Quantifying the Role of Financial Factors during the Great Recession

Rodrigo Heresi

May 2018

Abstract

The Great Recession spurred a new wave of models with financial frictions aimed to understand the particular features of such a crisis. Most of these models, however, can be mapped into prototype economies with “agnostic” intertemporal wedges, which are deemed to be not promising to explain U.S. business cycles (Chari, Kehoe, and McGrattan, 2007; CKM). Was this time different? Are these new wave of models well-suited to quantitatively account for the so called financial crisis? To answer these questions, I augment a real business cycle model with financial intermediaries that face an endogenously determined balance sheet constraint (Gertler and Karadi, 2011; GK). In order to capture the intrinsic nonlinear nature of the crisis and to give the friction the best chance to play a role, I allow for the financial constraint to bind only occasionally. This way I capture the idea of infrequent financial crises nested within typical business cycles without relying on unrealistically large shocks. I show that the model with microfounded friction is equivalent to a prototype economy with an (exogenous) intertemporal investment wedge, which is a function of the key endogenous variables associated with the friction in the baseline model. Consistent with CKM, I confirm that these type of frictions are unimportant to account for U.S. macroeconomic fluctuations over the five decades previous to the crisis. More surprisingly, I show that the CKM result is robust to (a) the extension to the Great Recession period, (b) the introduction of a nonlinear framework able to switch between “tranquil times” and “financial crises”, (c) the solution and filtering of structural shocks using nonlinear techniques, and (d) the introduction of spread data to inform the model about the severity of the friction.

Keywords: Business Cycles; Financial Markets; Financial Crises; Bayesian Analysis.
JEL classifications: E32, E44, G01, C11.
1 Introduction

There is a long debate in macroeconomics about what drives business cycles. Under the lens of stochastic general equilibrium frameworks, economic fluctuations arise from disturbances to the model’s equilibrium conditions, which can be interpreted as structural shifts in preferences and technology (e.g. Smets and Wouters, 2007), as reduced-form representations of frictions that manifest as time-varying wedges (e.g. Chari, Kehoe, and McGrattan, 2007), or simply as “convenient” representations of model misspecification (e.g. Primiceri, Schaumburg, and Tambalotti, 2006).

The Real Business Cycle literature pioneered by Kydland and Prescott (1982) and Long and Plosser (1983) strikingly illustrated that economic fluctuations can be largely accounted for by random technology shocks to the production function (King and Rebelo, 1999). On the downside, the neoclassical growth model predicts that the (tax-adjusted) household’s marginal rate of substitution (MRS) between consumption and leisure should equals the marginal product of labor (MPL), an equilibrium condition that fails miserably in the U.S. post-war data. The resulting labor wedge, the gap between MRS and MPL, varies significantly over the business cycle in a countercyclical way (Shimer, 2009; Karabarbounis, 2014), and is one of the key driving forces to explain U.S. business cycles (Chari, Kehoe, and McGrattan, 2007). Hall (1997) decompose the labor wedge emphasizing the distinction between intratemporal and intertemporal channels, finding that most of the movements in employment over the business cycle are due to intratemporal “preference” shocks.

On the other hand, Primiceri, Schaumburg, and Tambalotti (2006), and Justiniano, Primiceri, and Tambalotti (2010, 2011) find that intertemporal disturbances are the key source of macroeconomic fluctuations. They reach this conclusion using a New Keynesian setup with a rich set of real and nominal frictions. In recent years, motivated by the arguably prominent role of explicit financial frictions during the Great Recession episode of 2007-2009, there has been a renewed interest on the role of intertemporal disturbances in shaping business cycles. In fact, most prominent models with microfounded financial frictions used to study the recent financial crisis can be mapped into prototype economies with intertemporal wedges. For instance, in the present article I show that a real version of the Gertler and Karadi (2011) model is equivalent to an economy with intertemporal investment shocks. Likewise, Chari, Kehoe, and McGrattan (2007) show that an economy with the type of credit market frictions considered in Bernanke, Gertler, and Gilchrist (1999) is equivalent to a growth model in which there is a wedge in the Euler equation for capital. More recently, Ajello (2016) set up a model with Kiyotaki and Moore’s (2012) type of friction, in which financial intermediation disturbances act as intertemporal wedges. He finds that these financial shocks were a key business cycle driver not only at the onset of the Great Recession but also during most of the Great Moderation period.

This paper contributes to this literature by assessing quantitatively the importance played by financial frictions and financial shocks in the U.S. business cycle, with a particular focus on the Great Recession. The financial crisis of 2007-2009 is a particularly relevant episode to study the role of intertemporal disturbances, because as mentioned above, most financial frictions emphasized in the literature manifest themselves as intertemporal investment wedges. At the core of the analysis is a real business cycle model augmented to include financial intermediaries (banks, for short) facing endogenously determined balance sheet constraints. Banks take depos-
its from households and combine them with their own net worth to produce state-contingent loans to firms. Following Gertler and Kiyotaki (2011) and Gertler and Karadi (2011) (GK henceforth), the relationship between banks and households is characterized by a moral hazard problem, which ultimately limits banks’ ability to raise funds (borrow), and hence to acquire assets (lend). In equilibrium, a contraction of banks’ net worth may activate a financial accelerator effect, in which banks delever through fire-selling of assets, credit spreads rise, investment plummets, and the economy may face a protracted recession. Following much of the RBC literature, the competing shocks include standard productivity, labor wedge and government spending shocks, as well as a less standard disturbance to the quality of capital held by the banking sector. The latter shock is often used in the literature to trigger asset price dynamics, and is also interpreted here as a “financial shock” that mimics the significant losses from toxic assets observed in the U.S. banking sector during the Great Recession.

In the spirit of Chari, Kehoe, and McGrattan (2007), I show that the baseline RBC model with GK friction maps into a prototype economy with labor and investment wedges. Unlike that paper, however, and given the size of the shocks hitting the economy during the financial crisis, I use nonlinear techniques not only to solve the model but also to uncover the structural driving forces behind economic fluctuations. Moreover, in order to capture the intrinsic nonlinear nature of this type of crises, I allow the balance sheet constraint associated with the GK friction to bind only occasionally, typically when banks’ leverage is sufficiently high. In the spirit of Mendoza (2010), occasionally binding constraints can potentially capture the idea of infrequent financial crises nested within typical business cycles.

In a nutshell, the model has the ability to generate conditional amplification, giving rise to an asymmetry in the relationship between the net worth of the banking sector and economic activity. During “tranquil times”, when the balance sheet constraint is slack, credit spreads are low and the economy (conditional on the realization of other relatively benign shocks) is booming. I show that under a fully nonlinear solution the economy spends most of the time in the slack regime, because banks have the incentive to act with cautious and hold “precautionary equity capital”. In other words, forward-looking banks anticipate the possibility that future shocks may push them into a vulnerable zone (dangerously near the constraint), leading to precautionary deleveraging, a mission that is accomplished by cutting lending to firms. In this environment, the sensitivity of the financial system to shocks is relatively small and the economy behaves like a frictionless neoclassical benchmark. However, in some states (say, a 2007-2009 scenario), an unlikely but possible combination of bad shocks can push the banks to hit the leverage limit, and the economy shifts into a “financial crisis” regime. Credit spreads rise sharply, and non-financial firms respond by borrowing less, so the equilibrium amount of credit drops. Along the way, the financial accelerator mechanism embedded in the model amplifies the initial shock. Less borrowing translates into less investment, which in turn leads to a fall in output, consumption, and the price of capital. The fall in the return to capital feeds back into the balance sheets of banks, propagating the effects even after the initial shock has dissipated.

The remainder of the article is organized as follows. Section 2 makes a brief summary of the related literature. Section 3 describes the baseline model with an occasionally binding GK friction. Section 4 shows the mapping between the baseline model and a prototype economy with an intertemporal investment wedge. Section 5 discusses the main results of the paper and Section 6 concludes.
2 Relationship with the Literature

The present paper is related to the literature that explores quantitatively the main driving forces behind macroeconomic fluctuations. Stochastic general equilibrium models imply three broad classes of equilibrium conditions: intratemporal first-order conditions, intertemporal first-order conditions, and accounting relationships between inputs and outputs. Business cycles originate from disturbances hitting these equilibrium relationships.

Chari, Kehoe, and McGrattan (2007) (CKM henceforth) use their so-called Business Cycle Accounting method to conclude that neutral technology shocks (efficiency wedges in their nomenclature) and intratemporal preference shocks (labor wedges) together account for essentially all of the economic fluctuations during the Great Depression and the 1982 recession in the United States. More recently, Brinca, Chari, Kehoe, and McGrattan (2016) apply the same CKM method to update their results including the Great Recession, obtaining similar results: the intratemporal labor wedge played the dominant role during the recent financial crisis, while the intertemporal investment wedge played a decidedly tertiary role. These conclusions are reminiscent of classic results in the real business cycle literature emphasizing the role of productivity and preference shocks (e.g. King and Rebelo, 1999; Hall 1997).

Primiceri, Schaumburg, and Tambalotti (2006) (PST henceforth) emphasize two main problems about the studies mentioned above emphasizing the role of the labor wedge: first, they concentrate in environments in which physical capital is the only asset; second, they disregard asset market returns data to inform the model. Therefore, the only Euler equation implies very smooth dynamics on both the return to capital in the economy (as a function of the stable output-to-capital ratio) and the stochastic discount factor (measured through consumption growth), thus “fitting” very small intertemporal disturbances. By considering an economy in which a short-term nominal bond is traded along with physical capital, and by exploiting the bond pricing implications of an estimated state-of-the-art business cycle model, they obtain a prominent role for intertemporal disturbances. Similarly, Christiano, Eichenbaum, and Trabandt (2015) find that the vast bulk of movements in aggregate real activity during the Great Recession were due to intertemporal wedges introduced on the households’ Euler equations associated with both nominal risk-free bonds and capital accumulation.

This paper borrows insights from both stripes of the literature. First, in the spirit of CKM, I build a detailed economy with microfounded financial frictions that is observationally equivalent to a prototype economy with investment wedges. I allow for the financial constraint to bind only occasionally, and show that the associated intertemporal wedge in the prototype model is a function of the multiplier on the bankers’ inequality constraint. Intuitively, if the financial constraint in the baseline model is never to bind, then the investment wedge is always zero, and both models behave like a frictionless RBC benchmark. Unlike CKM who allow for correlated shocks, I assume that the model’s exogenous innovations are independent, a necessary condition for a meaningful structural interpretation of the shocks.

Second, following PST’s advice, I build a model in which a risk-free government bond is traded along with physical capital, giving rise to two intertemporal Euler equations, and inform them with data on credit spreads. Unlike PST, and given the focus on the Great Recession, I filter the structural innovations using a nonlinear filter that enforces the occasionally binding constraint.
This paper also builds on the growing body of literature that studies the role of financial frictions and financial shocks for business cycles. Much of the earlier research about financial frictions emphasized the role of non-financial firms’ balance sheets in the propagation of shocks. Seminal articles by Bernanke and Gertler (1989), Carlstrom and Fuerst (1997), and Bernanke, Gertler, and Gilchrist (1999) state that credit-market imperfections may significantly amplify shocks and hinder investment by worsening the terms at which firms can borrow. As asset values typically fall during downturns, the initial shock may be further amplified in subsequent periods through tightening collateral constraints (Kiyotaki and Moore (1997)). Recently, Christiano, Motto, and Rostagno (2014) use a BGG framework allowing the volatility of cross-sectional idiosyncratic uncertainty to fluctuate stochastically over time, and show that fluctuations in risk are the most important shock driving the cycle, including during the 2008 financial crisis.

On the other hand, the Great Recession gave rise to renewed research emphasizing frictions and shocks that originate directly in the financial sector.\footnote{See Gertler and Kiyotaki (2010), Christiano, Motto, and Rostagno (2010), Del Negro et al. (2010), Gertler and Karadi (2011, 2012), Jermann and Quadrini (2012), Kiyotaki and Moore (2012), Iacoviello (2015), Bigio (2015), among others. For recent surveys on financial frictions, see Brunnermeir, Eisenbach, and Sannikov (2012) and Quadrini (2011).} In the presence of financial frictions, fluctuations in credit spreads and overall lending standards may reflect shifts in the effective supply of funds offered to firms, with important spillovers to the real economy (Gilchrist and Zakrajsk (2012)). The so-called bank lending channel states that banks’ losses from “toxic” assets during the financial crisis forced them to delever by fire-selling securities and reducing lending, therefore shrinking the effective supply of credit available to non-financial firms (Gertler and Kiyotaki (2011), Gertler and Karadi (2011, 2012)). This papers builds on the insights of the latter articles, but allowing for an occasionally binding financial constraint. By doing so, I attempt to explain the mechanisms that caused small losses in the mortgage market (relative to the size of the economy) to amplify into such large dislocations in the financial markets as the ones observed in the summer of 2008.

By building a model with an occasionally binding constraint, this paper is also related to the new body of literature studying nonlinear models with endogenous switching between “normal times” and “financial crisis” regimes, as in Mendoza (2010). Unlike the latter paper, the occasionally binding constraint is derived from a micro-founded moral hazard problem and is imposed on the banking sector rather than the entrepreneurial sector, as in Akinci and Queralto (2014) and Bocola (2016). Unlike the present paper, Akinci and Queralto (2014) use a small open economy setup in which the interest rate evolves mechanically as an autoregressive process with debt-elastic feedback from the country’s international debt-to-output ratio. In the present paper, the interest rates (hence the lending-deposit spread) are fully determined in the general equilibrium. In that sense, this paper is closer to He and Krishnamurthy (2014), who also build a model with an occasionally binding constraint on the banks’ equity capital, but using a setup similar to the one proposed by Holmstrom and Tirole (1997). They calibrate the model for the U.S. economy and use it to characterize the transition from a “normal” state to what they label as a “systemic risk” state that apparently took place during the Great Recession episode. In turn, Bocola (2016) focuses on the effects of a sovereign default risk shock on financial intermediation, in the context of the Italian debt crisis of 2011.
3 The Model

I consider a real business cycle model augmented with a financial friction that limits the ability of the banking sector to acquire funds from savers, in the spirit of Gertler and Kiyotaki (2010) and Gertler and Karadi (2011). Such a friction ultimately gives rise to an endogenous limit on banks’ leverage ratio that may restrict their ability to channel funds efficiently from savers to bank-dependent agents. Unlike the above mentioned articles, the bank’s constraint binds only occasionally, typically when a sequence of bad shocks hits a relatively vulnerable and highly leveraged banking sector. I assume there is no friction in the relationship between banks and the corporate sector.

The economy is populated by four types of private agents: households, banks, capital goods producers, and final goods producers. There is also a government that finances its purchases of the final goods by levying lump-sum taxes and by issuing bonds. Regarding households, I use the “large family” metaphor in order to maintain the tractability of the representative agent approach. More specifically, there are constant fractions of workers and bankers within each household. Workers supply labor and return the wages they earn to the household. Bankers manage financial intermediaries and transfer any earnings back to their household. Within the family there is perfect consumption insurance. Final good producers combine labor and capital in order to produce the single final good in the economy. They need external finance from banks in order to buy physical capital from capital producers, which in turn combine old left-over capital with investment in order to produce new capital. Both types of firms are owned by households.

3.1 Households

Households consume, supply labor, and save. Households do not hold capital directly. Rather, they save by making deposits in competitive financial intermediaries or by purchasing government bonds. Both bank deposits and government debt are non-state-contingent, one-period real bonds that pay the gross return $R_t$ from $t-1$ to $t$. In the equilibrium considered here, both instruments are riskless and thus perfect substitutes. Therefore, I impose this condition in the budget constraint from the outset. The typical household solves the following problem:

$$\max_{\{C_t, H_t, D_t\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[ \frac{C_t^{\gamma-\gamma}}{1-\gamma} - \phi t \frac{H_t^{1+\zeta}}{1+\zeta} \right]$$

subject to

$$C_t + D_t + T_t = W_t H_t + R_{t-1} D_{t-1} + \Sigma_t$$

(1)

where $C_t$ is consumption, $H_t$ is hours worked, $D_t$ is total savings, $W_t$ is the real wage, $T_t$ is lump-sum taxes, $\Sigma_t$ is real dividends from the ownership of firms and banks (net of start-up transfers

\footnote{It is best to think of them as making deposits in banks other than the ones they own. The implicit assumption is that banks are specialists at evaluating, monitoring and enforcing loan contracts, which is why firms rely exclusively on banks to obtain funds. Gertler and Kiyotaki (2015) consider a model in which banks and households may extend loans to firms, but the latter are less efficient in doing so.}

\footnote{Since all households solve an identical problem I omit household subscripts.}
that households give to its members entering banking activities, more details below), \( \varphi_t \) is an exogenous preference (labor wedge) shock. Parameters \( \beta, \gamma, \zeta, \) and \( \chi \) are the discount factor, the risk aversion, the inverse of the Frisch elasticity of labor supply, and a scale parameter that affects the marginal rate of substitution between consumption and leisure, respectively. The first-order conditions are fairly standard:

\[
\varphi_t \chi H_t^\zeta = C_t^{-\gamma} W_t \\
1 = E_t [\Lambda_{t,t+1} R_t]
\]

where \( \Lambda_{t,t+i} \equiv \beta^i \left( \frac{C_{t+i}}{C_t} \right)^{-\gamma} \) is the household’s marginal rate of substitution.

### 3.2 Banks

Banks use their own net worth together with one-period deposits from households to provide equity finance to the final goods producers. In particular, they buy claims on the returns of physical capital that final goods producers purchase, period by period, from capital goods producers. Let \( N_{jt} \) be the end-of-period \( t \) net worth in the hands of bank \( j \), \( D_{jt} \) be deposits received from households, and \( S_{jt} \) be the number of claims purchased to firms at market price \( Q_t \). The balance sheet of bank \( j \) at the end of period \( t \) is given by

\[
Q_t S_{jt} = D_{jt} + N_{jt}.
\]

Financial intermediaries accumulate net worth through retained earnings. Bankers’ liabilities pay the non-state-contingent real gross return \( R_t \), and its assets earn the state-contingent real gross rate \( R_{K,t+1} \). Accordingly, bankers’ net worth evolves as follows:

\[
N_{j,t+1} = R_{K,t+1} Q_t S_{jt} - R_t D_{jt} = (R_{K,t+1} - R_t) Q_t S_{jt} + R_t N_{jt}
\]

where the second equality follows from equation (4). Intuitively, any increase in net worth above the riskless return is a function of the spread \( (R_{K,t+1} - R_t) \) and the market value of the securities purchased from firms. Under frictionless financial markets, as in the standard neoclassical model, banks always have enough funds to arbitrage away differences between the risk-adjusted lending and deposit rates:

\[
E_t \Lambda_{t,t+1+i} R_{t,t+1+i} = E_t \Lambda_{t,t+1+i} R_{t+i}, \quad i \geq 0.
\]

The key to the notion of financial factors affecting real activity is the existence of limits to this arbitrage, so that credit spreads may arise. Following Gertler and Kiyotaki (2011) and Gertler and Karadi (2011), I assume limited enforcement of contracts in the relationship between savers (households) and bankers. In particular, in each period, after portfolio decisions but before financial payouts are made, the banker can choose to divert a fraction \( \mu \) of total assets \( Q_t S_{jt} \). The cost is that depositors can then force the intermediary into bankruptcy and recover the
remaining fraction \((1 - \mu)\) of the assets. Because rational households recognize the bank’s option to divert assets, they will only be willing to supply them funds conditional on an incentive compatibility constraint: the continuation value of operating the bank, \(V_{jt}\), cannot be less than the outside option:

\[
V_{jt} \geq \mu Q_t S_{jt}.
\]  

(6)

Because bankers may face a binding financial constraint in their ability to obtain deposits from households (that is, when (6) holds with equality), they will retain earnings in order to accumulate net worth until escaping the constraint indefinitely. To allow financial frictions to remain a relevant threat over time, I assume that bankers have finite lifetimes. Specifically, each period a banker continues operating with exogenous i.i.d. probability \(\theta\) which is independent of history.\(^4\) This mechanism motivates dividend payouts upon exit, while the financial constraint is still binding or (with some positive probability) expected to be binding in the near future.

Then, a mass \((1 - \theta)\) of bankers exit each period (and become workers), and are replaced by an equal mass of workers that become new bankers, keeping the mass of agents in each occupation constant over time.\(^5\)

Therefore, given that the bank pays dividends only upon exit, the objective of bank \(j\) at the end of period \(t\) is to maximize the expected present value of terminal wealth:

\[
V_{jt} = \mathbb{E}_t \left\{ \sum_{i=1}^{\infty} (1 - \theta)\theta^{i-1} \Lambda_{t,t+i} N_{j,t+i} \right\}.
\]  

(7)

where \(\Lambda_{t,t+i}\) is the appropriate stochastic discount factor because the banker is ultimately a member of the household.

Switching to a recursive formulation, the bank problem can be written as follows:

\[
V_{jt} = \max_{\{S_{jt}, D_{jt}\}} \{ \mathbb{E}_t \Lambda_{t,t+1} [(1 - \theta) N_{j,t+1} + \theta V_{j,t+1}(N_{j,t+1})] \}
\]

subject to

\[
V_{jt} \geq \mu Q_t S_{jt}
\]  

(8)

\[
Q_t S_{jt} = D_{jt} + N_{jt}
\]  

(9)

\[
N_{j,t+1} = R^K_{t+1} Q_t S_{jt} - R_t D_{jt}
\]  

(10)

\(^4\)This implies an average survival time equal to \(1/\theta\).

\(^5\)The retained capital of exiting bankers \((1 - \theta) [R^K_t - R_t] Q_t A_{jt} + R_t N_{jt}\) is transferred back to households (as “dividends”), which in turn use part of it to provide new bankers with small start-up funds. These transactions are accounted for in the households’ budget constraint through the term \(\Sigma_t\). See Appendix I for more details on these transactions and how they wash out in deriving the aggregate resource constraint of the economy.
In order to solve the dynamic program, we guess (and then verify) that the value function is linear in net worth, \( V_{jt}(N_{jt}) = \psi_t N_{jt} \). Combining (9) and (10) to eliminate deposits \( D_{jt} \), and given the conjectured value function, the problem can be conveniently written as follows:

\[
V_{jt}(N_{jt}) = \psi_t N_{jt} = \max_{\{S_{jt}\}} \{ \mu_{K,t} Q_t S_{jt} + \mu_{N,t} N_{jt} \}
\]

subject to \( \psi_t N_{jt} \geq \mu Q_t S_{jt} \)

where

\[
\mu_{K,t} = E_t \Lambda_{t,t+1} \Omega_{t+1} (R^K_{t+1} - R_t) \tag{11}
\]

\[
\mu_{N,t} = E_t \Lambda_{t,t+1} \Omega_{t+1} R_t \tag{12}
\]

\[
\Omega_t = (1 - \theta) + \theta \psi_t \tag{13}
\]

Note that \( \psi_t = \frac{V_{jt}}{N_{jt}} \) corresponds to bank’s \( j \) value per unit of net worth and can be interpreted as the “Tobin’s Q” ratio of the franchise. The variable \( \Omega_t \) represents the value to the bank of an extra unit of net worth in, which equals \( \psi_t \) if the bank survives (with probability \( \theta \)), and one otherwise (probability \( 1 - \theta \)). Given the financial constraint, the Tobin’s Q ratio \( \psi_t \) will always exceed unity (see Gertler and Kiyotaki, 2015). We can think of \( \mu_{N,t} \) and \( \mu_{K,t} \) as the expected discounted marginal cost of funds, and the excess marginal return on assets over liabilities, respectively.

Letting \( \xi_t \) be the multiplier on the incentive compatibility (IC) constraint, the first-order and slackness conditions are:

\[
\mu_{K,t} = \mu \xi_t \tag{14}
\]

\[
\xi_t [\psi_t N_{jt} - \mu Q_t S_{jt}] = 0 \tag{15}
\]

Combining equations (11)-(14) it can be shown that:

\[
\psi_t = \frac{\mu_{N,t}}{1 - \xi_t} = \frac{E_t [\Lambda_{t,t+1}] [1 - \theta + \theta \psi_{t+1}] R_t}{1 - \xi_t} \tag{16}
\]

---

\(^6\)Gertler and Kiyotaki (2015), Akinci and Queralto (2015), and Bocola (2015) follow a similar strategy. Ultimately, this linearity result implies that banks’ heterogeneity does not affect aggregate dynamics, and therefore, we do not need to keep track of the wealth distribution. This feature of the model helps to maintain tractability in the numerical analysis below.
From equations (14) and (15), when the IC constraint binds, $\xi_t > 0$ and:

$$
\phi_t \equiv \frac{Q_t S_{jt}}{N_{jt}} \begin{cases} \frac{\psi_t}{\mu} = \frac{\mu N_{jt}}{\mu(1 - \xi_t)} \equiv \bar{\phi}_t & \text{if constraint is binding, } \xi_t > 0 \\ \frac{\psi_t}{\mu} = \frac{\mu N_{jt}}{\mu} & \text{if constraint is slack, } \xi_t = 0. \end{cases}
$$

(17)

The above expressions lie at the heart of the GK financial accelerator. Equation (16) tells us that the marginal value of wealth ($\psi_t$) is increasing in the IC multiplier ($\xi_t$). Furthermore, even when the constraint is slack ($\xi_t = 0$), we have $\psi_t > 1$ because the bank recognizes the possibility of a binding constraint in subsequent periods and would like to hold “precautionary capital”.

From equation (17), when the IC constraint binds there is an endogenous upper bound $\bar{\phi}_t$ on the bank’s leverage ratio $\phi_t$. Notice that from the linearity property of (17), we can easily aggregate to get $Q_t S_t \leq \bar{\phi}_t N_t$: total credit provided by the banking sector depends positively on aggregate net worth. A negative shock to banks’ wealth triggers an endogenous decline in their lending capacity, which reinforces itself in subsequent periods through fire-selling asset prices declines and the law of motion for aggregate net worth (equation (20), to be described below).

Also note that combining (11) and (14) we can write an Euler equation of the form:

$$
\mathbb{E}_t \left[ \Lambda_{t,t+1} \Omega_{t+1} R_{t+1}^K \right] = \mathbb{E}_t \left[ \Lambda_{t,t+1} \Omega_{t+1} R_t \right] + \mu \xi_t.
$$

(18)

Two features are noteworthy in equation (18). First, a binding leverage constraint introduces a wedge between the expected discounted return on loans and the risk-free rate on deposits. The tighter the constraint binds (the higher $\xi_t$), the more financial distress in the economy (the higher $\mu_{K,t}$) and the higher the lending-deposit spread (via (11)). Second, the bank’s stochastic discount factor (SDF) is “augmented” by the factor $\Omega_{t+1}$, which is ultimately a function of the bank’s leverage ratio. Intuitively, if the leverage constraint was never to bind in the future, we would have $\xi_{t+i} = 0$ and $\psi_{t+i} = \Omega_{t+i} = 1$ for all $i \geq 0$, and equation (18) would collapse to the neoclassical benchmark.

Defining the augmented SDF of banks as $\hat{\Lambda}_{t,t+1} = \Lambda_{t,t+1} \Omega_{t+1}$ we obtain an intuitive expression for the risk and liquidity premium required by banks as compensation for holding in order to be willing to hold the claims issued by the corporate sector:

$$
\mathbb{E}_t \left[ R_{t+1}^K - R_t \right] = \frac{\mu \xi_t}{\mathbb{E}_t \left[ \hat{\Lambda}_{t,t+1} \right]} - \frac{COV_t(\hat{\Lambda}_{t,t+1}, R_{t+1}^K)}{\mathbb{E}_t \left[ \hat{\Lambda}_{t,t+1} \right]}.
$$

(19)

Expected excess returns on capital may arise for two reasons. First, as in canonical equity premium models, high excess returns reflect a fair compensation that bankers demand for holding assets whose payouts covary negatively with the (augmented) SDF. Second, positive spreads may reflect the inability of bankers to increase their portfolio of assets (raise new profitable lending) due to the leverage constraint ($\xi_t > 0$).

The financial intermediaries characterization is closed with the law of motion for bankers’ net worth. Aggregate net worth in each period is the sum of the net worth of “surviving” bankers
(N_t^s) and the net worth of “new” bankers (N_t^n). Since the fraction of surviving bankers from $t-1$ to $t$ is $\theta$, we have:

$$N_t^s \equiv \theta N_t = \theta \left[ (R_t^K - R_{t-1})Q_{t-1}S_{t-1} + R_{t-1}N_{t-1} \right].$$

As noted earlier, new bankers receive a start-up transfer of funds from households corresponding to a small share $\iota$ of the value of the assets that bankers have intermediated in the previous period:

$$N_t^n \equiv \iota(1 - \theta)Q_{t-1}S_{t-1}.$$ 

Accordingly, the aggregate net worth of banks evolves according to:

$$N_t = N_t^s + N_t^n = \{ \theta \left[ (R_t^K - R_{t-1})Q_{t-1}S_{t-1} + R_{t-1}N_{t-1} \right] + \iota(1 - \theta)Q_{t-1}S_{t-1} \}.$$  \hspace{1cm} (20)

Equation (20) can be conveniently rewritten as:

$$N_t = \theta R_t^K Q_{t-1}S_{t-1} + P_{t-1}$$ \hspace{1cm} (21)

$$P_t = \theta \left[ R_t(N_t - Q_tS_t) \right] + \iota(1 - \theta)Q_tS_t$$ \hspace{1cm} (22)

where the state variable $P_{t-1}$ measures the interest deposits that bankers pay to households at the beginning of the period (net of startup transfers), and is sufficient to keep track of the evolution of aggregate net worth.

### 3.3 Final Goods Producers

There is a large set of competitive final goods producers that combine labor and capital to produce the single final good in the economy, using a constant returns-to-scale Cobb-Douglas technology. I introduce two relatively non-standard features on these agents. First, they need external financing from banks (described above) to purchase new physical capital from capital producers (to be described below). At the beginning of each period, they issue perfectly state-contingent claims to bankers in exchange for funds, which are used to purchase the capital to be used in production in the current period. A no-arbitrage condition implies that the latter two transactions are made at the capital market price $Q_t$. That is, in equilibrium, final goods firms pay $Q_t$ for each unit of physical capital to capital producers. In turn, for each security, bankers also pay $Q_t$ to final goods firms. Second, after purchasing the capital stock, the realization of the (aggregate) “quality of capital” shock $\Psi_t$ determines the effective amount of physical capital available for production. Therefore, the realized return on the securities issued by firms and purchased by banks is given by:

$$R_{t+1}^K = \left[ \frac{Z_{t+1} + (1 - \delta)Q_{t+1}}{Q_t} \right] \Psi_{t+1}$$ \hspace{1cm} (23)
where $Z_t$ denotes the net revenue from production per unit of effective capital. Anticipating the labor market clearing condition, profit maximization gives rise to the following first-order conditions:

$$
W_t = (1 - \alpha) \frac{Y_t}{H_t} \quad \text{(24)}
$$

$$
Z_t = \alpha \frac{Y_t}{\Psi_t K_{t-1}}. \quad \text{(25)}
$$

with

$$
Y_t = A_t (\Psi_t K_{t-1})^\alpha H_t^{1-\alpha}. \quad \text{(26)}
$$

where $A_t$ is a standard technology shock. Note that in the aggregate, $Q_tK_t$ is the total value of capital acquired and $Q_tS_t$ is the total value of claims against this capital (total credit in the economy). Then, by arbitrage, the capital market clearing condition implies $K_t = S_t$.

### 3.4 Capital Producers

There is an arbitrary large set of competitive capital producers that operate the technology to increase the economy-wide stock of capital. At the end of each period $t$, capital producers purchase from final goods producers the stock of undepreciated capital already used in production $(1 - \delta)\Psi_t K_{t-1}$, repair it, and then combine it with new investment $I_t$ to produce new capital $K_t$. The latter will be available for production next period. The newly produced capital is then sold to firms and any profit is transferred back to the households. Since the marginal rate of transformation (the repair stage) from previously used capital to new capital is unity, the market price of new and used capital are both equal to $Q_t$. Accordingly, the period $t$ profit of capital producers is given by:

$$
\Pi^K_t = Q_t K_t - Q_t (1 - \delta) \Psi_t K_{t-1} - I_t \quad \text{(27)}
$$

where the market price of capital $Q_t$ is taken as given. The implied law of motion for capital is given by:

$$
K_t = (1 - \delta) \Psi_t K_{t-1} + \Gamma \left( \frac{I_t}{\Psi_t K_{t-1}} \right) \Psi_t K_{t-1} \quad \text{(28)}
$$

where the function $\Gamma(.)$ is of the form:

$$
\Gamma(x) = a_1 x^{1-\varrho} + a_2. \quad \text{(29)}
$$

The constants $a_1$ and $a_2$ are set in order to ensure that in the steady state $\Gamma(\delta) = \delta$ and $\Gamma'(\delta) = 1$ (so that $Q = 1$). The parameter $\varrho \in [0, 1]$ governs the amount of adjustment costs in
the economy through the elasticity of Tobin’s $Q$ with respect to the investment-capital ratio, as in Bernanke, Gertler, and Gilchrist (1999).

Capital producers choose $I_t$ in order to maximize (27) subject to (28). The first-order condition is:

$$Q_t = \left[ \Gamma' \left( \frac{I_t}{\Psi_t K_{t-1}} \right) \right]^{-1} = \left[ \frac{I_t}{\delta \Psi_t K_{t-1}} \right]^{\varphi}. \quad (30)$$

### 3.5 Government

The government finances its purchases of the final good ($G_t$) by levying lump-sum taxes ($T_t$), and by issuing one-period risk-free bonds ($B_t$). Lump-sum taxes adjust every period to compensate any difference between spending and net bonds issuance. The budget constraint is of the form:

$$G_t + R_{t-1} B_{t-1} = T_t + B_t. \quad (31)$$

The ratio of government spending-to-output evolves exogenously as follows:

$$G_t = \left( 1 - \frac{1}{g_t} \right) Y_t \quad (32)$$

where the spending shock $g_t$ follows an AR(1) process.

### 3.6 Market Clearing and Driving Forces

Appendix I shows that the aggregate resource constraint can be written as follows:

$$Y_t = C_t + I_t + G_t. \quad (33)$$

The labor market clearing condition requires that hours supplied by households equal hours demanded by final producers. In turn, the capital markets clearing condition states that:

$$K_t = S_t.$$

The driving forces in the model are the TFP shock $A_t$, the labor wedge shock $\varphi$, the government-spending shock $g_t$, and the quality of capital shock $\Psi_t$. These variables follow stationary AR(1) processes in logs. The competitive equilibrium and the full system of equilibrium conditions are described in Appendix II.

---

7The required constants in (29) are $a_1 = \frac{\delta \varphi}{1-\varphi}$ and $a_2 = -\frac{\delta \varphi}{1-\varphi}$, expressions that are used to obtain the second equality in (30).
4 Prototype RBC Model with Wedges

In the spirit of Chari, Kehoe, and McGrattan’s (2007) business cycle accounting method, in what follows I show that the GK financial friction described above maps into an agnostic intertemporal wedge distorting the capital Euler equation in a prototype RBC model. In turn, as is well-known, the preference shock to the disutility of work maps into a labor wedge distorting the intratemporal optimality condition between the household’s marginal rate of substitution and firm’s marginal product of labor.

The productive sector in the prototype economy with wedges is the same as in the baseline model. The difference is that there are no financial intermediaries. Rather, households provide loans directly to final goods producers. As before, households can also save by purchasing government securities. Therefore, the representative household solves the following problem:

\[
\max_{\{C_t, H_t, L_t, B_t\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[ \frac{C_t^{1-\gamma}}{1-\gamma} - \frac{H_t^{1+\zeta}}{1+\zeta} \right]
\]

subject to \[C_t + L_t + B_t = (1 - \tau_t^H)W_tH_t + (1 - \tau_t^K)R_t^K L_{t-1} + R_{t-1}B_{t-1} + \Sigma_t + T_t \tag{34}\]

where \(C_t\) is consumption, \(H_t\) is hours worked that earn the real wage \(W_t\), \(L_t\) is loans to final goods producers that earn the state-contingent real gross rate \(R_t^K\), \(B_t\) is savings in government bonds that pay the non-state-contingent real gross return \(R_t\), \(\Sigma_t\) are real dividends from the ownership of firms, and \(T_t\) are lump-sum transfers received from the government. \(\tau_t^H\) and \(\tau_t^K\) are exogenous processes that resemble taxes on labor and capital income, and play the role of labor and investment wedges, respectively. Redefining the wedges as \(\eta_t^H = 1 - \tau_t^H\) and \(\eta_t^K = 1 - \tau_t^K\), the first-order conditions are:

\[
\chi H_t = \eta_t^H C_t^{-\gamma} W_t \tag{35}
\]

\[
1 = \mathbb{E}_t \Lambda_{t,t+1} R_t \tag{36}
\]

\[
1 = \mathbb{E}_t \Lambda_{t,t+1} \eta_{t+1} R_{t+1}^K \tag{37}
\]

where \(\Lambda_{t,t+i} \equiv \beta^i \left( \frac{C_{t+i}}{C_t} \right)^{-\gamma}\) as in the baseline model. In equilibrium, total credit from households to final goods producers are \(L_t = Q_t K_t\).

The government finances its purchases of the final good \((G_t)\) by levying capital \((\tau_t^K)\) and labor \((\tau_t^H)\) income taxes, and by issuing one-period risk-free bonds \((B_t)\). Any difference between spending, tax revenues, and net bonds issuance is rebated back to households in a lump-sum fashion via \(T_t\). The budget constraint is of the form:

\[
G_t + T_t + R_{t-1}B_{t-1} = \tau_t^H W_t H_t + \tau_t^K R_t^K L_{t-1} + B_t. \tag{38}
\]

Tax rates \(\tau_t^K\) and \(\tau_t^H\) follow exogenous AR(1) processes. Public spending evolves as described in equation (32).
Mapping from frictions to wedges

Comparing equation (2) in the baseline model with microfounded frictions with equation (35) in the prototype RBC model, it is clear that the labor wedge maps as follows:

\[ \eta^H_t = 1 - \tau^H_t = \frac{1}{\varphi_t}. \]  

(39)

Combining equations (11)-(14) in the baseline model, we can write:

\[ 1 = \mathbb{E}_t \left[ \Lambda_{t,t+1} \left( \frac{1 - \theta + \theta \psi_{t+1}}{\mu \xi_t + \psi_t (1 - \xi_t)} \right) R_{t+1}^K \right]. \]  

(40)

Comparing the resulting Euler equation for capital in the baseline model (40) with equation (37) in the prototype RBC model, the investment wedge maps into the financial friction as follows:

\[ \eta^K_{t+1} = 1 - \tau^K_{t+1} = \left( \frac{1 - \theta + \theta \psi_{t+1}}{\mu \xi_t + \psi_t (1 - \xi_t)} \right). \]  

(41)

The agnostic intertemporal investment wedge in the prototype model is a function of the banks’ Tobin’s Q ratio and the multiplier on the occasionally binding constraint. Intuitively, if the leverage constraint in the model with explicit financial frictions is never to bind (so that the multiplier on the incentive constraint \( \xi_t = 0 \ \forall t \), and the banks’ marginal utility of net worth \( \psi_t = 1 \ \forall t \)), then \( \eta^K_t = 1 \) (and \( \tau^K_t = 0 \)) \( \forall t \). In other words, both models converge to the frictionless RBC benchmark, in which the expected discounted returns on all assets in the economy are equalized:

\[ 1 = \mathbb{E}_t [\Lambda_{t,t+1} R_t] = \mathbb{E}_t [\Lambda_{t,t+1} R^K_{t+1}]. \]

This result will be used in the quantitative section below, in order to provide a direct structural estimate of the impact of the financial friction on business cycle fluctuations.

5 Quantitative Results

This section presents a series of numerical experiments that shed light on the main features of the model dynamics, its ability to account for the U.S. business cycles, and the role of financial factors (the “financial shock” \( \Psi_t \) and the GK-type financial friction itself) during the Great Recession. The first subsection briefly describes the computational strategy. The second subsection presents the calibration/estimation strategy and results. Next, I present several experiments to illustrate the dynamics of the model. Finally, the last subsection reports the main results of the paper.
5.1 Computational Strategy

The empirical strategy combines both calibrated and estimated parameters. Some parameters are calibrated before the estimation step because the likelihood function is not informative about their value. On the other hand, estimating the fully nonlinear model subject to the occasionally binding financial constraint is computationally challenging because it requires the solution of the nonlinear model to be computed for a large number of parameter vectors. Instead I estimate the parameters using a log-linearized approximation of the model equilibrium conditions, and characterize the posterior distribution using a Random Walk Metropolis-Hastings algorithm, as described in An and Schorfheide (2007). Conditional on the estimated parameter vector, I solve the model and extract the underlying states and structural shocks enforcing the occasionally binding constraint by means of fully nonlinear methods.

In particular, the model is solved using a global method based on Chebyshev approximations of the decision rules along the lines of Judd (1992). Given the minimum set of state variables associated with the DSGE model, $S_t$, the solution algorithm requires to choose a grid of points $G = \{S_1, ..., S_M\}$ in the model’s state-space and determining the coefficients on the Chebyshev polynomials by minimizing the unweighted sum of squared residuals associated with the Euler equations of the model. Following Aruoba, Cuba-Borda, and Schorfheide (2016), the solution algorithm involves two non-standard tools. First, because the occasionally binding constraint potentially introduces kinks in the policy functions, I use a piecewise smooth representation of the approximated decision rules. Second, the solution grid $G$ is chosen using an iterative procedure based on a simulation-based clustered-grid-algorithm (CGA) first proposed by Judd, Maliar, and Maliar (2010). The solution algorithm and the accuracy of the numerical approximations are described in detail in Appendix III. Given the model solution, I uncover the hidden state and disturbances that best fits the data over the sample using the Bootstrap Particle Filter, as described in Herbst and Schorfheide (2016).

5.2 Model Estimation

Table 1 reports the parameters and targeted steady state values used in the experiments. The model includes four conventional parameters ($\beta, \gamma, \alpha, \delta$), for which I choose standard values and steady state targets used in related studies. I target an steady state risk-free annual interest rate of 4%. I set the inverse of the intertemporal elasticity of substitution $\gamma$ equal to one, which implies log utility. I set the capital share $\alpha = 0.33$ and the quarterly depreciation rate of capital $\delta = 0.025$, as in GK. There are three parameters that are specific to bankers. First, the divertable share $\mu$ is calibrated to generate a frequency of financial crises of about 0.5% (two systemic financial crisis in 100 years). Second, the start-up share $i$ is set to target an ergodic mean leverage ratio of four, which is the typical value used in the literature (Gertler and Karadi, 2011). Finally, a value of $\theta = 0.96$ for the banks’ survival probability is used, following Bocola (2016) who uses a similar GK model subject to an occasionally binding constraint.

The remaining parameters which govern the dynamics of the model are estimated using Bayesian techniques. Namely, the Frisch elasticity of labor supply, the adjustment costs, and the persistence and standard deviations of all exogenous AR(1) processes. The sample period is 1954:I-2015:IV. The observables used are real output, real investment, hours worked, and a measure
of the credit spread. Real variables are in per capita terms. Following Del Negro, Schorfheide, Smets, and Wouters (2004), real variables are constructed scaling nominal variables by the GDP deflator. Consumption includes nondurable goods and services, while investment includes durable consumption and gross private domestic investment. Because I work with a closed economy model, the data on output excludes net exports. The spread measure is the difference between the Moody’s seasoned Baa and Aaa corporate bond yield. All observables are HP-filtered, and the associated HP-cycle time series are matched with the model-implied variables in percent deviation from their respective ergodic means.

Table 1: DSGE Model Parameters

<table>
<thead>
<tr>
<th>Calibrated</th>
<th>Description</th>
<th>Source or Target</th>
<th>Value</th>
<th>Posterior [5% 95%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>Disc. factor</td>
<td>$R = 4%$ annual</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Risk aversion</td>
<td>Gertler and Karadi (2011)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>Depreciation</td>
<td>Gertler and Karadi (2011)</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Capital share</td>
<td>Gertler and Karadi (2011)</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>$g$</td>
<td>Steady State $G/Y$</td>
<td>$G/Y$ Avg. 1955-2015</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>Divert share</td>
<td>Freq. Fin. Crises 0.5%</td>
<td>0.256</td>
<td></td>
</tr>
<tr>
<td>$\iota$</td>
<td>Start-up share</td>
<td>$TEV = 4.$</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>Survival rate</td>
<td>Bocola (2016)</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>$\zeta$</td>
<td>Inverse Frisch</td>
<td>Gamma(1.4, 1)</td>
<td>2.40</td>
<td>[1.20-3.50]</td>
</tr>
<tr>
<td>$\varrho$</td>
<td>Adjustment Cost</td>
<td>Beta(0.25, 0.1)</td>
<td>0.03</td>
<td>[0.01-0.05]</td>
</tr>
<tr>
<td>$\rho_A$</td>
<td>AR TFP</td>
<td>Beta(0.5, 0.2)</td>
<td>0.75</td>
<td>[0.68-0.83]</td>
</tr>
<tr>
<td>$\rho_\varphi$</td>
<td>AR Labor Wedge</td>
<td>Beta(0.5, 0.2)</td>
<td>0.85</td>
<td>[0.81-0.89]</td>
</tr>
<tr>
<td>$\rho_g$</td>
<td>AR Gov. Spending</td>
<td>Beta(0.5, 0.2)</td>
<td>0.78</td>
<td>[0.64-0.92]</td>
</tr>
<tr>
<td>$\rho_q$</td>
<td>AR Quality K</td>
<td>Beta(0.5, 0.2)</td>
<td>0.16</td>
<td>[0.03-0.29]</td>
</tr>
<tr>
<td>$100\sigma_A$</td>
<td>Std. TFP</td>
<td>InvGamma(1, 2)</td>
<td>0.52</td>
<td>[0.46-0.58]</td>
</tr>
<tr>
<td>$100\sigma_\eta$</td>
<td>Std. Labor Wedge</td>
<td>InvGamma(1, 2)</td>
<td>2.27</td>
<td>[1.27-3.21]</td>
</tr>
<tr>
<td>$100\sigma_g$</td>
<td>Std. Gov. Spending</td>
<td>InvGamma(1, 2)</td>
<td>0.213</td>
<td>[0.15-0.28]</td>
</tr>
<tr>
<td>$100\sigma_q$</td>
<td>Std. Quality K</td>
<td>InvGamma(1, 2)</td>
<td>0.211</td>
<td>[0.19-0.23]</td>
</tr>
</tbody>
</table>

Notes: Priors and posteriors based on 100,000 draws from the Metropolis-Hastings algorithm (discarding the first 50,000) applied to the log-linearized model in which the constraint always binds. Sample period: 1954:I-2015:IV.

Fairly agnostic priors are used for the autoregressive processes. I assume a Beta distribution with mean 0.5 and standard deviation 0.2 for the autoregressive coefficients. The prior for the inverse Frisch elasticity is a Gamma distribution centered at 1.4 ($\approx 1/0.72$), which is the baseline figure used by Rios-Rull et al (2012) who extensively study appropriate values for this parameter using U.S. data in the context of DSGE models. The prior for the elasticity of Tobin’s Q with respect to the investment-to-capital ratio (the adjustment cost parameter) is

8Similar results can be found using the spread between the Baa corporate bond yield and the yield on long-term Treasury bonds, or using the Gilchrist and Zakrajsek’s (2012) excess bond premium. I choose to use the Baa-Aaa spread because it allows me to extend the data until 1947 (the beginning of the NIPA sample), and have several years of pre-sample data that is useful to train the filters in the experiments below. The effective sample used is 1954:I-2015:IV.
assumed to be a Beta distribution centered at 0.25, that is, the value calibrated by Bernanke, Gertler, and Gilchrist (1999). Overall the data proved to contain significant information about the estimated parameters, which is reflected in the considerably different posterior densities relative to the chosen priors.

5.3 Model Dynamics

5.3.1 Goodness of fit

Table 2 compare some selected model-implied second moments with their analogs in the data over the sample period 1954:I-2015:IV. Overall, the model is able to reproduce several key volatilities and correlations that are consistent with observed U.S. business cycle moments. The model slightly underpredict the absolute volatility of output (1.5%) relative to the data (1.7%), because it generates less volatility on investment. However, the relative standard deviation of output over investment is reasonably well-aligned around three in both the model and the data, consistent with standard results in the business cycle literature. The model also does a fair job in capturing the moments of the financial variable included in estimation (spread), and also those for consumption which is not observable. As expected, the model is able to produce countercyclical spreads, but only half the size observed in the data. As explained above, positive spreads may arise in the model (and in the data) because of two reasons: first, a standard countercyclical risk-premium, and second, the financial constraint. As will become clear below, the model presented here displays very low risk premium (a canonical result in Euler-equation based macro frameworks), preventing the model from generating the full countercyclicality of spreads observed in the data.

Table 2: Second Moments

<table>
<thead>
<tr>
<th></th>
<th>$\sigma(X)$</th>
<th>$\sigma(X)/\sigma(Y)$</th>
<th>$\text{corr}(X,Y)$</th>
<th>$\text{corr}(X, X_{-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td>1.73</td>
<td>1</td>
<td>1</td>
<td>0.87</td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>1.53</td>
<td>1</td>
<td>1</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>Output ($Y$)</strong></td>
<td>0.85</td>
<td>0.49</td>
<td>0.80</td>
<td>0.84</td>
</tr>
<tr>
<td><strong>Consumption</strong></td>
<td>0.91</td>
<td>0.60</td>
<td>0.67</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Investment</strong></td>
<td>5.79</td>
<td>3.34</td>
<td>0.94</td>
<td>0.86</td>
</tr>
<tr>
<td><strong>Hours</strong></td>
<td>1.88</td>
<td>1.08</td>
<td>0.85</td>
<td>0.92</td>
</tr>
<tr>
<td><strong>Spread</strong></td>
<td>0.07</td>
<td>0.04</td>
<td>-0.58</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Notes: Model-implied moments compared to data. Sample period: 1954:I-2014:IV.

5.3.2 Sample Decision Rules

In order to illustrate how the occasionally binding constraint drives an asymmetry in the economy, this subsection presents decision rules for selected variables and the three most important shocks analysed in the model, the technology shock, the labor wedge shock, and the disturbance to the quality of capital.

Figure 1 show the results for slices of the decision rules in which only one exogenous state variable varies along a fine grid. The range of the grids is wide enough (+ and - three standard
deviations) to cover cases in which the IC constraint is both slack and binding. Each column of the figure moves along the corresponding exogenous state variable (TFP, labor wedge, quality of capital), while keeping all other states at their ergodic means. Each row contains the decision rules for investment ($I_t$), the bank Tobin’s Q ratio ($\psi_t = V_t/N_t$), and the spread ($E_t [R^K_{t+1} - R_t]$).

Figure 1: Policy Functions for Selected Variables

Notes: In each panel, only one exogenous state variable varies on the horizontal axis. The other state variables are fixed at their ergodic means. “Nonlinear” corresponds to the piecewise nonlinear solution algorithm described in detail in Appendix III. “OccBin” is the piecewise linear solution using the OccBin toolkit described in Guerrieri and Iacoviello (2015). “Linear” corresponds to a first-order perturbation solution which does not respect the occasionally binding constraint.

Each panel includes the policy functions for three different solution methods. “Nonlinear” corresponds to the piecewise nonlinear solution algorithm described in detail in Appendix III. “OccBin” is the piecewise linear solution using the OccBin toolkit described in Guerrieri and
Iacoviello (2015). 9 “Linear” corresponds to a first-order perturbation solution which does not respect the occasionally binding constraint. The comparison with the other solution methods commonly used in the literature sheds light on the importance of the nonlinear solution in capturing both the kinks in the policy functions (not captured in the linear solution) and precautionary motives (not captured in the piecewise linear solution).

Figure 1 shows that the economy experience a binding financial constraint after large negative (positive) TFP (labor wedge) shocks. Not surprisingly, the economy tends to switch into a “financial crisis” regime after relatively small negative quality of capital shocks, given the fact that these disturbances directly affect the net worth of the banking sector, hereby increasing the leverage ratio and activating the GK financial accelerator. Likewise, for high positive values of both TFP and quality of capital shocks (or negative values for the labor wedge shock) the economy is in a slack regime. In this regime, the level of banks’ net worth is high, the leverage ratio tends to be low relative to the mean of the ergodic distribution and relative to the leverage limit imposed by the financial friction ($\bar{\phi}_t$ in equation (17)). Because the multiplier of the incentive constraint is equal to zero and the model generates a tiny amount of risk premium, the lending-deposit interest rate spread is around zero (see equation (19)). Cheap credit fuels a rise in both investment and asset prices, and the economy booms.

Notice that the marginal value of wealth in the slack regime is always lower than in the binding regime. Intuitively, due to the bank’s financial constraint, the bankers’ marginal value of an extra unit of net worth is higher during periods of financial distress. It is also noteworthy that the bank’s Tobin’s Q ratio in the fully nonlinear decision rule is always greater (or equal) than the values under the piecewise linear (OccBin) solution in both regimes. This is precisely the “precautionary capital” motive embedded in the model: under the fully nonlinear solution, even under a slack regime, banks realize that future shocks might push them into the leverage limit. Hence, the marginal value of net worth is higher in every state of the economy. This is an important mechanism because it implies the economy will spend less time under the binding regime, as is clear in the last row of the figure that shows a wider range in which the spread (a direct function of the multiplier on the IC constraint, see again equation (19)) is near zero. Therefore, the model has the potential to capture the idea of financial crises being relatively rare events nested within typical business cycles.

---

9Guerrieri and Iacoviello (2015) provide a fast an efficient Dynare-based toolkit (OccBin) to solve dynamic models with occasionally binding constraints. They adapt a first-order perturbation approach and applies it in a piecewise fashion. Importantly, the piecewise solution is not just linear - with two different set of policy functions depending on whether the constraint is binding or not - but rather, it can be highly nonlinear. The dynamics in each regime depends on how long one expects to be in that regime, which in turn depends on the state vector. This interaction produces the high nonlinearity, and allows to capture the kinks in the decision rules accurately. However, there is a limitation of such a method. Just as any linear approximation method, the algorithm discards all information regarding the realization of future shocks. Therefore, the algorithm is not able to capture precautionary behaviour linked to the possibility that a constraint become binding in the future, as a result of shocks yet unrealized. I use the piecewise linear solution in order to find a reasonable initial guess for my fully nonlinear algorithm.
5.3.3 Crises Experiments: Impulse Responses

This section presents several experiments to show how the model dynamics work. Figure 2 shows the responses of key variables in the model to the three main disturbances: the TFP shock, the labor wedge shock, and the “financial” shock. In each case, the direction of the shock is designed to generate a recession, and the size of the impulse is two standard deviations, which is consistent with the size of the filtered innovations uncovered during the Great Recession (see below). The rows in the figure present the responses of output, consumption, investment, the banking sector’s leverage ratio, and the spread.

Figure 2: Impulse Responses

Notes: Responses to negative 2-standard deviation shocks. All variables are in percent deviation from the baseline unshocked path.

The negative shocks generate an immediate increase in the leverage ratio until a point in which the financial constraint is activated in every case. Therefore the multiplier on the incentive com-
patibility constraint becomes increasingly positive, implying an increase in the spread between the lending and the risk-free deposit interest rate. In a context of severe financial distress, banks are forced to delever through a protracted cut in lending to the corporate sector in order to escape from the binding constraint, which in turn translate into a significant investment slump at impact. Consumption also falls persistently in all cases but in a much smoother fashion. While the initial recession is much smaller under the quality of capital shock, the economy tends to stay below trend for a larger period.

5.4 Inference of Unobserved States from a Nonlinear Filter

5.4.1 Quantifying the role of the financial shock

In this section I use a nonlinear filter to backout the hidden state and structural innovations that best fits the U.S. data over the sample 1954:I-2015:IV, under the lens of the model subject to the occasionally binding financial constraint. The DSGE model has a nonlinear state-space representation of the form

\[ y_t = \Psi(s_t) + u_t, \quad u_t \sim N(0, \Sigma_u) \]  
\[ s_t = \Phi(s_{t-1}, \epsilon_t), \quad \epsilon_t \sim N(0, \Sigma_\epsilon) \]

where \( y_t \) is the vector of observables in period \( t \), \( s_t \) stacks the hidden state vector, \( \epsilon_t \) is the vector of structural shocks, while \( u_t \) are measurement errors. (42) is the measurement equation that links the model state variables with the observable time series used to inform the model. (43) is the transition equation given by the piecewise nonlinear solution of the model represented here by the nonlinear function \( \Phi(\cdot) \). I use the bootstrap particle filter to conduct inference about the unobserved state \( (s_t) \) and shocks \( (\epsilon_t) \) over the sample. The details of the algorithm can be found in Herbst and Schorfheide (2016).

Figure 3 presents the filtered i.i.d. innovations \( (\epsilon_t) \) uncovered from the filter, using the same observable variables as in the estimation step (real output, real investment, per capita hours, and the spread). By construction, feeding these structural shocks back to the nonlinear state-space system recovers the observable data used to inform the model (up to a small measurement error assumed to be 10% of the sample variance of the respective observable time series). Likewise, counterfactual experiments can be ran by turning on and off one or more of these driving forces at a time.

In general, the sequence of shocks extracted from the filter are consistent with historical accounts and previous findings: productivity shocks are procyclical (Basu and Fernald, 2000), while labor wedge shocks are highly countercyclical (Hall, 1997; Shimer, 2009; Karabarbounis, 2014), and government spending shocks tend to be less important for business cycle fluctuations in the U.S. (Chari, Kehoe, and McGrattan, 2007). As expected, the quality of capital or “financial” shock seems to be specially relevant during the Great Recession, and also during previous recessions arguably caused by other unmodeled disturbances (e.g. the oil price shocks during
the seventies). With the exception of some spikes around particular recession episodes, all the innovations tend to be within two standard deviations.

Figure 3: Structural Innovations (in number of standard deviations)

![TFP innovation](image1)

![Labor Wedge innovation](image2)

![Government Spending innovation](image3)

![Quality of Capital innovation](image4)

**Notes:** Structural innovations scaled by the standard deviation of each shock. The gray areas indicate NBER recession dates.

One way to externally validate the results of the filter is to compare some of the extracted states to analogous objects in the data. Figure 4 compares the underlying autoregressive processes for TFP and the labor wedge implied by the filter, with available empirical counterparts provided by Fernald (2012) and Karabarbounis (2014), respectively. For both TFP and the labor wedge the comovement between the model-implied states and their external data counterparts is striking (correlation of 0.88 and 0.91, respectively), despite the fact of being obtained through completely different methodologies and using different observable time series.
What were the main driving forces behind the Great Recession? Figure 5 presents the evolution of output, investment, hours, and spread in the data and decomposes each quarterly observed realization into the positive (above the x axis) and negative (below the x axis) contributions of the structural shocks in the model. To focus the attention on the Great Recession episode, the figures show a zoom of the period 2000-2015. The long-run analogs starting at the beginning of the sample are available in Appendix IV.

According to the structural model the economic downturn during the Great Recession was mainly a result of negative productivity and labor wedge shocks. On the one side, TFP shocks and to a lesser extent the quality of capital or “financial” shock were key to explain the investment slump at the very onset of the crisis. Interestingly, the positive productivity forces on output were fading out a couple of quarters before the recession, when the economy was
already showing significant signals of distress. On the other side, the labor wedge was key to explain the slow recovery of output an investment. In fact, the long-run shock decompositions reported in Appendix IV are consistent with this same pattern: the TFP process tend to lead the cycle while the labor wedge tend to lag the cycle, a result reminiscent of the findings by CKM in the context of the 1982 crisis.

Not surprisingly, hours worked are overwhelmingly explained by the labor wedge shock, while around half of the spike in the spread in 2008-2009 was due to the effects of the financial shock. However, the incidence of the financial shock in the spread does not translate into a significant effect of this structural force into output and investment, a result that is in part due to the quick mean reversion of the observable spread times series.

Figure 5: Baseline Model: Historical Decomposition: Real Variables and Spread

| Notes: | Output, investment, and hours worked are in quarterly terms, expressed in percent deviations from their ergodic means (solid line). Spread is expressed in annualized percent terms. The solid lines correspond to the model-implied filtered observable variables, which are up to a small measurement error equal to the data described above. The bars decompose each filtered variable into the contributions of each structural shock. The gray areas indicate NBER recession dates. |
5.4.2 Assessing the role of the financial friction

What was the role of the GK financial frictions itself during the Great Recession? The relatively little role of the financial shock does not mean that the effects of the financial friction was not relevant. In fact, the GK financial constraint was binding during the crisis, likely triggered by the full combination of disturbances reported in Figure 3. One way to estimate the direct effect of the friction in the economy is to use the prototype RBC model with wedges and the equivalence result presented in Section 4. In that model, we find that all the endogenous effect of the friction (captured in the baseline model by the marginal value of wealth $\psi_t$, and the multiplier on the occasionally binding constraint $\xi_t$) can be structurally captured by an agnostic exogenous wedge in the spirit of CKM. In contrast to CKM whose shocks are allowed to be correlated, I use the particle filter to uncover the model’s independent innovations, a necessary condition for a meaningful structural interpretation of the shocks.

Figure 6: Pure RBC Model: Historical Decomposition: Real Variables and Spread

Notes: See notes from Figure (5).
Figure 6 presents the results. Overall, the results are consistent with the baseline model. Productivity and labor wedge shocks are still the main drivers of the economy during the boom-bust cycle. However, the financial forces appear to be more important under this metric, explaining around one third of the slump in investment by the end of 2008 and the beginning of 2009.

Another difference in the RBC model with exogenous frictions is that now the investment wedge absorb the full effect not only of the financial friction but also the financial shock. This result can be better explained by combining equations (36) and (37):

\[ 1 = E_t \Lambda_{t,t+1} R_t = E_t \Lambda_{t,t+1} \eta^K_{t+1} R^K_{t+1}. \]

Whenever there is a difference between the risky and the riskless return in the economy, the model can explain it directly through the investment wedge \( \eta^K_{t+1} \) or through the financial shock embedded in \( R^K_{t+1} \). In the prototype RBC model, the data tend to favor the direct investment wedge effect because the financial shock imply a sudden decrease in the supply of capital in the economy, causing a counterfactual increase in the price of capital (\( Q_t \)) during the recession.

6 Conclusions

This paper studies quantitatively the role of financial factors in U.S. business cycle, with a particular focus on the Great Recession. To do so, I augment an otherwise standard real business cycle model with a non-trivial banking sector, in which financial intermediaries face an occasionally binding endogenous limit on their leverage ratio. The asymmetry induced by the occasionally binding constraint generates non-monotone dynamics, therefore capturing the idea of infrequent financial crises nested within typical business cycles. At the same time, the framework is still tractable enough to allow for the introduction of several standard features typically used in the DSGE literature.

In the spirit of Chari, Kehoe, and McGrattan (2007), I show that the baseline model with GK friction is equivalent to a prototype economy with an intertemporal investment wedge that distorts the Euler equation for capital. Moreover, the investment wedge is shown to be a function of the key endogenous variables associated with the friction in the microfounded model: that is, the multiplier associated with the occasionally binding constraint, and the marginal value of net worth (a measure of the “precautionary capital” motive that arises in the nonlinear solution). Unlike that paper, however, I back out the structural economic shocks that drove the economy into the Great Recession using the bootstrap particle filter.

Consistent with previous literature, the results suggest that financial frictions that manifest as intertemporal wedges are relatively unimportant to understand U.S. business cycles over the five decades previous to the crisis. More surprisingly, while the GK friction (or investment wedge) was indeed quantitatively more relevant during the Great Recession, its effects are still of second-order importance relative to other driving forces like productivity or labor wedge shocks.
References


18. Gertler and Karadi (2012). “QE1 vs. 2 vs. 3... A Framework for Analyzing Large Scale Asset Purchases as a Monetary Policy Tool”. mimeo.


Appendix I. Aggregate Resource Constraint

In the aggregate, $Q_tK_t$ is the total value of capital acquired by final goods producers in period $t$, while $Q_tS_t$ is the total value of claims issued against that capital (total credit in the economy). By arbitrage, the capital market clearing condition implies:

$$K_t = S_t \quad (44)$$

To get the economy-wide resource constraint, we start with the budget constraint of the households:

$$C_t + D_t + T_t = W_t H_t + R_{t-1} D_{t-1} + \Sigma_t \quad (45)$$

where $\Sigma_t$ includes the profits from the ownership of productive firms (final good producers, $\Pi^C_t$, and capital goods producers $\Pi^K_t$) and the net transfers between households and banks (exiting and new). Therefore, we have:

$$\Sigma_t = \Pi^C_t + \Pi^K_t + (1 - \theta) \left[ R^K_t Q_{t-1} K_{t-1} - R_{t-1} D_{t-1} \right] - (1 - \theta) t Q_{t-1} K_{t-1} \quad (46)$$

Next, combining the balance sheet and the law of motion for net worth of the aggregated banking sector, and imposing the capital market clearing (44), we have:

$$Q_t K_t = D_t + N_t = D_t + \theta \left[ R^K_t Q_{t-1} K_{t-1} - R_{t-1} D_{t-1} \right] + (1 - \theta) t Q_{t-1} K_{t-1} \quad (47)$$

where $R^K_t = \left[ \frac{Z_t + (1 - \delta) Q_t}{Q_{t-1}} \right] \Psi_t$. Combining (45), (47), the budget constraint of the government (31), and (46) yields:

$$C_t + G_t + Q_t K_t = W_t H_t + R_{t-1} D_{t-1} + \Sigma_t + \theta \left[ R^K_t Q_{t-1} K_{t-1} - R_{t-1} D_{t-1} \right] + (1 - \theta) t Q_{t-1} K_{t-1}$$

$$C_t + G_t + Q_t K_t = W_t H_t + R_{t-1} D_{t-1} + \Pi^C_t + \Pi^K_t + (1 - \theta + \theta) \left[ R^K_t Q_{t-1} K_{t-1} - R_{t-1} D_{t-1} \right]$$

$$C_t + I_t + G_t + Q_t K_t = W_t H_t + R^K_t Q_{t-1} K_{t-1} + Y_t - W_t H_t - Z_t \Psi_t K_{t-1} + Q_t K_t + Q_t(1 - \delta) \Psi_t K_{t-1}$$

$$C_t + I_t + G_t = [Z_t + (1 - \delta) Q_t] \Psi_t K_{t-1} + Y_t - Z_t \Psi_t K_{t-1} - Q_t(1 - \delta) \Psi_t K_{t-1}$$

Finally:

$$C_t + I_t + G_t = Y_t.$$

$$\quad (48)$$
Appendix II. Competitive Equilibrium

The rational expectations equilibrium of the model is a set of sequences for the 17 endogenous variables

\[ \{ C_t, I_t, Y_t, H_t, K_t, N_t, P_t, R_t, R^K_t, Q_t, \phi_t, \psi_t, \xi_t, SPR_t, G_t, W_t, Z_t \} \]

such that for given initial conditions and exogenous sequences

\[ \{ A_t, \varphi_t, g_t, \Psi_t \} \]

the following conditions are satisfied:

- Households maximize utility subject to their budget constraint, that is, the following equations hold:

  \[
  1 = \mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} R_t \right] \\
  \varphi_t \chi H^\xi_t = C_t^{-\gamma} W_t
  \]

- Banks maximize their expected terminal wealth subject to the IC constraint, that is, the following equations hold:

  \[
  \begin{align*}
  \phi_t &= \frac{Q_t K_t}{N_t} \\
  \psi_t &= \mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} R_t \right] \left[ 1 - \theta + \theta \psi_{t+1} \right] \\
  \mu \xi_t &= \mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} \right] \left[ 1 - \theta + \theta \psi_{t+1} \right] \left[ R^K_{t+1} - R_t \right] \\
  0 &= \left[ \psi_t - \mu \phi_t \right] \xi_t \\
  N_t &= \theta R^K_t Q_{t-1} K_{t-1} + P_{t-1} \\
  P_t &= \theta \left[ R_t (N_t - Q_t K_t) \right] + (1 - \theta) \psi_t Q_t K_t
  \end{align*}
  \]

- Capital producers maximizes profits subject to their technology, that is, the following equations hold:

  \[
  \begin{align*}
  K_t &= (1 - \delta) \Psi_t K_{t-1} + \left[ a_1 \left( \frac{I_t}{\Psi_t K_{t-1}} \right)^{1-\gamma} + a_2 \right] \Psi_t K_{t-1} \\
  Q_t &= \left[ \frac{I_t}{\delta \Psi_t K_{t-1}} \right]^\theta
  \end{align*}
  \]
Final good producers maximizes profits subject to their technology, and markets clear, that is:

\[
R^K_t \equiv \left[ \frac{Z_t + (1-\delta)Q_t}{Q_{t-1}} \right] \Psi_t \quad (59)
\]

\[
Y_t = A_t (\Psi_t K_{t-1})^\alpha H_t^{1-\alpha} \quad (60)
\]

\[
W_t = (1-\alpha) \frac{Y_t}{H_t} \quad (61)
\]

\[
Z_t = \alpha \frac{Y_t}{\Psi_t K_{t-1}} \quad (62)
\]

\[
Y_t = C_t + I_t + G_t \quad (63)
\]

\[
G_t \equiv \left( 1 - \frac{1}{g_t} \right) Y_t \quad (64)
\]

\[
SPR_t \equiv E_t R^K_{t+1} - R_t \quad (65)
\]
Appendix III. Solution Algorithm

I solve the model using a global approximation method based on Chebyshev approximations of decision rules along the lines of Judd (1992). Following Aruoba, Cuba-Borda, and Schorfheide (2018), the solution algorithm involves two non-standard tools: (i) a piecewise smooth representation of the approximated decision rules, and (ii) an iterative procedure of choosing grid points based on a clustered-grid-algorithm (CGA) proposed by Judd, Maliar, and Maliar (2010).

The set of equilibrium conditions for \( \{ C_t, I_t, Y_t, H_t, K_t, N_t, P_t, R_t, Q_t, \phi_t, \psi_t, \xi_t \} \) can be written as follows:

\[
1 = \mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} R_t \right] (66)
\]

\[
\psi_t = \frac{\mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} R_t \right] \left[ 1 - \theta + \theta \psi_{t+1} \right]}{1 - \xi_t} (67)
\]

\[
\mu \xi_t = \mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} \right] \left[ 1 - \theta + \theta \psi_{t+1} \right] \left( \frac{\alpha \eta_{t+1} + (1 - \delta)Q_{t+1}}{Q_t} \right) \Psi_{t+1} - R_t \right] (68)
\]

\[
0 = \left[ \psi_t - \mu \phi_t \right] \xi_t (69)
\]

\[
H_t = \left[ \frac{(1 - \alpha)A_t(\Psi_t K_{t-1})^\alpha}{\chi \phi_t^2 C_t^\gamma} \right]^{\frac{1}{1 - \gamma}} \xi_t (70)
\]

\[
\phi_t = \frac{Q_t K_t}{N_t} (71)
\]

\[
N_t = \theta R_t^K Q_{t-1} K_{t-1} + P_{t-1} (72)
\]

\[
P_t = \theta [R_t(N_t - Q_t K_t)] + (1 - \theta)\eta Q_t K_t (73)
\]

\[
K_t = (1 - \delta)\Psi_t K_{t-1} + \left[ a_1 \left( \frac{I_t}{\Psi_t K_{t-1}} \right) ^{1 - \rho} + a_2 \right] \Psi_t K_{t-1} (74)
\]

\[
Q_t = \left[ \frac{I_t}{\delta \Psi_t K_{t-1}} \right] ^{\rho} (75)
\]

\[
Y_t = A_t (\Psi_t K_{t-1})^\alpha H_t^{1 - \alpha} (76)
\]

\[
Y_t = C_t + I_t + \left( 1 - \frac{1}{g_t} \right) Y_t (77)
\]

The model has two endogenous and four exogenous state variables: \( S_t = \{ K_{t-1}, P_{t-1}; A_t, \eta_t, g_t, \Psi_t \} \). I approximate the decision rules for consumption \( C_t \), the risk-free interest rate \( R_t \), and the banking sector Tobin’s \( q \) ratio \( \psi_t \), in a piecewise fashion as follows. Define the set of approximated control variables in period \( t \) as \( \mathcal{X}_t = \{ C_t, R_t, \psi_t \} \). The piecewise smooth functions are parameterized by the set \( \Theta = \{ \Theta_x^x, \Theta_x^b \} \) as follows:
\[ \mathcal{X}_t = (1 - I_b) \cdot \Theta^X_t \mathbb{T}(S_t) + I_b \cdot \Theta^X_t \mathbb{T}(S_t) \]

where \( I_b \) is an indicator function that takes the value of one when the economy is under the binding regime \( (\xi_t > 0) \) and zero otherwise, and \( \mathbb{T}(\cdot) \) is a vector collecting complete combinations of Chebyshev polynomials. The controls \( \mathcal{X}_t = \{C_t, R_t, \psi_t\} \) solve the set of residual functions given by:

\[
\mathcal{R}_{1t} = \mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} R_t \right] - 1
\]

\[
\mathcal{R}_{2t} = \mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} \right] \left[ 1 - \theta + \theta \psi_{t+1} \right] \left[ \left( \frac{\alpha \Psi_{t+1} + (1 - \delta) Q_{t+1}}{Q_t} \right) \Psi_{t+1} - R_t \right] - \mu \xi_t
\]

\[
\mathcal{R}_{3t} = \begin{cases} 
\mu \phi_t - \psi_t & \text{if constraint is binding, } \xi_t > 0 \\
\mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} R_t \right] \left[ 1 - \theta + \theta \psi_{t+1} \right] - \psi_t & \text{if constraint is slack, } \xi_t = 0
\end{cases}
\]

I assess the accuracy of the numerical solution by computing Euler equations errors. Figure 7 show a histogram of the three Euler equation residuals computed over a CGA grid based on a long simulation of the model. I express the errors in decimal log scale as is common in the literature. The Euler errors are small, averaging -4.9, -4.8, and -4.5 for \( \mathcal{R}_1, \mathcal{R}_2, \) and \( \mathcal{R}_3, \) respectively.

Figure 7: Histogram of Euler Equation Errors (log10(abs(EE)) scale)

Notes: The histograms report the Euler equation errors over a simulation of 10,000 periods in decimal log basis. The dotted vertical line corresponds to the mean of the residuals over the simulation.
Appendix IV. Historical Decompositions 1965-2015

Figure 8: Baseline Model: Historical Decomposition: Real Variables and Spread

Notes: Output, investment, and hours worked are in quarterly terms, expressed in percent deviations from their ergodic means (solid line). Spread is expressed in annualized percent terms. The solid lines correspond to the model-implied filtered observable variables, which are up to a small measurement error equal to the data described above. The bars decompose each filtered variable into the contributions of each structural shock. The gray areas indicate NBER recession dates.
Notes: Output, investment, and hours worked are in quarterly terms, expressed in percent deviations from their ergodic means (solid line). Spread is expressed in annualized percent terms. The solid lines correspond to the model-implied filtered observable variables, which are up to a small measurement error equal to the data described above. The bars decompose each filtered variable into the contributions of each structural shock. The gray areas indicate NBER recession dates.