scenarios to show how the substitutability of banknotes of different denominations and currencies determines whether the central bankers move simultaneously or sequentially and whether cooperation could make Pareto improvements possible.[15] The state of the world in both scenarios is as defined in section 4.1: We consider a world where the natural rate at time $t = 0$ is still at a relatively high level, i.e., $r_0 = r > g - \omega > r^{ss}$. This high level implies that at time $t = 0$ and for some time thereafter, the costs of calling in large notes in the form of lost seignorage revenues will exceed the benefits from relaxing the ELB-constraint. So, optimality will require the central bankers to defer their calling-in moves until the natural rate has fallen to a sufficiently low level.

The optimal choice of times by the second- and first-moving central banker is stated by equations (16) and (21) with $T^{second^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega-e^{second}}\right)$ and $T^{first^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega+e^{first}}\right)$. To analyze whether central bankers would be better off cooperating and coordinating the timings of their calling-in moves, we compare the first and second mover's payoff functions subject to the first central banker moving at T^{first} for some $T^{first} \geq 0$ given that the second mover behaves optimally, as well as both central bankers' payoff functions if they move simultaneously at some $T^{first} = T^{second} > 0$.[16] By comparing the respective payoff functions we can state whether cooperation can lead to Pareto improvements. So, the first mover's payoff if she moves at T^{first}, given that the second central banker behaves optimally and moves at T^{second^*}, is the present value of the stream of period flow utilities

$$F^{first}(r, T^{first}, T^{second^*}) = \int_{T^{first}}^{T^{second^*}} e^{first} \cdot \exp(-\delta t)dt + \int_{T^{first}}^{\infty} (g - r_t - \omega) \cdot \exp(-\delta t)dt \quad (23)$$

[15] We will not discuss a collective's decision problem of the kind $\max(F^{first} + F^{second})$ with an objective function of the collective since this would lead us to issues of transfer payments or compensations between the two central bankers.

[16] In methodical regards, in analyzing the respective payoff functions of a leader, a follower, and of simultaneous movers, we drew on Fudenberg and Tirole (1985, pp. 386–389).

$$
\begin{aligned}
&= \frac{1}{\delta} e^{first} \cdot \Big(\exp(-\delta T^{first}) - \exp(-\delta T^{second^*}) \Big) \\
&+ \frac{1}{\delta}(g - r^{ss} - \omega) \cdot \exp(-\delta T^{first}) \\
&- \frac{1}{\delta + \theta}(r - r^{ss}) \cdot \exp(-(\delta+\theta) T^{first})
\end{aligned}
$$

(note, that $T^{second^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega - e^{second}}\right)$ if $T^{first} \leq \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega - e^{second}}\right)$ and that $T^{second^*} = T^{first}$ otherwise).[17] The second mover's payoff as a function of the first central banker moving at T^{first}, given that the second mover behaves optimally, is

$$
\begin{aligned}
F^{second}(r, T^{first}, T^{second^*}) &= \int_{T^{first}}^{T^{second^*}} e^{second} \cdot \exp(-\delta t) dt \\
&+ \int_{T^{second^*}}^{\infty} (g - r_t - \omega) \cdot \exp(-\delta t) dt \qquad (24) \\
&= \frac{1}{\delta} e^{second} \cdot \Big(\exp(-\delta T^{first}) - \exp(-\delta T^{second^*}) \Big) \\
&+ \frac{1}{\delta}(g - r^{ss} - \omega) \cdot \exp(-\delta T^{second^*}) \\
&- \frac{1}{\delta + \theta}(r - r^{ss}) \cdot \exp(-(\delta+\theta) T^{second^*}),
\end{aligned}
$$

and each central banker's payoff as a function of T^{first} if both agree and stick to moving simultaneously at $T^{first} = T^{second} \geq 0$ is

$$
\begin{aligned}
F^{sim}(r, T^{first}, T^{second} = T^{first}) &= \int_{T^{first}}^{\infty} (g - r_t - \omega) \cdot \exp(-\delta t) dt \qquad (25) \\
&= \frac{1}{\delta} \cdot (g - r^{ss} - \omega) \cdot \exp(-\delta T^{first}) \\
&- \frac{1}{\delta + \theta} \cdot (r - r^{ss}) \cdot \exp(-(\delta+\theta) T^{first})
\end{aligned}
$$

(note, that F^{sim} just corresponds to the payoff function F in the spillover-free benchmark). Figure 1 illustrates the three payoff functions for specific parameter values (the idea of this kind of illustration is taken from Fudenberg and Tirole, 1985, p. 387).

Let us now explore the two scenarios. The scenarios we consider correspond to the two scenarios described in appendix A. Of course, an endless number of alternative scenarios is possible. But since our main goal is to explain the idea of how the substitutability of different banknotes determines the outcome of the central bankers' timing game, we focus on two cases: Scenario I describes a state of the world where banknotes of different

[17] By assuming that $e^{first} \in [-e^{second}, 0]$ we make sure that optimality requires the first-moving central banker to choose $T^{first^*} \leq \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega - e^{second}}\right)$.

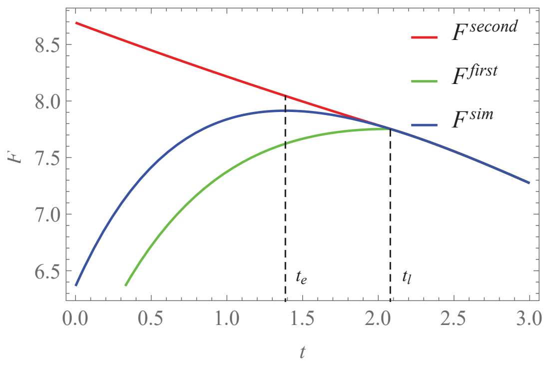


Figure 1: Central bankers' payoff functions as defined by (23), (24), and (25) for $r = 4$, $r^{ss} = 0$, $\theta = 1$, $\delta = 0.1$, $g = 1$, $\omega = 0$, $e^{first} = -0.5$, and $e^{second} = 0.5$. t_e corresponds to the optimal timing in the spillover-free benchmark. If $e^{first} = 0$, t_e thus also corresponds to the optimal timing of the first mover. t_l corresponds to the optimal timing of the second mover in the presence of spillovers. If $e^{first} = -e^{second}$, t_l also corresponds to the optimal timing of the first mover. If $e^{first} = 0$, the first mover's payoff function will correspond to the payoff function of simultaneous movers (and thus to the payoff function in the frictionless benchmark) and will thus be represented by the blue curve.

denominations and currencies are substitutes from the perspective of private hoarders or criminals. We show in appendix A that $e^{second} > 0$ together with $e^{first} = -e^{second}$ describes such a world.[18] The crucial point in this scenario is that the first mover's seignorage losses are larger if she decides to move strictly before instead of with the second central banker. In the appendix, we explained this larger loss in seignorage revenues by a shift in the demand for the first mover's discontinued large notes at T^{first} to the second mover's small notes. If we assume that demand shifts are symmetric, the shift of banknote demand and thus of seignorage from the leader to the follower would be balanced by a symmetric shift from the follower's demand for large notes to the leader's demand for small notes if both central bankers move simultaneously (this is also our explanation for why an external effect e^{first} or e^{second} only arises when T^{first} and T^{second} diverge).

Scenario II describes a state of the world where small and large banknotes are not substitutable at all whereas home and foreign notes of the same denomination are substitutes from the perspective of private hoarders or criminals. In the appendix, we argue that such a scenario can be described by $e^{first} = 0$ together with $e^{second} > 0$ if the demand for both central bankers' small notes remains completely unaffected by the removal of large notes at T^{first} and T^{second} and if the demand for the first mover's discontinued large notes com-

[18]Since we want to avoid making our argumentation overly complex, we introduce an external third country in the appendix and only consider cash demand by private hoarders and criminals in this third country.

pletely shifts to the second mover's large notes between T^{first} and T^{second} (see appendix A).

Scenario I (cross-currency and cross-denomination substitutability such that $e^{second} > 0$ and $e^{first} = -e^{second}$): If the first mover and the second mover are equally affected by spillovers in absolute value, i.e., if $e^{first} = -e^{second}$, they move simultaneously at $T^{first^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega+e^{first}}\right) = T^{second^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega-e^{second}}\right)$. Between T^{first} and T^{second} (if the two dates were to diverge), the second mover's period opportunity costs of calling in large banknotes were just equal to the first mover's period loss she incurred as a first mover. In concrete terms, this means that the first mover's additional loss in seignorage revenues after having made an unilateral calling-in move would correspond exactly to the second mover's additional seignorage revenues she would have from deferring her calling-in move. So, the first mover's disadvantage from moving first would be of the same magnitude as the second mover's advantage from moving second. Consequently, both central bankers require the period utility u_t from calling in their large notes to be at the level $u_t = -e^{first} = e^{second}$ to be compensated for the loss, respectively opportunity cost, and to be willing to make a move. As a result, both defer their moves until the natural rate hits the level $\underline{r}^{first} = \underline{r}^{second} = g - \omega + e^{first} = g - \omega - e^{second}$ and then move simultaneously.

The lack of cooperation results in a Pareto inefficient delay by both central bankers. The first central banker defers her move to avoid the disadvantage from moving first, which would arise as a consequence of the second central banker delaying her move and enjoying the free ride in the form of additional seignorage revenues between T^{first} and T^{second} if the two dates did diverge. So, as stated above, both central bankers move at $T^{first^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega+e^{first}}\right) = \frac{1}{\theta} \cdot \ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega-e^{second}}\right) = T^{second^*}$ which is later than in the spillover-free benchmark where the calling-in move is optimally made at $T^* = \frac{1}{\theta} \cdot \ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega}\right)$. The point is that now, due to the simultaneity of the moves, neither of the central bankers is hit by spillovers at all. So, both receive the payoff $F^{sim}(r, T^{first^*}, T^{second^*})$ which is just the payoff in the spillover-free benchmark $F(r, T = T^{first^*} = T^{second^*})$. By delaying her own move the first central banker can avoid negative spillovers and ultimately prevent the second central banker from enjoying a free ride. But as a consequence, both central bankers' payoffs are equal to $F^{sim}(r, T^{first^*}, T^{second^*})$ and are thus lower than the maximum payoff in the benchmark case $F(r, T^*)$. Both central bankers are worse off compared to moving simultaneously at T^* which means that cooperation would make a Pareto improvement possible.

Figure 1 illustrates this problem. The central bankers' payoffs from making their moves at t_l which corresponds to $T^{first^*} = T^{second^*}$ in this scenario, i.e., their actual payoffs without cooperation, are lower than the potential payoffs from moving simultaneously at any $t \in [t_e, t_l)$. So, while choosing t_e which is the optimal timing in the spillover-free case, T^*, leads to the maximum payoff for simultaneously moving central bankers, cooperating and moving simultaneously at any $t \in [t_e, t_l)$ already leads to a Pareto improvement compared to the non-cooperative outcome. In section 4.3, we argue that fully symmetric central bankers will face a prisoner's dilemma in a corresponding timing game if banknotes of different denominations and currencies are substitutes.

Scenario II (cross-currency substitutability but no cross-denomination substitutability such that $e^{second} > 0$ and $e^{first} = 0$): In the appendix we argue that if small and large notes are not substitutable at all, the first-moving central banker will have no disadvantage from moving first compared to the spillover-free benchmark and $e^{first} = 0$. However, we also argue that if banknotes of different currencies but of the same denomination are substitutes from the perspective of private hoarders or criminals, there will be an advantage for the second-moving central banker. Concretely, the second mover will benefit from a shift of the demand for the leader's discontinued large notes to the demand for her own large notes and make higher seignorage profits. This benefit is reflected in $e^{second} > 0$. So, if $e^{first} = 0$ and $e^{second} > 0$, the two central bankers will move sequentially at $T^{first} < T^{second}$ (see equations (16) and (21)). Without a specific disadvantage from moving first, the first central banker will choose the same timing as in the spillover-free benchmark and make the calling-in move already at a relatively high natural rate $\underline{r}^{first} = \underline{r} = g - \omega$. The first mover's optimal threshold is thus higher than the second mover's optimal threshold $\underline{r}^{second} = g - \omega - e^{second}$.

In this scenario, with the first and second mover being determined exogenously, there is no need for cooperation between the two central bankers. They move sequentially at $T^{first^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega}\right)$ respectively at $T^{second^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega - e^{second}}\right)$ and are unable to achieve a Pareto improvement by changing their timing. In the absence of a spillover for the first mover ($e^{first} = 0$) she has no disadvantage specifically from moving first so that her payoff at T^{first^*} equals the maximum payoff in the spillover-free benchmark.

The first central banker would always be worse off by deviating from this optimal timing whereas the second central banker would indeed profit from the first mover calling in large notes even earlier. Figure 1 illustrates this outcome: If $e^{first} = 0$, the first mover's payoff function is illustrated by the blue line and thus corresponds to the payoff function of simultaneous movers, respectively to the payoff function in the spillover-free benchmark. While the first mover maximizes her payoff by choosing $T^{first^*} = t_e$, the second mover's payoff (the red line) is increasing in the length of the period between T^{first} and T^{second^*} and could thus be maximized by the first mover calling in large notes at $t = 0$ (given that the second mover behaves optimally and moves at T^{second^*} which in figure 1 corresponds to t_l).

This scenario hints at the coordination problem that two fully symmetric central bankers face: If both central bankers behave optimally, the increased demand for the second mover's large banknotes during the period between T^{first} and T^{second} will imply that the second mover receives a higher payoff than the first mover does. Figure 1 shows that the second mover's payoff given that the first mover chooses $T^{first^*} = t_e$ and given that the second mover behaves optimally and chooses $T^{second^*} = t_l$ is greater than the first mover's payoff for this timing profile. If only one of the central bankers deviates from this timing profile, at least one of them will be worse off. If the first mover chooses to move at t_l, both central bankers will receive a smaller payoff while only the follower will receives a smaller payoff if she moves early with the leader at t_e. We discuss the coordination problem in a corresponding timing game between two symmetric central bankers in the next section.

4.3 Timing in a Setting with Symmetric Central Bankers

4.3.1 Rules of the Game

We now restore symmetry between the *home* and *foreign* central banker and drop the assumption made in section 4.2 of the first and second mover being exogenously appointed. The state of the world and the scenarios we consider in the following remain the same as considered in section 4.2 and defined in section 4.1. In particular, we assume that $r_0 = r > g - \omega$ (short-run losses will be incurred if a calling-in move is made shortly after $t = 0$), $u^{ss} > 0$ (net benefits from calling in large notes will be positive in the long run), and $e^{second} < u^{ss}$ (no eternal free ride is possible). Our goal is to explore the equilibria of timing games between two symmetric central bankers where the order of the two central bankers' calling-in moves can evolve endogenously.

If symmetric central bankers can choose in any period $t \geq 0$ whether to make their calling-in move immediately or wait and defer that decision to a later point in time they play a dynamic game in continuous time. However, and with regard to the European Central Bank (ECB) that communicated its decision to stop the issuance of the 500-euro note almost three years before it was definitely implemented in 2019, we add to the central bankers' game the possibility to make a commitment to future actions. This additional model ingredient simplifies the analysis and reduces the dynamic game to a static game.[19] The commitment is made at a public "press conference" by announcing the timing of a planned calling-in move. It is credible and the central bankers stick to the commitment as the reputational costs of deviating are prohibitively high.

The number of plausible assumptions on when or on how frequently the central bankers are able to make their commitment is large and, obviously, the outcome of the central bankers' game depends on the specification in this respect. While periodic commitment opportunities can be given with press conferences in the course of regular meetings of a central bank's decision-making body, permanent commitment opportunities are given if a policymaker can choose the date to announce major policy decisions at her own discretion. With regard to our main goal in this section, which is to show how the substitutability of banknotes of different denominations and currencies can determine whether central bankers face a prisoner's dilemma or a coordination problem and to explore the equilibria of these games, we take a pragmatic approach: We assume that the commitment can and must be made in period $t = 0$, and only in this period, and we define the strategy set of central banker $i \in \{H, F\}$ as $\mathcal{T}^i = \{t_e, t_l\}$ where t_e equals the optimal timing in the spillover-free benchmark such that $t_e := T^* = \frac{1}{\theta} \cdot \ln\left(\frac{r - r^{ss}}{g - r^{ss} - \omega}\right)$ (which also corresponds to the optimal timing by simultaneously moving central bankers) and where t_l is the

[19] The introduction of commitment into the basic game was inspired by Hamilton and Slutsky (1990) who (in an industrial organization context) extend a basic game between duopolists by a preplay stage with action commitment in order to endogenize the sequence of moves in the basic game between the two firms. However, as outlined in Hamilton and Slutsky (1990, p. 31), their basic game is a static game where the strategic variables are interpreted as price, quantity, or product type – in contrast to our basic game which is a dynamic game where the players' strategic variable is their timing, respectively. In this regard, we would like to thank Hans-Theo Normann for a fruitful discussion after a seminar talk and for suggesting Hamilton and Slutsky (1990) to tackle the problem of endogenizing the central bankers' roles as first and second movers (although far away from our subject, Lambertini (1996) is a paper that already has and previously made use of Hamilton and Slutsky (1990) in the context of international monetary policy coordination).

optimal timing a preassigned second mover would choose with $t_l := T^{second^*} = \frac{1}{\theta} \cdot \ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega-e^{second}}\right)$. This setting defines a static game the home and foreign central banker play in period $t=0$ where a central banker $i \in \{H, F\}$ has to choose and commit to a point in time $T^i \in \{t_e, t_l\}$ to call in her large-denomination banknotes.

The *home* and *foreign* central banker's payoffs in this static game are defined as follows: The payoff $\pi^i(T^i, T^j)$ of central banker $i \in \{H, F\}$ is a function of her own timing T^i and the timing of her peer, T^j, for $j \in \{H, F\}$ with $j \neq i$ where

$$\pi^i(T^i, T^j) = \begin{cases} F^{first}(r, T^{first} = T^i, T^{second} = T^j) & \text{if } T^i < T^j, \\ F^{second}(r, T^{first} = T^j, T^{second} = T^i) & \text{if } T^i > T^j, \\ F^{sim}(r, T^{first} = T^i, T^{second} = T^j) = F(r, T = T^i) & \text{if } T^i = T^j, \end{cases} \tag{26}$$

given that the natural rate at $t = 0$ is on the level $r_0 = r$. The functions $F^{first}(r, T^{first}, T^{second})$, $F^{second}(r, T^{first}, T^{second})$, and $F(r, T)$ are just the payoffs as defined in the leader-follower setting and the spillover-free benchmark, respectively (recall, that the payoffs describe the present value of the stream of period utilities a central banker receives from calling in her large notes plus the present value of the stream of externalities during an implementation-gap period if $T^i \neq T^j$). The central bankers' payoff matrix is visualized by table 1.

		F: t_e	F: t_l
H	t_e	$\left(\pi^H(t_e, t_e),\ \pi^F(t_e, t_e)\right)$	$\left(\pi^H(t_e, t_l),\ \pi^F(t_l, t_e)\right)$
	t_l	$\left(\pi^H(t_l, t_e),\ \pi^F(t_e, t_l)\right)$	$\left(\pi^H(t_l, t_l),\ \pi^F(t_l, t_l)\right)$

Table 1: Central bankers' payoff matrix.

We evaluate the central bankers' payoffs as given in table 1 in the following for the same state of the world and the same two spillover scenarios that we consider in section 4.2. We only consider pure strategies and use Nash equilibrium as the solution concept for the central bankers' timing game. The equilibria of this game in the two scenarios are discussed in the next section.

4.3.2 Equilibria

Scenario I (cross-currency and cross-denomination substitutability such that $e^{second} > 0$ and $e^{first} = -e^{second}$): The *home* and *foreign* central banker will face a prisoner's dilemma if *home* and *foreign* as well as *small* and *large* banknotes are substitutes from the perspective of private hoarders or criminals, i.e., if $e^{first} = -e^{second} < 0$. We have already argued in section 4.2 that the prospect of additional seignorage revenues, respectively losses, induces a preassigned second, respectively first, mover in a non-cooperative setting to defer a calling-in move. We have also already explained why both central bankers choose the same timing and delay their moves to the same extent if $e^{first} = -e^{second}$. It remains to be shown that this outcome with late simultaneous moves emerges as an equilibrium of the game between the two symmetric central bankers and that cooperation can lead to a Pareto improvement. We show this by proving the next proposition arguing that moving late and making sure not to be a first mover is a strictly dominant strategy:

Proposition 1 (inefficient delay)**.** *Let the state of the world be* $u^{ss} > 0$*,* $r > g - \omega$*,* $e^{second} \in (0, u^{ss})$*, and* $e^{first} = -e^{second} < 0$*. Then,* (t_l, t_l) *is the unique pure strategy Nash equilibrium of the central bankers' timing game* $\Gamma = (\{H, F\}, \mathcal{T}, \pi)$ *where* $\mathcal{T} = \mathcal{T}^H \times \mathcal{T}^F$ *and* $\pi = (\pi^H, \pi^F)$*. The equilibrium strategy profile* (t_l, t_l) *is Pareto inferior to the strategy profile* (t_e, t_e)*.*

Proof. It is sufficient to show that moving late at t_l is the strictly dominant strategy for the *home* central banker. Since H and F are symmetric this implies that t_l is also the strictly dominant strategy for the *foreign* central banker so that (t_l, t_l) is the unique pure strategy Nash equilibrium. Let us first consider H's payoffs if F moves early at t_e. To see why H is better off being a second mover and moving late at t_l, i.e., to see why $\pi^H(t_l, t_e) > \pi^H(t_e, t_e)$, consider H's flow utility during the period from t_e to t_l and during the period from t_l to ∞. If H moves late, she has a free ride in the form of a second mover's additional seignorage revenues and receives e^{second} until she makes her own move. So H's net utility U_t for $t \in [t_e, t_l]$ is just e^{second} and $u_t = g - r_t - \omega$ thereafter. If H moves early and at the same time as F at t_e, she has no free ride but receives $u_t = g - r_t - \omega$ for $t \in [t_e, t_l]$ and also $u_t = g - r_t - \omega$ thereafter. However, during the period $[t_e, t_l]$, the natural rate r_t is still on a relatively high level with $g - \omega \geq r_t \geq g - \omega - e^{second}$. In turn, for $t \in [t_e, t_l]$, H's utility u_t from calling in large notes would still be relatively small with $0 \leq u_t \leq e^{second}$. So, H is better off if she moves late at t_l and thus receives e^{second} during $[t_e, t_l]$.

Let us now consider H's payoffs if F moves late at t_l. If H moves late at t_l, too, she will receive $u_t = g - r_t - \omega \geq e^{second} > 0$ for $t \in [t_l, \infty)$. However, if H moves first at t_e while F moves late at t_l, she will receive utility from calling in large notes already from

t_e on but incurs an additional loss of seignorage revenues during $[t_e, t_l]$, i.e., she receives the first mover's (period flow) externality $e^{first} = -e^{second}$ during that period. As argued above, with a relatively high natural rate during $[t_e, t_l]$ the utility during that period from calling in large notes would still be relatively small with $0 \leq u_t \leq e^{second} = -e^{first}$. So, H's net period utility $U_t = u_t + e^{first} = u_t - e^{second}$ would be negative for $t \in [t_e, t_l)$. Thus, H is better off choosing t_l.

Finally, for the last part of the proposition, it remains to be shown that $\pi^H(t_e, t_e) > \pi^H(t_l, t_l)$ which implies that cooperation makes a Pareto improvement possible. This statement is true since $u_t \geq 0$ for $t \geq t_e$ which means that each of the central bankers forgoes cumulative utility to the amount of $\int_{t_e}^{t_l} u_t \cdot \exp(-\delta t)dt > 0$ if they both move late at t_l instead of moving early together at t_e (if both move simultaneously at t_e or at t_l, there will be no additional seignorage losses/revenues as in the case of sequentially moving central bankers). □

Scenario II (cross-currency substitutability but no cross-denomination substitutability such that $e^{second} > 0$ and $e^{first} = 0$): In the appendix, we argue that the net benefit of a first mover from calling in large banknotes strictly before her counterpart calls in large notes does not depend on the timing of the second mover if banknotes of different denominations are not substitutable at all. However, we argue that if banknotes of different currencies in the same denomination are substitutes from the perspective of private hoarders or criminals, a second mover will be able to enjoy a free ride in the form of additional seignorage revenues during an implementation-gap period if $T^{first} \neq T^{second}$. $e^{first} = 0$ together with $e^{second} > 0$ describe such a scenario. All in all, this leads to the sequential-move outcome in the leader-follower setting we analyze in section 4.2.3. The crucial point is that the free ride which makes a second mover better off is only possible when the two central bankers move sequentially. In the game between two symmetric central bankers this means that moving late at t_l is no dominant strategy so that two sequential-move equilibria exist, as stated in proposition 2.

Proposition 2 (coordination problem). *Let the state of the world be $u^{ss} > 0$, $r > g - \omega$, $e^{second} \in (0, u^{ss})$, $e^{first} = 0$, and $e^{second} > 0$. Then, there are two pure strategy Nash equilibria of the central bankers' timing game $\Gamma = (\{H, F\}, \mathcal{T}, \pi)$ where $\mathcal{T} = \mathcal{T}^H \times \mathcal{T}^F$ and $\pi = (\pi^H, \pi^F)$. The central bankers' equilibrium strategies are to move sequentially and the two equilibria are (t_e, t_l) and (t_l, t_e).*

Proof. As for the proof of proposition 1, due to the symmetry of H and F, it is sufficient to show that the *home* central banker's best response to F moving at t_e is to choose t_l, and that the best response to F moving at t_l is to choose t_e. Obviously, the first statement has already been shown in the proof of proposition 1: Moving late is H's best response to F moving early since H's utility as a second mover from additional seignorage revenues during $[t_e, t_l]$ is greater than the utility during $[t_e, t_l]$ from calling in large notes already together with F at t_e would be.

On the other hand, moving early is H's best response to F moving late since H forgoes cumulative utility to the amount of $\int_{t_e}^{t_l} u_t \cdot \exp(-\delta t)dt$ if she moves late at t_l instead of early at t_e. The reason is that, as argued above, $e^{first} = 0$ implies that a first mover's utility from calling in large notes does not depend on the actions of a second mover, so optimality requires H to choose the same timing as in the spillover free benchmark. □

This scenario thus shows that there is a need for central bankers to coordinate the timings of their calling-in moves. If they coordinate their actions and manage to agree on one of the two equilibrium strategy profiles, this scenario also shows that states of the world exist where optimality requires even fully symmetric central bankers to call in their large-denomination banknotes sequentially and not at the same time.

5 Policy Implications

Our analysis has shown that the substitutability of banknotes of different denominations and currencies – and thus the absolute amount of cross-country currency demand shifts that occur after unilateral calling-in moves – determine the timing and the sequence of central bankers' calling-in moves. The non-cooperative game between two fully symmetric central bankers can have a simultaneous move equilibrium in a prisoner's dilemma or multiple sequential move equilibria such that a coordination problem arises. These two cases show that cooperation and policy coordination can lead to Pareto improvements or help to avoid a coordination failure. However, from an empirical perspective, the question is how large potential Pareto improvements and thus gains from policy coordination will actually be. We have shown that the inefficiencies in the absence of cooperation have their root in suboptimal timing decisions by non-cooperative central bankers. So, the empirical questions are (1) how large the net benefits from calling in large notes that central bankers will lose in the absence of cooperation or coordination actually are, and

(2) to what extent a Pareto superior timing profile diverges from a non-cooperative or suboptimal equilibrium.

So far, these are open research questions. But with regard to the first question it is natural to assume that if the net benefits from calling in large-denomination banknotes are large, calling-in moves should be coordinated internationally even if the non-cooperative timings diverge only slightly from Pareto superior timings. Within our stylized model, the answer to the second question depends on how fast the natural rate reverts to its long-run steady state and on the absolute value of the external effects e^{first} and e^{second} that capture the additional seignorage losses, respectively gains, if central bankers call in large notes sequentially. In the prisoner's dilemma we consider in section 4.3.2, the divergence of the non-cooperative timing from the Pareto superior timing profile where both central bankers move early is just reflected in the interval $t_l - t_e = \frac{1}{\theta}\ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega-e^{second}}\right) - \frac{1}{\theta}\ln\left(\frac{r-r^{ss}}{g-r^{ss}-\omega}\right) = \frac{1}{\theta}\ln\left(\frac{u^{ss}}{u^{ss}-e^{second}}\right)$ (recall that we ruled out the possibility of an eternal free ride for a follower and assumed $e^{second} < u^{ss}$ where $u^{ss} > 0$). It is easy to see that this interval is increasing in e^{second} and decreasing in the speed at which the natural rate reverts to its steady state, θ.[20] In the coordination problem we consider in section 4.3.2 with two sequential move equilibria, a coordination failure would occur if both central bankers called in large notes simultaneously, i.e., if one central banker moved $(t_l - t_e)$ periods "too early" or "too late". The interval $(t_l - t_e)$ thus also reflects the extent to which a coordination failure outcome diverges from a Pareto superior sequential move equilibrium. So, taken all together, this means that the gains from policy coordination will be larger the more seignorage revenues are shifted between central bankers if they make their calling-in moves sequentially. Clearly, if there are significant gains, calling-in moves should be coordinated internationally. It should thus be an empirical goal to find estimates for these seignorage shifts. Conducting further effort to estimate the stock of currency that is held abroad is only one part in achieving this goal. Another crucial part – at least with respect to major reserve currencies – is to find estimates on the specific demand for large banknotes and on the substitutability of banknotes of different denominations and currencies from the perspective of different types of agents in the economy.

[20] To see this, consider $\frac{\partial(t_l-t_e)}{\partial e^{second}} = \frac{1}{\theta}\frac{1}{u^{ss}-e^{second}} > 0$ for $e^{second}, u^{ss} > 0$ with $e^{second} < u^{ss}$ and $\frac{\partial(t_l-t_e)}{\partial\theta} = -\theta^{-2}\ln\left(\frac{u^{ss}}{u^{ss}-e^{second}}\right) < 0$ in the region of the parameter space we consider.

6 Concluding Remarks

Our analysis has pointed out the extent to which the expectation of cross-country shifts in seignorage revenues, in case large banknotes are called in unilaterally, can induce central bankers to postpone their calling-in moves to a later point in time. In Link (2019) we argue that uncertainty over the future path of the natural rate of interest can also induce a central banker to postpone a calling-in move. Taken together, there are thus several reasons which suggest that major central banks – if ever – will eliminate large banknotes only in times of exceptionally low natural rates of interest. In some circumstances, international coordination could accelerate this process.

We have left out a couple of questions raised directly by our analysis. For instance, a natural way to carry our analysis forward is to explore the joint effect of strategic interactions and uncertainty over future states of the world on a central banker's decision to call in large notes. Such an analysis could shed light on the question of whether the individual effects of uncertainty and strategic interactions will reinforce or dampen each other. There is also uncertainty along many other dimensions that we have not taken into account – uncertainty over the actual demand for large-denomination banknotes, and thus about the magnitude of potential seignorage losses, is just one of them. Another interesting task for further research is to explore the behavior of central bankers that play a fully dynamic timing game with more than one point in time where they can commit to future policies. What we have also left out for future research is an exploration of the central bankers' timing game in other scenarios and for alternative paths of the natural rate. Eventually, the exploration of all these key issues will do the groundwork for analyzing the optimal timing of calling in large-denomination banknotes in an open economy macro model.

Appendix

A A Network Model of Banknote Demand and Seignorage Shifts

We set up a small network model of banknote demand in this appendix. Our goal is to illustrate how the external effect e^{first} respectively e^{second} that a first, respectively a second, mover receives can be interpreted as the loss, respectively the gain, in seignorage revenues that occur if the two central bankers make their calling-in moves sequentially. We argue that the ratio of the first and second mover's externality can be related to the degree of substitutability of banknotes of different denominations and currencies from the perspective of private hoarders or criminals.

Suppose that in our stylized framework of section 2 there is an external third country. We introduce a third country to avoid the unnecessary complexity that would emerge if we had to specify what would happen to the *foreign* demand for *home* currency if the *home* central banker called in *large* notes and vice versa. For instance, the third country can be thought of as a developing country where private individuals have a specific demand for *home* and *foreign* currency as a store of value or as a means to facilitate illegal transactions.[21] So, let us further suppose that there is a third country demand for *home* and *foreign* currency in both banknote denominations from private hoarders or criminals, that cash supply is fully elastic, and that the quantity of cash circulating in the third country is thus completely demand-driven (with private hoarders or criminals being the only demanders of cash). We denote the total value of all banknotes of size $k \in \{s, l\}$ of currency $j \in \{H, F\}$ circulating in the third country at time t by $M_{j,k}(t)$ where s denotes *small*-denomination and l *large*-denomination banknotes, respectively (H and F abbreviate *home* and *foreign*).[22] Furthermore, let us abstract from all other determinants of third country cash demand, like income fluctuations, inflation, exchange rate volatility, opportunity costs of holding cash, etc. With these assumptions, third country cash demand depends only on the available banknote denominations and different currencies such that each $M_{j,k}(t)$ is constant until the points in time T^H and T^F when the home and foreign central banker make their calling-in moves, respectively. The changes at T^H and T^F occur instantly and comprise simple, discrete demand shifts $\epsilon_{j,k}^{m,n}(t) \geq 0$ from $M_{j,k}$ to $M_{m,n}$ that occur at time $t \in \{T^H, T^F\}$. Importantly, we do not require the total value of currency circulating in the third country to remain constant at T^H and T^F. The shrinkage in total third country cash holdings due to the calling-in of *large*-denomination banknotes of currency j at time $t \in \{T^H, T^F\}$ is denoted by $\epsilon_{j,l}^{\text{shrinkage}}(t) \geq 0$.[23] For the sake of

[21] See, for instance, for the store-of-value motive Fischer, Köhler, and Seitz (2004); Feige (2012); Bartzsch, Seitz, and Setzer (2015) or for a discussion of the specific demand for *large* notes from the underground economy Rogoff (1998) or Rogoff (2016, chapter 13).

[22] We abstract from differentiating explicitly between real and nominal money holdings but M can be thought of as a real variable.

[23] An increase in total global cash holdings at T^H or at T^F is ruled out.

clarity, figure 2 summarizes the possible paths of demand shifts from one denomination to another.

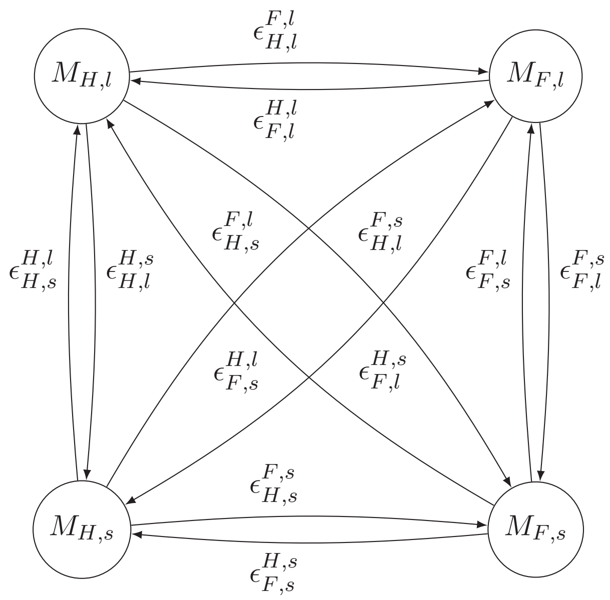


Figure 2: Possible paths of shifts in banknote demand, respectively currency substitution effects, illustrated as a network: The nodes of this network represent banknote demand before demand shifts occur. The demand shifts, in turn, are represented by the edges.

We are aware of the oversimplification but this stylized framework is everything we need to describe different scenarios of shifts in seignorage revenues from one central banker to her counterpart. To obtain a definition of seignorage profits we start with the assumption that *home* and *foreign* currency is brought into circulation in the third country in the following way: Both central bankers are ready to fully meet third country demand for their currencies, which is $M_{j,s} + M_{j,l}$ for $j \in \{H, F\}$. To meet demand, a central banker buys one unit of a specific third country asset (for instance, a third country government bond) with one unit of her currency. So, a central banker just swaps one unit of newly created currency for one unit of the third country asset. One unit of the third country asset yields a flow return $\kappa \in \mathbb{R}_{\geq 0}$. κ is constant over time and, in particular, it is independent of the (*home* and *foreign*) natural rate of interest. The present value of the stream of flow returns from holding $M_{j,k}$ units of the third country asset from time t_1 to t_2 is thus

$$\mathcal{S}_{t_1}^{t_2}(j,k) := \int_{t_1}^{t_2} \kappa M_{j,k}(t) \cdot \exp(-\delta t) dt. \tag{27}$$

We refer to $\mathcal{S}_{t_1}^{t_2}(j,k)$ as seignorage profits from issuing $M_{j,k}$ units of currency to the third country from t_1 to t_2 and use this definition of seignorage in the following.[24]

There are countless constellations of banknote demand and seignorage shifts across central bankers but we shall limit our attention to two scenarios – one with and one without a disadvantage for a first mover. Whether there is a disadvantage or not will depend on the substitutability of *small/large* and *home/foreign* banknotes. The first mover disadvantage will be reflected in $e^{first} \in (-\infty, 0]$. We set up the scenarios such that a second mover will always be better off in both cases, which means that $e^{second} > 0$. Our goal is to illustrate how seignorage shifts after calling-in moves are described by e^{first} and e^{second} and how the ratio of the externalities reflects the substitutability of different banknotes and currencies.

Scenario I (cross-currency and cross-denomination substitutability such that $e^{second} > 0$ and $e^{first} = -e^{second}$): Suppose that the state of the world is such that banknotes of different currencies and denominations are substitutes from the perspective of private hoarders and criminals in the third country. Suppose further that the allocation of total third country cash demand to the available banknotes is just driven by a preference to split cash holdings equally in terms of value across the available denominations and currencies. Without loss of generality, assume that $T^H < T^F$ and that third country cash demand (and thus cash in circulation) before any central banker has called in her *large*-denomination banknote is $M_{H,s}(t), M_{H,l}(t), M_{F,s}(t), M_{F,l}(t) = 3\mu$ for all $t \in [0, T^H)$ and for some arbitrary $\mu > 0$. Total third country cash demand is thus 12μ. In this scenario, we assume that a discontinued *large* denomination can be perfectly replaced by the remaining denominations such that total third country cash demand will stay constant at T^H and T^F with the demand for the discontinued denomination, respectively, being reallocated in equal shares to the remaining banknote alternatives. So, the shifts from *large home* currency banknotes at T^H are $\epsilon_{H,l}^{H,s}(T^H), \epsilon_{H,l}^{F,l}(T^H), \epsilon_{H,l}^{F,s}(T^H) = \mu$. This implies $M_{H,s}(t), M_{F,s}(t), M_{F,l}(t) = 4\mu$ for all $t \in [T^H, T^F)$. The demand for the discontinued denomination at T^F is reallocated equally as well with $\epsilon_{F,l}^{H,s}(T^F), \epsilon_{F,l}^{F,s}(T^F) = 2\mu$, so for all $t \geq T^F$ it is $M_{H,s}(t), M_{F,s}(t) = 6\mu$. Thus, the *foreign* central banker is able to satisfy an additional share of third country cash demand of 2μ during $t \in [T^H, T^F)$ while the *home*

[24] This measure of seignorage thus, in principle, corresponds to the one Buiter (2007, p. 5) (and the citation therein) refers to as *"Central Bank revenue"*. See also Buiter (2007) for other definitions or measures of seignorage.

central banker loses an equal share during this period. The additional, respectively lost, flow seignorage profits during $[T^H, T^F)$ are thus $\kappa \cdot 2\mu$. We relate $e^{second} = -e^{first}$ to this value.[25] Figure 3 illustrates banknote demand and the demand shifts that occur at T^H and T^F.

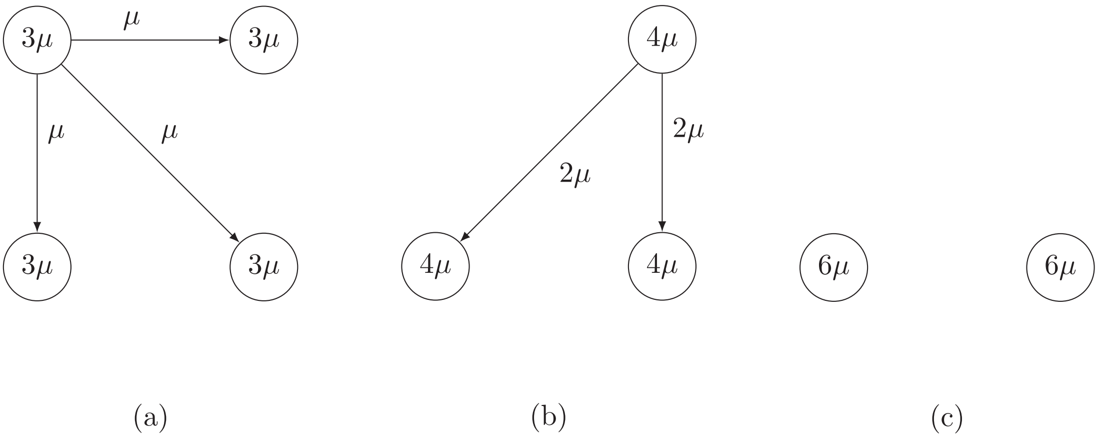


Figure 3: Banknote demand and demand shifts in scenario I illustrated as networks. The structure (i.e., the nodes and edges) of these networks corresponds to the structure of the network defined in figure 2: Subfigure (a) illustrates banknote demand until T^H, which is 3μ for each denomination, and the demand shifts μ that occur due to the *home* central bankers calling-in move at T^H. Subfigure (b) shows banknote demand between T^H and T^F and the demand shifts that occur at time T^F. Subfigure (c) shows the allocation of third country banknote demand after *home* and *foreign* banknotes have been called in.

Scenario II (cross-currency substitutability but no cross-denomination substitutability such that $e^{second} > 0$ and $e^{first} = 0$): Suppose that the world at the beginning (in $t = 0$) is exactly like in scenario I with the crucial difference that banknotes of different denominations are not substitutable at all from a third country demand-side perspective in scenario II. Instead, suppose that *small* and *large* notes are demanded for completely different reasons and that third country demand for the discontinued *large*-denomination *home* currency banknotes completely shifts to *large*-denomination *foreign* currency banknotes at T^H with $\epsilon_{H,l}^{F,l} = 3\mu$ so that $M_{H,s}(t), M_{F,s}(t) = 3\mu$ and $M_{F,l}(t) = 6\mu$ for all $t \in [T^H, T^F)$. Since *small*-denomination banknotes are no alternative for the discontinued *large*-denomination *foreign* currency banknotes, total third country cash demand decreases by 6μ at T^F and $M_{H,s}(t), M_{F,s}(t) = 3\mu$ for all $t \geq T^F$. The point is that the

[25] Therewith, the present value of the additional, respectively lost, flow seignorage profits is thus $\int_{T^H}^{T^F} \kappa \cdot 2\mu \cdot \exp(-\delta t) dt$ and consequently, $\int_{T^H}^{T^F} e^{second} \cdot \exp(-\delta t) dt = -\int_{T^H}^{T^F} e^{first} \cdot \exp(-\delta t) dt$ can be related to this value.

first mover's seignorage profits once she has made her calling-in move do not depend on the second mover's timing nor whether a potential second mover moves at all. So, there is no difference between being a first mover or being one of two simultaneous movers and thus $e^{first} = 0$. On the other hand, e^{second} can be related to the second mover's additional flow seignorage profits of $\kappa \cdot 3\mu$.[26] Figure 4 illustrates how banknote demand changes due to the central bankers' calling-in moves.

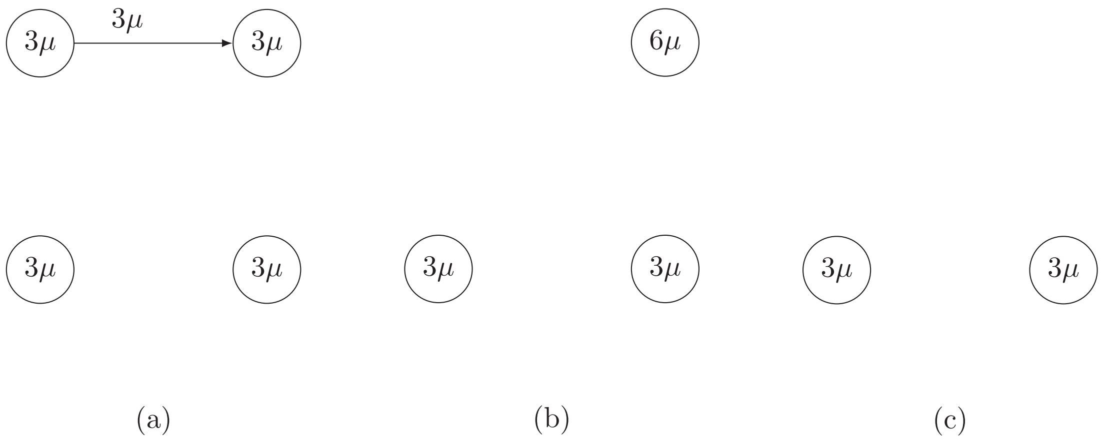


Figure 4: Banknote demand and demand shifts in scenario II illustrated as networks. The structure (i.e., the nodes and edges) of these networks corresponds to the structure of the network defined in figure 2: Subfigure (a) illustrates banknote demand until T^H, which is 3μ for each denomination, and the demand shift 3μ from *large home* notes to *large foreign* notes that occurs due to the *home* central bankers calling-in move at T^H. Subfigure (b) shows banknote demand between T^H and T^F. At T^F, total third country banknote demand decreases by 6μ, so there are no demand shifts from *large foreign* notes to the remaining denominations. Subfigure (c) shows the allocation of third country banknote demand after *home* and *foreign* banknotes have been called in.

[26] The present value of the second mover's additional flow seignorage profits is thus $\int_{T^H}^{T^F} \kappa \cdot 3\mu \cdot \exp(-\delta t)dt$.

Bibliography

AGARWAL, R., AND M. S. KIMBALL (2019): "Enabling Deep Negative Rates to Fight Recessions: A Guide," IMF Working Paper 19/84, International Monetary Fund.

ALVAREZ, F., AND A. DIXIT (2014): "A Real Options Perspective on the Future of the Euro," *Journal of Monetary Economics*, 61, 78–109.

BARTZSCH, N., G. RÖSL, AND F. SEITZ (2013): "Currency Movements within and outside a Currency Union: The Case of Germany and the Euro Area," *The Quarterly Review of Economics and Finance*, 53(4), 393–401.

BARTZSCH, N., F. SEITZ, AND R. SETZER (2015): "The Demand for Euro Banknotes Issued in Germany: Structural Modelling and Forecasting," ROME Discussion Paper Series, No. 15-03, Research On Money in the Economy (ROME).

BUITER, W. H. (2007): "Seigniorage," *Economics: The Open-Access, Open-Assessment E-Journal*, 1(2007-10), 1–49.

——— (2009): "Negative Nominal Interest Rates: Three Ways to Overcome the Zero Lower Bound," *The North American Journal of Economics and Finance*, 20(3), 213 – 238.

CALZA, A., AND A. ZAGHINI (2016): "Shoe-Leather Costs in the Euro Area and the Foreign Demand for Euro Banknotes," *International Journal of Central Banking*, 12(1), 231–246.

DIXIT, A. K., AND R. S. PINDYCK (1994): *Investment under Uncertainty*. Princeton University Press, Princeton, New Jersey.

ENGEL, C. (2016): "International Coordination of Central Bank Policy," *Journal of International Money and Finance*, 67, 13–24.

EUROPEAN CENTRAL BANK (2019): "The International Role of the Euro," European Central Bank, Frankfurt am Main, June 2019, available at `https://www.ecb.europa.eu/pub/pdf/ire/ecb.ire201906~f0da2b823e.en.pdf`.

FEIGE, E. L. (2012): "New Estimates of U.S. Currency Abroad, the Domestic Money Supply and the Unreported Economy," *Crime, Law and Social Change*, 57(3), 239–263.

FISCHER, B., P. KÖHLER, AND F. SEITZ (2004): "The Demand for Euro Area Currencies: Past, Present and Future," Working Paper Series No. 330, European Central Bank.

FUDENBERG, D., AND J. TIROLE (1985): "Preemption and Rent Equalization in the Adoption of New Technology," *The Review of Economic Studies*, 52(3), 383–401.

GOODFRIEND, M. (2000): "Overcoming the Zero Bound on Interest Rate Policy," *Journal of Money, Credit and Banking*, 32(4), 1007–1035.

HAMILTON, J. H., AND S. M. SLUTSKY (1990): "Endogenous Timing in Duopoly Games: Stackelberg or Cournot Equilibria," *Games and Economic Behavior*, 2(1), 29–46.

KRÜGER, M., AND F. SEITZ (2017): *The Benefits of Cash: Costs and Benefits of Cash and Cashless Payment Instruments (Module 2)*. Fritz Knapp Verlag, Frankfurt am Main.

Lambertini, L. (1996): "The International Coordination of Monetary Policy: A Game-Theoretic Reformulation," Quaderni - Working Paper DSE, No. 250, Alma Mater Studiorum – Università di Bologna, Dipartimento di Scienze Economiche (DSE), Bologna.

Link, T. (2019): "Optimal Timing of Calling In Large-Denomination Banknotes under Natural Rate Uncertainty," DICE Discussion Paper, No. 327, Düsseldorf Institute for Competition Economics (DICE), Düsseldorf, available at `http://www.dice.hhu.de/fileadmin/redaktion/Fakultaeten/Wirtschaftswissenschaftliche_Fakultaet/DICE/Discussion_Paper/327_Link.pdf`. The latest version of this paper is included as the second paper in this dissertation.

Liu, Z., and E. Pappa (2008): "Gains from International Monetary Policy Coordination: Does It Pay to be Different?," *Journal of Economic Dynamics and Control*, 32(7), 2085–2117.

Maller, R. A., G. Müller, and A. Szimayer (2009): "Ornstein–Uhlenbeck Processes and Extensions," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch, pp. 421–437. Springer, Berlin Heidelberg.

Pappa, E. (2004): "Do the ECB and the Fed Really Need to Cooperate? Optimal Monetary Policy in a Two-Country World," *Journal of Monetary Economics*, 51(4), 753–779.

Reinganum, J. F. (1981): "On the Diffusion of New Technology: A Game Theoretic Approach," *The Review of Economic Studies*, 48(3), 395–405.

Rognlie, M. (2016): "What Lower Bound? Monetary Policy with Negative Interest Rates," unpublished manuscript, July 2016, available at `http://mattrognlie.com/negative_rates.pdf`.

Rogoff, K. (1998): "Blessing or Curse? Foreign and Underground Demand for Euro Notes," *Economic Policy*, 13(26), 262–303.

——— (2013): "Comment on 'International Policy Coordination: Present, Past and Future' by John B Taylor," released within BIS Working Paper No. 437, Bank for International Settlements, December 2013, available at `https://www.bis.org/publ/work437.pdf`.

——— (2015): "Costs and Benefits to Phasing out Paper Currency," *NBER Macroeconomics Annual*, 29, 445–456.

——— (2016): *The Curse of Cash*. Princeton University Press, Princeton and Oxford.

——— (2017): "Dealing with Monetary Paralysis at the Zero Bound," *Journal of Economic Perspectives*, 31(3), 47–66.

Taylor, J. B. (2013): "International Monetary Policy Coordination: Past, Present and Future," BIS Working Papers, No. 437, Bank for International Settlements.

Thiele, C.-L., D. Niepelt, M. Krüger, F. Seitz, R. Halver, and A. F. Michler (2015): "Diskussion um das Bargeld: Hätte eine Abschaffung von Banknoten und Münzen wirklich Vorteile?," *ifo Schnelldienst*, 68(13), 3–18.

Weeds, H. (2002): "Strategic Delay in a Real Options Model of R&D Competition," *The Review of Economic Studies*, 69(3), 729–747.