DISCUSSION PAPER SERIES

DP15428
(v. 2)

BEWLEY BANKS

Tommaso Monacelli and Rustam Jamilov

INTERNATIONAL MACROECONOMICS AND FINANCE

MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

BEWLEY BANKS

Tommaso Monacelli and Rustam Jamilov

Discussion Paper DP15428
First Published 05 November 2020
This Revision 16 November 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

BEWLEY BANKS

Abstract

We develop a non-linear, quantitative macroeconomic model with heterogeneous monopolistic financial intermediaries, incomplete markets, default risk, endogenous bank entry, and aggregate uncertainty. The model generates a bank net worth distribution fluctuation problem analogous to the canonical Bewley-Huggett-Aiyagari-Imrohoglu environment. Our framework nests Gertler-Kyiotaki (2010) and the standard Real Business Cycle model as special cases. We present four general results. First, relative to the GK benchmark, banks' balance sheet-driven recessions can be significantly amplified, depending on the interaction of endogenous credit margins, the cyclicality of a precautionary lending motive and the (counter-) cyclicality of intermediaries' idiosyncratic risk. Second, equilibrium responses to aggregate exogenous shocks depend explicitly on the conditional distributions of bank net worth and leverage, which are endogenous time-varying objects. Aggregate shocks to banks' balance sheets that hit a concentrated and fragile banking distribution cause significantly larger recessions. A persistent consolidation in the U.S. banking sector that matches the one observed over 1980-2020 generates a large economic contraction and an increase in financial instability. Third, we document, and match, novel stylized facts on both the cross-section of credit margins and the cyclical properties of the first three moments of the cross-sectional distributions of financial intermediary assets, net worth, leverage, loan margins, and default risk. We find that shocks to capital quality and to leverage constraint tightness (``financial shocks'') can match fluctuations in the U.S. financial sector very well. Finally, we use the model to identify and characterize episodes of systemic banking crises. Such events are associated with large economic recessions, spikes in bank leverage, and large drops in the number of intermediaries.

JEL Classification: E44, E32

Keywords: financial intermediaries, Heterogeneity, incomplete markets, monopolistic competition, Aggregate fluctuations

Tommaso Monacelli - tommaso.monacelli@unibocconi.it
Bocconi University and CEPR

Rustam Jamilov - jamilovrustam@gmail.com
London Business School

Bewley Banks*

Rustam Jamilov
London Business School

Tommaso Monacelli
Bocconi University, IGIER, and CEPR

October 15, 2020

Abstract

We develop a non-linear, quantitative macroeconomic model with heterogeneous monopolistic financial intermediaries, incomplete markets, default risk, endogenous bank entry, and aggregate uncertainty. The model generates a bank net worth distribution fluctuation problem analogous to the canonical Bewley-Huggett-Aiyagari-Imrohoglu environment. Our framework nests Gertler and Kiyotaki (2010) (GK) and the standard Real Business Cycle model as special cases. We present four general results. First, relative to the GK benchmark, banks' balance sheet-driven recessions can be significantly amplified, depending on the interaction of endogenous credit margins, the cyclicality of a precautionary lending motive and the (counter-) cyclicality of intermediaries' idiosyncratic risk. Second, equilibrium responses to aggregate exogenous shocks depend explicitly on the conditional distributions of bank net worth and leverage, which are endogenous time-varying objects. Aggregate shocks to banks' balance sheets that hit a concentrated and fragile banking distribution cause significantly larger recessions. A persistent consolidation in the U.S. banking sector that matches the one observed over 1980-2020 generates a large economic contraction and an increase in financial instability. Third, we document, and match, novel stylized facts on both the cross-section of credit margins and the cyclical properties of the first three moments of the cross-sectional distributions of financial intermediary assets, net worth, leverage, loan margins, and default risk. We find that shocks to capital quality and to leverage constraint tightness ("financial shocks") can match fluctuations in the U.S. financial sector very well. Finally, we use the model to identify and characterize episodes of systemic banking crises. Such events are associated with large economic recessions, spikes in bank leverage, and large drops in the number of intermediaries.

*We thank Marco Bellifemine for excellent research assistance. Jamilov thanks the AQR Asset Management Institute and the Wheeler Institute for Business and Development for financial support.

1 Introduction

The 2007-2008 Global Financial Crisis has transformed the way the profession thinks about the role of financial intermediaries in the economy. A large, new literature that followed has recognized the importance of financial frictions in a rich variety of setups that span constraints on risk-taking, imperfect competition, credit cycles, and moral hazard. However, the vast majority of existing papers are studies of the *first moment* rather than of the full *distribution* of financial intermediaries. In this paper, we lay out a tractable, quantitative macroeconomic framework with a banking sector where the time-varying distributions of financial intermediary net worth and leverage are at the core of the analysis. We supplement our quantitative analysis with novel stylized facts on both the cross-section and the cyclicality of different moments of the U.S. banking distribution.

Our model builds on the work of Jamilov (2020) and introduces aggregate uncertainty and heterogeneous banks into an environment with monopolistic competition in financial intermediation, incomplete markets, default risk, and endogenous bank entry. Bank credit is intermediated through a time-varying mass of local credit markets. Each market features a unique financial variety or amenity that is desired by the household. A single bank intermediates all markets and can charge market-specific rate margins. The elasticity of substitution across credit markets is constant trough time and states of nature, in a way similar to the goods market structure in Dixit and Stiglitz (1977) or Melitz (2003). Imperfect competition is the source of an aggregate credit supply externality, in the spirit of Blanchard and Kiyotaki (1987): the bank does not internalize the impact of market-specific loan margins on aggregate investment demand.

In addition to monopolistic financial intermediation, we assume that the bank faces local, partially uninsurable idiosyncratic rate of return risk in the spirit of Benhabib et al. (2018). Idiosyncratic risk, jointly with credit market power, creates a banks' net worth distribution fluctuation problem analogous to the canonical Bewley-Huggett-Aiyagari-Imrohoglu environment (Bewley, 1977; Huggett, 1990; Aiyagari, 1994; Imrohoglu, 1996). Importantly, our modelling approach eliminates *scale invariance*: all dynamic choices in the financial sector depend on bank-specific characteristics such as the level of net worth.[1] The number of local credit markets is determined in equilibrium through endogenous entry, similarly to the heterogenous non-financial firms model of Melitz (2003). Equilibrium yields a non-trivial, dynamic distribution of bank assets. The presence of aggregate risk makes this distribution, in principle an infinitely-dimensional object, a relevant "state variable". Aggregate state-dependency on the distribution is thus achieved explicitly, a result that is not feasible in other environments that feature scale invariance and complete markets. Under

[1]Eliminating scale invariance is a crucial step that separates our paper from the rest of the literature where a "representative" intermediary is the commonplace assumption. An important exception is Coimbra and Rey (2019) who study ex-ante heterogeneous banks. In contrast, our model delivers ex-post heterogeneity in returns and bank size due to market incompleteness and loan market power.

perfectly competitive credit markets, and in the absence of idiosyncratic risk, our "Bewley Banks" setup nests the canonical Real Business Cycle model and the Gertler and Kiyotaki (2010); Gertler and Karadi (2011); Gertler et al. (2016) macro-banking frameworks as special cases ("GK" models henceforth).[2] Such high tractability allows us to conduct a variety of benchmarking exercises.

A key advantage of our Bewley Banks framework is that we can target cyclical properties of *higher-order* moments of any banking characteristic. To that end, we document a comprehensive set of stylized facts on the distribution of U.S. financial intermediaries, both in the cross-section and over the business cycle . In particular, we focus on the mean, dispersion, and concentration of intermediaries' assets, net worth, leverage ratio, loan margins, and default risk. Nailing down empirical moments of the bank distribution turns out to be a non-trivial task. There is multi-modality in the aggregate data and rich heterogeneity across *sub-industries* of the broader financial sector. We therefore provide additional sets of industry-specific facts on depository institutions, brokers and dealers, insurance companies, etc. We also report new stylized facts for other developed economies including Australia, Canada, France, Germany, and the United Kingdom.

As in the data, our model generates a right-skewed ergodic distribution of banks' asset and leverage, and an inverse relationship between credit margins and bank size. The model also generates business cycle statistics that approximate the cyclical properties of the different moments of the U.S. banking distribution rather well. Relative to the GK benchmark, and in response to banks' balance sheet ("capital quality") shocks, our baseline model generates equilibrium dynamics for key aggregate variables that are considerably dampened. This result is due to two reinforcing channels. First, a precautionary lending motive, due to market incompleteness, makes each intermediary accumulate more equity capital than in the GK counterfactual. Greater assets and net worth, coupled with lower aggregate leverage in the precautionary stochastic steady state, leave the financial sector in a less fragile initial condition when aggregate shocks hit. Second, credit market power is an additional margin of adjustment in response to adverse shocks that allows the bank to boost profits in high marginal-utility states by raising prices and reducing quantities by less. As a result of both mechanisms, aggregate contractions get dampened.

In order to counteract the precautionary lending motive, we allow idiosyncratic risk to be state-dependent and *counter-cyclical*. Among many others, Bloom et al. (2018) document that non-financial microeconomic risk rises in recessions. In our model, because all non-financial firms rely on external bank-driven financing, non-financial sector risk translates immediately into lower expected portfolio returns of the banker. In low aggregate states, a larger mass of credit markets can now experience low idiosyncratic return draws. The direct impact of business cycle fluctuations on bank balance sheets dominates the ex-ante precautionary lending motive. Recessions get amplified considerably, both in terms of real macroeconomic aggregates and in the financial sector. Notably,

[2]To enable the cleanest possible comparison, we solve GK with fully non-linear methods as well.

the number of active intermediaries falls by an order of magnitude more than in the acyclical risk counterfactual. Overall, in the Bewley Banks framework crises could be either amplified or dampened depending on whether idiosyncratic bank risk is acyclical or not.

Macroeconomic and financial dynamics with Bewley Banks are driven by the interaction of *three main forces*: credit market power, idiosyncratic return risk, and endogenous entry. The tractability of the model allows us to isolate the differential contribution of each force sequentially. From this exercise we gain three main insights. First, shutting down credit market power considerably amplifies the aggregate sensitivity to exogenous shocks. Second, without idiosyncratic return risk the economy is less risky and less responsive to aggregate shocks. Third, eliminating endogenous entry dampens the responsiveness by a negligible amount. In the baseline model with acyclical idiosyncratic risk, the large dampening role of credit market power dominates the amplifying effect of market incompleteness. The extensive margin is muted and does little to affect the responses of either output or consumption.

An important result of the paper pertains to the general aggregate *state dependency* on the dynamic distributions of bank net worth and leverage. In the model, equilibrium responses of aggregate output and consumption to exogenous aggregate shocks depend explicitly on the conditions in the banking cross section. A negative bank capital quality shock that hits the economy in a state with a more concentrated and fragile distribution of net worth and leverage generates a significantly larger cumulative loss in output, consumption, bank assets, and considerably greater levels of bank leverage and loan margins.

Recent papers by Jamilov (2020) and Corbae and D'Erasmo (2020) document a multifold increase in both the degree of dispersion and concentration in the U.S. banking sector over the 1980-2020 period, coupled with a steady decline in the absolute number of depository institutions. We explore the role of second moment shocks as a source of business cycle fluctuations. We find that positive transitory shocks to the dispersion of bank assets, the magnitude of which matches the data, have large negative effects both on the real economy and the financial sector. Specifically, aggregate output, consumption, bank assets and net worth each fall by substantial amounts whereas bank leverage, credit margins, and default risk all rise considerably. We implement this particular exercise by allowing the agents in the model to explicitly track and forecast higher-order moments of the distribution of bank net worth. This computational approach follows the original idea in the seminal works by Krusell and Smith (1996, 1998). That is, we keep track of the first two moments of the banking distribution as an approximation for what is otherwise an infinitely-dimensional object. In general, we find that persistent shocks to higher-order moments of the banking distribution have a significant and lasting effect. The key for this result is that our model generates concentrated, right-skewed distributions of bank assets and leverage.

Another avenue that we explore in the paper is the identification and characterization of systemic *banking crises* using event study methods. We apply tools from the open-economy macroeconomics literature that looks at financial crises in emerging economies (Mendoza, 2010). We simulate a long time-series from our model and define an economic crisis as an episode with unusually low measured TFP. We then collect all instances of such events and compute period-specific averages of key macro-financial aggregates. We find that economic crises occur in conjunction with banking crises. In relative terms, banking crises in the Bewley economy lead to more dampened aggregate contractions than in the GK counterfactual with perfect competition and no idiosyncratic risk. Crises in the Bewley economy are also less financially unstable - i.e., bank leverage increases by less. This is a variant of the financial competition-stability trade-off (Hellman et al., 2000). Imperfect credit-market competition acts as a buffer against shocks to financial stability but at the price of high loan margins and a greater steady-state loss in consumer welfare.

Crises in the Bewley Banks economy with *counter-cyclical idiosyncratic risk*, however, are significantly amplified. These episodes correspond to greater contractions in output, consumption, bank assets and net worth. Furthermore, the number of active intermediaries falls by an order of magnitude more than in the baseline model with acyclical idiosyncratic risk. In terms of financial stability, the financial sector is more fragile than in the baseline: aggregate leverage and default risk both increase by a greater amount. Interestingly, crises in the Bewley Banks economy with counter-cyclical risk are characterized by a persistent *decline* of credit margins in the build-up periods. Meanwhile, crises in the baseline case with acyclical idiosyncratic risk are associated with a *rise* of credit margins in the the same build-up phase. This is an interesting testable implication of our model. Generally speaking, this result suggests that competition and market power in the financial sector could be leveraged for the ex-post measurement of financial conditions as well as forward-looking diagnostics and forecasts of distress and crises.

Up to this point, the sole source of aggregate uncertainty in all quantitative experiments has been a (so-called) capital quality shock. In the final exercise, we take business cycle moments in the data as given and run a horse race across six different types of aggregate shocks with the purpose of matching as many unconditional correlations as possible. We consider aggregate shocks to Hicks-neutral total factor productivity, quality of aggregate capital, intermediary dividend payouts, credit markups, leverage constraint tightness, and degree of market incompleteness. For each shock type, we solve our model under the assumption that it is the only source of aggregate uncertainty in the environment. We find that shocks to capital quality and to leverage constraint tightness ("financial shocks") can match fluctuations in the U.S. financial sector very well.

Solving our model numerically is potentially a challenging task for at least four reasons. First, without scale invariance the distribution of financial intermediary assets, an infinitely-dimensional object, is now a "state variable" that needs to be kept track of. We solve this issue with a variant of the Krusell and Smith (1998) algorithm, where we assume that the bank forms linear forecasts for a limited set of moments of the cross-sectional distribution. Along these lines, the bank must also rationally forecast future aggregate prices (loan rates) which are set at the level of a local credit market. The two forecasts taken together allow the bank to pin down the projected return on aggregate capital. Second, the market for deposit holdings must clear on each point of the aggregate state space. Third, dynamics of the cross-sectional bank distribution must be consistent with the dynamic problem of the incumbents *and* new entrants. Finally, the intermediary faces an occasionally binding constraint on leverage. Importantly, this constraint may bind on any part of the idiosyncratic or aggregate state space. In Section 3.13 we describe our computational algorithm in detail.

Related Literature. This paper builds on the recent work by Jamilov (2020) and Corbae and D'Erasmo (2019, 2020) who explore aggregate implications of the rise of U.S banking concentration. De Loecker et al. (2020) and Diez et al. (2018) document a large, 50%+ increase in markups (margins) in the U.S. financial industry over 1980-2016. The apparent consolidation of market power in the financial sector calls into question the assumption of perfect competition that is made in most existing quantitative macro-banking models. Gerali et al. (2010) and Cuciniello and Signoretti (2015), among others, studied the role of imperfect competition in general equilibrium macroeconomic models. Our paper is different in our assumption of persistent ex-post heterogeneity, which allows us to achieve a smooth equilibrium ergodic distribution of bank leverage. Our financial intermediaries are also exposed to insolvency-driven default risk, which is priced into a distribution of deposit rates.

Among studies that explore banking industry dynamics in general equilibrium, Corbae and D'Erasmo (2019) is the paper that is closest to ours but uses a different approach.[3] Authors focus on dynamic capital requirements in a quantitative model of oligopolistic financial competition with dynamic interactions between one large bank with market power and many small perfectly-competitive institutions. We differ from Corbae and D'Erasmo (2019) in at least three main respects. First, our approach to modelling imperfect banking competition follows a large literature that works with CES aggregation (Dixit and Stiglitz, 1977; Melitz, 2003). This approach is highly tractable and "portable" - our monopolistically competitive banking bloc can be readily enhanced to allow for nominal rigidities or open-economy features[4]. Second, in our model local credit market-specific actions are not internalized in the aggregate. In the oligopostic banking setup of Corbae and D'Erasmo (2019), actions of the "lead bank" and "fringe banks" are fully internalized. Our

[3]Banking industry dynamics have been explored in several other recent papers including Capelle (2019), Rios Rull et al. (2018), Nguyen (2015), Christiano and Ikeda (2013), Davydiuk (2020), Martinez-Miera and Repullo (2010). This broad literature builds on the first generation of mostly partial-equilibrium models on the financial competition-stability trade-off (Allen and Gale, 1998; Hellman et al., 2000; Allen and Gale, 2004).

[4]We extend our framework to include nominal rigidities and a role for monetary policy in Jamilov and Monacelli (2020).

modelling approach yields a powerful aggregate credit supply externality, which acts as an additional channel of amplification on firm investment demand. Jamilov (2020) shows that internalization of the credit supply externality moves steady-state aggregates such as welfare, output, and bank credit by double-digit percentage points. Finally, our model nests explicitly the canonical Real Business Cycle model and the GK models as special cases, which adds to its tractability. Reverting to GK entails simply a re-calibration of three structural parameters.

We also build on the important work by Coimbra and Rey (2019) who study the impact of ex-ante financial heterogeneity on systemic macroeconomic risk and financial stability. Their model features both intensive and extensive margins, similarly to ours. Our approach differs from theirs in several substantive ways. First, we achieve persistent "ex-post" financial heterogeneity due to incomplete markets and exposure to idiosyncratic risk, while Coimbra and Rey (2019) assume ex-ante heterogeneity in Value-at-Risk constraints. Second, in our setup, trading markets are incomplete and the banking distribution is a time-varying, endogenous state variable which must be kept track of. Our model also, importantly, allows for credit market power and a mismatch between the cost of funds and the price of bank credit. Finally, the number of active intermediaries in our model is a time-varying object that moves with the business cycle.[5]

Our assumption of idiosyncratic rate of return risk in banking is not ad-hoc or far-stretched. Several recent studies document that intermediaries are not perfectly insured against non-systematic shocks at various layers of aggregation. Galaasen et al. (2020) show that financial intermediaries in Norway are exposed to idiosyncratic borrower-level risk at multiple layers of aggregation. In particular, they show that bank portfolios are highly regionally concentrated due to regional "home bias" in lending and that local market-specific risk is hard to hedge. Their finding maps closely to our modelling approach where idiosyncratic risk is tied to the spatial distribution of locally differentiated credit markets. Paravisini et al. (2020) find that persistent specialization of banks by export market leaves them vulnerable to idiosyncratic shocks originating with foreign partner-countries. Agarwal et al. (2020) find that banks which over-exposed to the Mexican energy sector were much more likely to suffer from the industry-specific negative shock of 2014. Although rigorously micro-founding idiosyncratic risk is beyond the scope of our quantitative framework, our parsimonious modelling approach can capture and operationalise the general idea reasonably well.

The rest of the paper is structured as follows. Section 2 reports stylized facts on the banking distribution in the cross-section and over the business cycle. In Section 3, we lay out our model. Section 4 describes our calibration strategy, shows the model policy functions and ergodic distributions, and demonstrates the responsiveness to aggregate fluctuations. Section 5 inspects the model

[5]Other papers such as Korinek and Nowak (2016), Boissay et al. (2016), Goldstein et al. (2020), Begenau and Landvoigt (2018) also work with equilibrium models of ex-ante heterogeneity in the financial sector.

mechanism by isolating each key moving part. Section 6 presents our main quantitative results and experiments. Section 7 explores different types of exogenous aggregate shocks. Finally, Section 8 concludes.

2 Stylized Facts on the Distribution of Financial Intermediaries

In this section we document key stylized facts on the banking industry in the cross-section and over the business cycle. In Section 2.2 we report facts on the cross-sectional distribution. In Section 2.3 we look at the behavior of the banking distribution over the business cycle. We focus on the whole financial sector in main text and discuss sub-industry heterogeneity in Appendix A.2. We also report, subject to data availability, relevant statistics for non-U.S. countries in Appendix A.3. Our data Appendix A.1 provides further details on raw data and our data cleaning approaches. Alternative robustness checks are documented in Appendix A.4.

2.1 Data Description

Variable Definition

For every financial characteristic that we describe below, we are interested in computing the first three moments of the time-varying cross section. We begin with the first moment which we proxy with the (unweighted) mean μ_t of a generic panel x_{jt} with N_t being the size of the population:

$$\mu_t(x) = \sum_{j}^{N_t} \frac{1}{N_t} x_{jt}$$

For the second moment, we compute the time-varying standard deviation σ_t:

$$\sigma_t(x) = \sqrt{\frac{\Sigma_j^{N_t}\left(x_{jt} - \mu_t(x)\right)^2}{N_t}}$$

For the third moment, depending on the characteristic, we compute either the Herfindahl index or the statistical skewness. We use the HHI primarily for bank assets and net worth, and skewness for the leverage ratio. The Herfindal H_t of x_{jt} is defined using the usual formula, where s_{jt} is the share of bank j in market x in time t:

$$H_t(x) = \sum_{j}^{N_t} s_{jt}(x)^2$$

Finally, for statistical skewness S_t we use the Pearson's standardized third moment:

$$S_t(x) = \frac{\frac{1}{N_t}\sum_j^{N_t}\left(x_{jt} - \mu_t(x)\right)^3}{\left[\frac{1}{N_t-1}\sum_j^{N_t}\left(x_{jt} - \mu_t(x)\right)^2\right]^{\frac{3}{2}}}$$

Bank balance sheets

Our main source of bank balance sheet information is the Compustat database. We start by extracting financial intermediary assets and net worth for the financial sector in the U.S. Our definition of (book) leverage is the ratio of bank assets to net worth. The aggregate distribution of intermediary leverage is multi-modal, because different sub-industries of the broader sector have heterogeneous business models. We therefore also report statistics for six sub-sectors: depository credit institutions, non-depository credit institutions, brokers and dealers, insurance companies, real estate companies and brokers, and holding companies and investors. We define these sectors based on the SIC classification. We also work with Compustat Global, which has information for non-U.S. institutions. Existing data for non-U.S. developed economies is, however, limited. Our main aggregate time-series for assets, net worth, and leverage runs from 1985q1 until 2020q1. We explore samples with alternative starting dates in Appendix A.4.

Loan margins

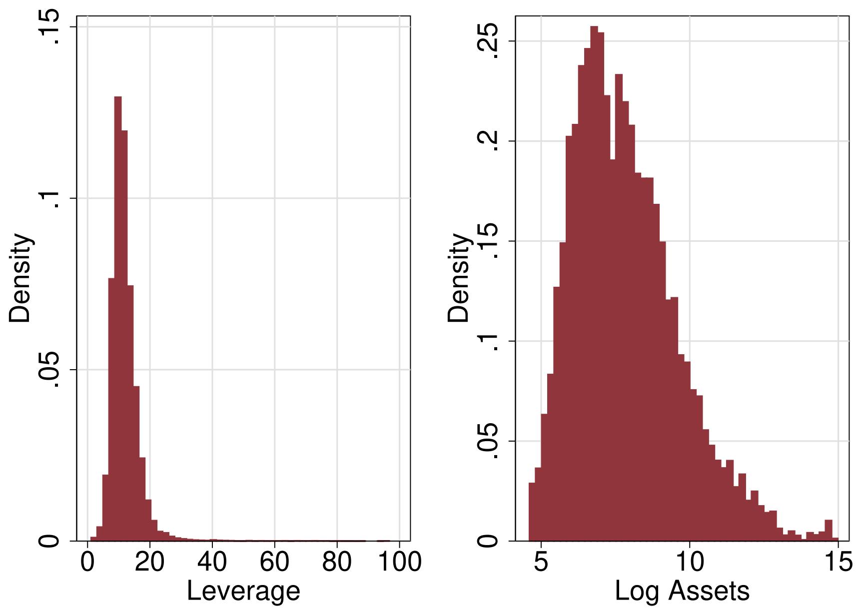
We construct our own, bottom-up measure of credit margins. To that end, we employ the Compustat Banks dataset. Our baseline definition of a credit margin is the bank-level ratio of Total Interest and Related Income over Total Interest and Related Expenses. Our approach is similar to Corbae and D'Erasmo (2019), who measure credit margins by the ratio of returns over the marginal cost of funds. Our main loan margin series runs from 1993q1 until 2020q1, and refers to depository institutions. For each quarter, as with the balance sheet data, we compute the mean, standard deviation, and skewness based on the raw bank-level panel data.

Bank default risk

We proxy bank-level default risk with data on Credit Default Swaps (CDS) provided by Markit. Our baseline measure is the 5-year CDS spread, which is the most liquid. Raw data is available at daily frequency, and we aggregate it to the bank-quarter level. We were only able to compute CDS spreads for the financial sector as a whole; industry-level information is not available. Our measures are available for both U.S. and non-U.S. based institutions. The final dataset runs over 2002q1-2020q1. For each quarter, we calculate the mean, standard deviation, and skewness of the distribution of CDS spreads.

Each variable, unless explicitly noted otherwise, is logged and linearly detrended. Our main proxy of the "business cycle" is real U.S. Gross Domestic Product (GDP), which is detrended and seasonally adjusted. The same filtering is performed on the model-simulated time-series as well,

Figure 1: **Banking Distributions in the Data**



Notes: Histograms of bank leverage and (log) total assets in the US data. Leverage is defined as book assets over book equity. Source: Compustat.

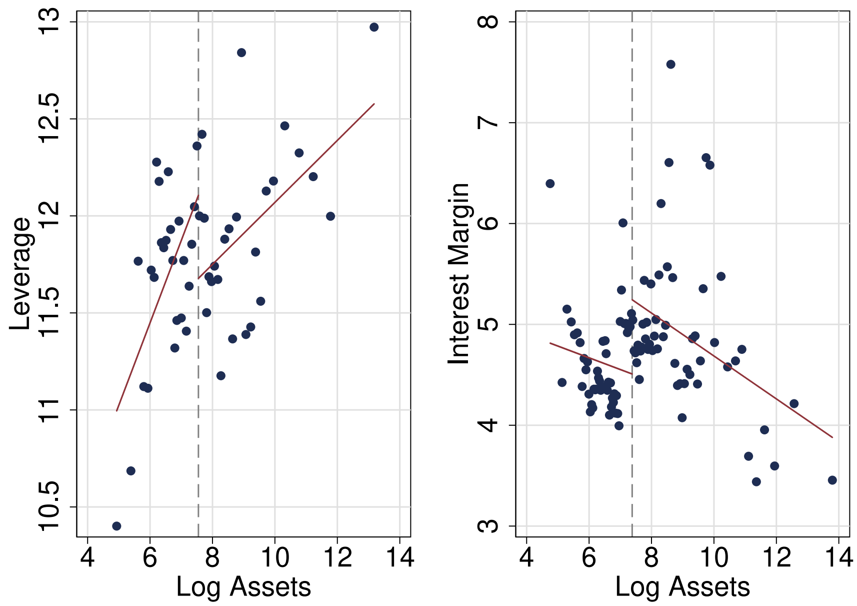
so the model and the data are directly comparable.

2.2 Banking in the Cross Section

We begin by first documenting facts on the banking distribution by analyzing the cross-section. In Figure 1 we plot the unconditional histograms of bank leverage (defined as the ratio of book assets over book equity) and the log of assets. We observe that both distributions are highly right-skewed. The vast majority of intermediaries have a leverage ratio somewhere in the 5-15 region. However, there is a small mass of highly risky banks. An important feature of our structural model will be the ability to generate such a skewed distribution thanks to the combination of a persistent idiosyncratic return process and market incompleteness.

In Figure 2 we plot binned scatterplots of bank size against book leverage and our measure of the interest margin. We define bank size as total book assets, which is also in line with a similar empirical exercise done in Coimbra and Rey (2019). We construct these plots in three steps. First, we residualize both y-axis and x-axis variables from the time fixed effect in order to absorb any common time-varying factors. Second, we construct 100 equally-sized bins of log assets. Each bin has at least 300 observations. For each bin we compute unweighted averages of log assets, book leverage, and the interest margin. Finally, we fit the data points separately for "small" and

Figure 2: **Bank Size, Leverage and Interest Margins**



Notes: Binned scatterplots of (log) assets against bank leverage and interest margins. Leverage is defined as book assets over book equity. Interest margins are defined as the ratio of total interest income to total interest and related expenses. Vertical dashed lines represent discontinuity at the median of log assets. Red lines of fit are separate lines of linear fit for values above and below the discontinuity. See main text for further details. Source: Compustat.

"large" intermediares, which we separate based on the median of the distribution of assets. This discontinuity on the graph is captured by a dashed vertical line.

From Figure 2 we document two empirical regularities. First, the conditional correlation of book leverage and bank size is positive. Larger banks are also more levered, in line with the findings in Coimbra and Rey (2019). Second, the conditional correlation of bank size and interest rate margins is negative. It is statistically significant at the 5% level. Larger intermediaries have lower interest rate margins, on average. The relationship is particularly strong for banks in the top half of the size distribution.

2.3 Banking over the Business Cycle

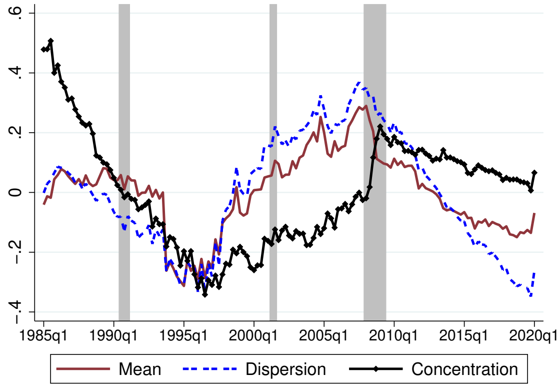
Next we present a series of facts documenting the cyclical properties of the first three moments of the banking distribution.

Figure 3 plots the time-series of the first three moments of financial intermediary assets, net worth (equity), leverage, margins, and default risk (CDS spreads). The underlying sample is for the U.S. only. In panel (a) we report the mean, standard deviation, and HHI of total intermediary *assets*. We see clearly from the picture that average assets are pro-cyclical, which is consistent with observations in Nuno and Thomas (2016). The second moment is also (highly) pro-cyclical. The third moment, however, is counter-cyclical. In panel (b) we plot the time-series for bank *equity*. Cyclical properties of the first three moments of bank equity are the same: positive, positive, and negative.

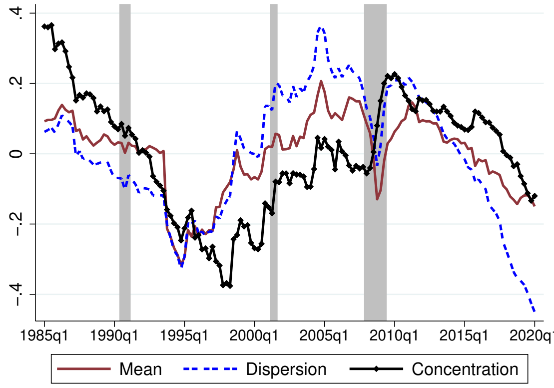
In panel (c) we report statistics on the bank *leverage ratio* (defined as book assets over book equity). The first moment is highly pro-cyclical, consistent with the evidence in Adrian and Shin (2010, 2011). It is important to note, however, that there is substantial variation in risk-taking management across different types of financial intermediaries, a point we will return to again in Appendix A.2. We find that the second moment of leverage is slightly pro-cyclical and the third moment (skewness) is strongly counter-cyclical. Overall, for the total financial sector the mean and dispersion of bank assets, net worth, and leverage are pro-cyclical while concentration is highly counter-cyclical. Counter-cyclicality of the third moment appears to be a robust feature of the data.

In panel (d) we plot the first three-moments of the distribution of *loan margins*. We find that average margins are strongly counter-cyclical, and so is the dispersion of margins. Concentration, however, is very pro-cyclical. Consistently with our results, Corbae and D'Erasmo (2019) also document that average credit margins are counter-cyclical but they do not explore higher-order moments like us. Finally, in panel (e) we document properties of *CDS spreads*. The first two moments of the distribution are counter-cyclical, a fact that is largely consistent with canonical theory. The third moment is slightly pro-cyclical but this result is sensitive to sample selection.

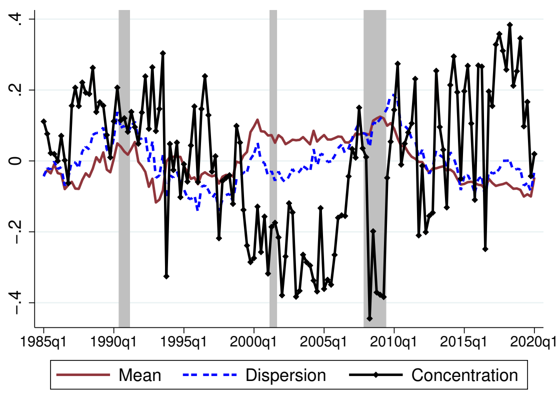
Figure 3: **Cyclical Components of the Distribution of U.S. Financial Intermediaries**



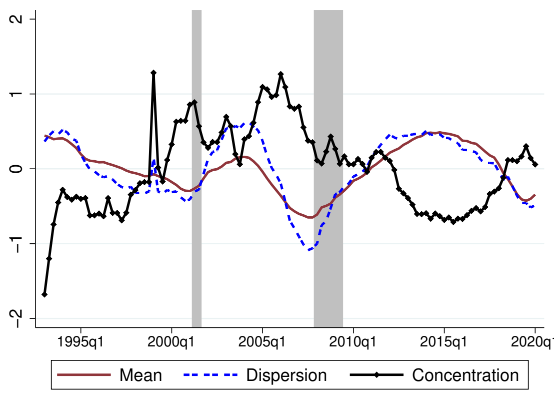
(a) Assets



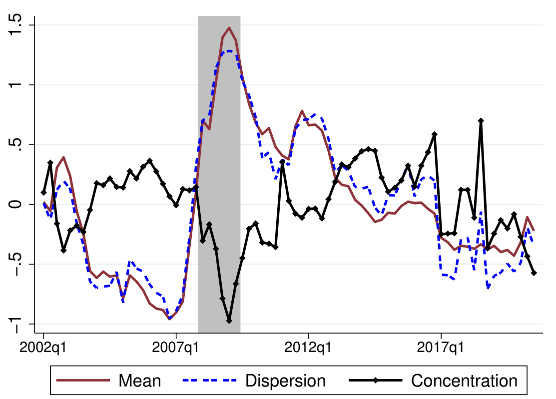
(b) Equity



(c) Leverage



(d) Credit Margins



(e) CDS Spreads

Notes: Every variable has been logged (except the skewness of leverage) and linearly detrended. Shaded areas represent US recessions based on the NBER classification. Bank balance sheet data is from Compustat. CDS data is from Markit. See Appendix A.1 for variable definitions and further details.

Table 1 displays a summary of our banking business cycle facts. Hence we see that intermediary

Table 1: **Business Cycle Statistics - Aggregate U.S. Data**

	Mean of	Dispersion of	Concentration of
Correlation			
Assets - GDP	0.498	0.642	-0.568
Net Worth - GDP	0.211	0.544	-0.472
Leverage - GDP	0.701	0.043	-0.641
Margins - GDP	-0.563	-0.370	0.725
CDS Spreads - GDP	-0.325	-0.309	0.033
Standard Deviation (%)			
Assets	13.383	19.371	18.281
Net Worth	11.268	18.076	16.640
Leverage	6.036	6.855	20.157
Margins	31.046	42.404	56.595
CDS Spreads	57.751	58.496	32.021

Notes: Dispersion is measured as time-varying standard deviation. Concentration is measured with skewness. For every variable except CDS spreads the sample is 1985q1:2020q1. For CDS spreads the sample is 2002q1:2020q1. Every variable has been logged (except the skewness of leverage) and linearly detrended. Bank balance sheet data is from Compustat. CDS data is from Markit. See Appendix A.1 for variable definitions and further details.

balance sheet quantities have a pro-cyclical mean and dispersion, and a strongly counter-cyclical HHI. Book leverage is pro-cyclical in the first two moments, and counter-cyclical in the third moment. The mean and standard deviation of credit margins and CDS spreads are both counter-cyclical, while skewness is pro-cyclical. In the lower panel of Table 1 we report standard deviations of time-series fluctuations of our main variables. The results can be summarized into two key facts. First, the volatility of balance sheet quantities - assets, net worth, leverage - is far smaller than the volatility of either credit margins or default risk. Second, higher-order moments appear to be more volatile than the mean. In particular, measures of concentration are especially volatile.

3 Model

In this section we lay out our quantitative model with a dynamic, endogenous distribution of bank characteristics.

3.1 Overview

Time is discrete and infinite. The economy is populated by *four agents*: a representative household, a capital goods producer, a final goods producer, and a financial intermediary. The economy is subject to *aggregate uncertainty* in the form of shocks to the quality of aggregate capital. This shock proxies exogenous perturbations to the balance sheet of financial intermediaries. The role of the bank is to intermediate funds between a unit mass of identical households and productive capital. The bank is owned by the household and ultimately redistributes back all its accumulated wealth (net worth) via dividends. Dividends are paid out only upon exit. Every period, the bank finances its operations via accumulated net worth or through a nationwide retail market, where it obtains deposits from households. There are no wholesale funding markets available.

Bank credit is intermediated on a time-varying mass J_t of local credit markets. This mass is determined endogenously through entry and exit. A single financial intermediary intermediates funds across all of these markets. Credit markets are differentiated by unique local features such as amenities or variety in financial services. The elasticity of substitution between local credit markets, $\theta > 1$, is constant across time and aggregate states of nature. Imperfect substitutability across credit markets allows the intermediary to charge localized marked-up *margins* over the cost of funds.

The portfolio return of the banker consists of a systematic component and a persistent, idiosyncratic, *uninsurable* returns process. Idiosyncratic risk is market-specific and cannot be hedged because trading markets are incomplete. A long enough sequence of negative idiosyncratic return shocks can drive the bank into insolvency. There is no deposit insurance. Bank default risk gets competitively priced into the interest rates on deposits by the homogeneous, risk-averse household.

Entry into credit markets is endogenous. There is a large number of potential entrants - financiers - who become financial varieties conditional on entry. Entering financiers pay a fixed startup cost and obtain a one-time idiosyncratic return draw. Having observed the draw, financiers can either decide to operate or to immediately exit. The mass of entering financiers grows until bank profits (in expectation) remain above the startup costs. Involuntary exit occurs at rate $0 < \sigma < 1$.

Importantly, our model breaks down scale invariance, thereby generating a dynamic, endogenous cross-section of intermediary assets. The distribution of bank assets (loan portfolio), an infinitely dimensional object, is a new state variable in the model. We describe how we deal with the curse of dimensionality in Section 3.13. The equilibrium is also associated with the dynamic endogenous distributions of bank net worth, leverage, default risk, margins, and deposit rates.

3.2 Aggregate Technology

There is a continuum of perfectly competitive firms that produce the final good using an identical constant returns to scale Cobb-Douglas production function with capital and labor as inputs. Labor is supplied inelastically, for tractability. Output Y_t is the following function of aggregate capital K_t

and labor L_t:

$$Y_t = AK_t^{\alpha}L_t^{1-\alpha} \tag{1}$$

with $0 < \alpha < 1$.

In the baseline model, the only source of aggregate uncertainty is a shock to the quality of capital ψ_t (Merton, 1973; Gertler and Kiyotaki, 2010; Gertler and Karadi, 2011). This shock captures fluctuations in the value of capital - its sudden obsolescence or valuation. We will explore alternative forms of aggregate uncertainty in the latter sections. We assume that ψ_t follows an AR(1) process.

$$\psi_{t+1} = (1-\rho_{\psi}) + \rho_{\psi}\psi_t + \epsilon_{t+1}^{\psi} \tag{2}$$

Capital accumulates over time according to the law of motion:

$$K_{t+1} = \psi_{t+1}\Big(I_t + (1-\delta)K_t\Big)$$

where I_t is aggregate investment of non-financial firms and $0 < \delta < 1$ is the constant depreciation rate. Wages are competitive and follow directly from the production function and firms' optimization. Return on aggregate capital R_t^k is:

$$R_{t+1}^k = \psi_{t+1}\left(\frac{A\alpha K_{t+1}^{\alpha-1} + (1-\delta)P_{t+1}}{P_t}\right) \tag{3}$$

which comprises profits and capital gains. The latter depend on the dynamics of the *aggregate* price of capital, P_t, which is determined in equilibrium by financial intermediary activities.

3.3 Local Credit Markets

There exists a time-varying mass J_t of local credit markets. The elasticity of substitution across markets is $\theta > 1$. Credit markets are differentiated by unique features that (local) borrowers derive utility from. Differentiated capital goods are assembled by a representative capital producing firm with a Dixit-Stiglitz aggregation technology from the mass J_t of available financial varieties k(j) where $j \in [0, J_t]$.

$$K_t = \left[\int_0^{J_t} k_t(j)^{\frac{\theta-1}{\theta}}dj\right]^{\frac{\theta}{\theta-1}} \tag{4}$$

Financial variety-specific demand functions are obtained from the following maximization problem:

$$\max_{k_t(j)}\Big[K_t - \int_0^{J_t} p_t(j)k_t(j)dj\Big] \tag{5}$$

subject to technology (4). This yields the demand function for banks' lending activities:

$$k_t(j) = \left(\frac{p_t(j)}{P_t}\right)^{-\theta} K_t \tag{6}$$

where $p_t(j)$ is the price of capital in the local credit market (j) and P_t is the aggregate price of capital consistent with the competitive capital producing firm earning zero profits:

$$P_t := \left[\int_0^{J_t} p(j)^{1-\theta} dj\right]^{\frac{1}{1-\theta}} \tag{7}$$

3.4 Financial Intermediary

The monopolistic credit demand system (4)-(7) is taken as given by the intermediary. The intermediary starts each period with an initial stock of net worth $n \in \mathbf{N} \subset \mathbf{R}_+$ and must choose the stock of assets k(j), deposits d(j), and price of varieties p(j) while satisfying the balance sheet constraint:

$$d_t(j) + n_t(j) = p_t(j)k_t(j)^{\beta} \tag{8}$$

where $\beta > 1$ is a parameter that governs local decreasing returns to scale. The bank can borrow deposits d(j) from the household, subject to the market-specific interest rate $\bar{R}_t(j)$. The incumbent banker earns a portfolio return $R^T(j)$ that consists of the return on aggregate capital R^k - common across local markets - and an idiosyncratic component $\xi(j)$ which is specific to each local credit market:

$$R_t^T(j) = \kappa\xi_t(j) + (1-\kappa)R_t^k \tag{9}$$

Where $0 < \kappa < 1$ is a parameter that governs the ability to hedge local market-specific return risk. For tractability, assume that $\xi \in \Xi$ follows an AR(1) process:

$$\xi_t(j) = (1-\rho_\xi)\mu_\xi + \rho_\xi\xi_{t-1}(j) + \sigma_\xi\epsilon_t(j) \tag{10}$$

Let the finite state Markov representation of (10) be $\mathbf{G}_{\xi_{t+1},\xi_t}$. The law of motion of bank net worth is therefore:

$$n_{t+1}(j) = R_{t+1}^T(j)p_t(j)k_t(j) - \bar{R}_t(j)d_t(j) \tag{11}$$

Following Gertler and Karadi (2011) and Gertler and Kiyotaki (2010), the bank-household relationship is subject to a moral hazard friction. The bank has an exogenous incentive to divert bank assets. It has the capacity to divert no more than a fraction λ of the total value of local assets p(j)k(j). If considering to divert, the banker always manages to escape but the bank enters bankruptcy the following period. In equilibrium, it must be that the bank stays indifferent between operating honestly and diverting. This yields the following incentive constraint that puts a limit on bank leverage:

$$\lambda\ p_t(j)k_t(j) \leq\ V_t(j) \tag{12}$$

where $V_t(j)$ is the franchise value of the intermediary, whose recursion is defined below.

The market-specific probability of default is $\upsilon(j)$. Default risk is due to fundamental insolvency, i.e., when bank net worth at normal market prices is non-positive:

$$\upsilon_t(j) = \Pr\Big(n_{t+1}(j)(n_t(j), \xi_t(j)) \leq 0\Big) \tag{13}$$

Conditional on insolvency, the household recovers only an endogenous fraction of promised payments $x_t(j)$, to be defined later. This risk is priced by the household through equilibrium deposit rates. Remaining assets get transferred to the capital producing firm who produces K_t as normal.

The distribution of financial varieties ("banks", for short) is summarized by the probability measure μ defined on the Borel algebra B that is generated by open subsets of the product space $\mathbf{B} = \mathbf{N} \times \Xi$, corresponding to the distribution of incumbent banks with net worth n and idiosyncratic return realization ξ. The aggregate state of the economy is (ψ, μ, M) with M_t the total mass of new entrants in period t. The law of motion for the distribution is:

$$\mu_{t+1} = \Gamma(\psi_t, \mu_t, M_{t+1}) \tag{14}$$

We define Γ below. The evolution of the distribution thus depends on bank entry and exit as well as on the decisions of the incumbent banks-varieties.

3.5 Credit Margins and Markups

Private bank-level *margins* $\chi(j)$ are defined as the ratio of capital relative prices p(j) to the cost of funds $\bar{R}(j)$:

$$\chi(j) = \frac{p(j)}{\bar{R}(j)} \tag{15}$$

This is also precisely the definition of margins from our empirical analysis. Given the aggregate state vector $\mathbf{S}$ that we define below, the aggregate margin $X(\mathbf{S})$ is defined as the ratio of the aggregate price of capital to the average interest rate on deposits:

$$X(\mathbf{S}) = \frac{P(\mathbf{S})}{\bar{R}(\mathbf{S})} \tag{16}$$

In our model, $X(\mathbf{S})$ is a dynamic, endogenous object. That is, each $\chi(j)$ is determined conditional on the forward-looking expectations about the evolution of idiosyncratic return risk, the tightness of the leverage constraint, and the dynamic distribution of assets. The aggregate margin, in turn, is a time-varying average of the cross-section. Because of non-linearities and aggregate uncertainty, this average does *not* correspond necessarily to the margin of the average intermediary.

Although in order to compute margins in the full model we need to resort to numerical methods, we can derive several analytical results to show how credit margins are determined in a simpler version of the model. First, assume that we solve a static problem instead of the dynamic one. That is, ignore aggregate uncertainty. Second, treat v(j), $\bar{\mathrm{R}}(\mathrm{j})$, and $\mathrm{R}^{\mathrm{T}}(\mathrm{j})$ as given. Third, assume the occasionally binding leverage constraint is always slack. We can solve for the bank-level rate-setting rule:

Proposition 1 (Bank Price-Setting Rule)**.** *The price-setting rule* $\frac{\mathrm{p(j)}}{\mathrm{P}(\boldsymbol{S})}$ *on each local credit market j is:*

$$\frac{\mathrm{p(j)}}{\mathrm{P}(\boldsymbol{S})} = \left[\frac{\beta\theta - 1}{\theta - 1}\frac{\bar{\mathrm{R}}(\mathrm{j})}{\big(1 - \upsilon(\mathrm{j})\big)\mathrm{R}^{\mathrm{T}}(\mathrm{j})}\mathrm{K}(\boldsymbol{S})^{\beta-1}\right]^{\frac{1}{\theta(\beta-1)}} \tag{17}$$

and the marginal cost is:

$$\mathrm{MC(j)} := \frac{\beta\theta - 1}{\theta}\mathrm{p(j)}\frac{\bar{\mathrm{R}}(\mathrm{j})}{\big(1 - \upsilon(\mathrm{j})\big)\mathrm{R}^{\mathrm{T}}(\mathrm{j})}\left[\left(\frac{\mathrm{p(j)}}{\mathrm{P}(\boldsymbol{S})}\right)^{-\theta}\mathrm{K}(\boldsymbol{S})\right]^{\beta-1} \tag{18}$$

Proof: Appendix B

The proposition clarifies an important distinction between credit *margins* and *markups*. A credit margin, χ(j), is a measurable ratio of the credit rate p(j) over the deposit rate $\bar{\mathrm{R}}(\mathrm{j})$. In our model, χ(j) is endogenous, heterogeneous across local credit markets, and aggregate state-dependent. The markup, however, given the CES assumption, is *constant* across all dimensions and depends only on the structural parameter θ.

In addition, the proposition clarifies the determinants of bank j's endogenous marginal cost. First, it is a function of the *relative* cost of funds $\frac{\bar{\mathrm{R}}(\mathrm{j})}{\mathrm{R}^{\mathrm{T}}(\mathrm{j})}$. This is due to the fact that the marginal revenue, unlike standard models of monopolistic competition in product markets, depends not only on revenues p(j)k(j) but also on the return to investment $\mathrm{R}^{\mathrm{T}}(\mathrm{j})$. Second, there is a scale effect factor in $\mathrm{K}(\mathbf{S})^{\beta-1}$. Finally, the marginal cost is negatively related to the probability of default ν(j).

3.6 The Incumbent Banker Problem

We now detail the dynamic problem of the incumbent intermediary. We follow recursive notation from now on: the solution does not depend on the specific local credit market j but only on the relevant state variables. Define $\mathbf{s} = \{\mathrm{n}, \xi\}$ and $\mathbf{S} = \{\psi, \mu, \mathrm{M}\}$ as the bankers' idiosyncratic and aggregate state vectors, respectively. Conditional on the state vector $\{\mathbf{s}, \mathbf{S}\}$, each banker maximizes its franchise value which is defined as the discounted stream of future flows of net worth. The bank discounts the future by adopting and augmenting the household's stochastic discount factor $\Lambda(\mathbf{S})$, which is determined in equilibrium together with household optimization. Each banker takes as given the aggregate uncertainty process $\{\psi\}$, aggregate quantities $\{\mathrm{K}(\mathbf{S}), \mathrm{N}(\mathbf{S})\}$, aggregate prices $\{\mathrm{P}(\mathbf{S}), \mathrm{R}^{\mathrm{k}}(\mathbf{S})\}$, variety-specific deposit rates $\bar{\mathrm{R}}(\mathbf{s}, \mathbf{S})$, and the law of motion of the distribution Γ. The generic incumbent banker therefore solves:

$$\mathrm{V}(\mathbf{s}, \mathbf{S}) = \max_{\{\mathrm{k},\mathrm{p},\mathrm{d}\}} \left\{ \mathbb{E}_{\mathbf{S}'|\mathbf{S}} \left[\Lambda(\mathbf{S}') \Big((1-\sigma)\mathrm{n}' + \sigma \mathrm{V}(\mathbf{s}', \mathbf{S}') \Big) \right] \right\} \tag{19}$$

s.t. conditions 1-14.

We can simplify the problem substantially by rewriting it as a one-argument problem. Each bank now chooses the leverage ratio $\phi = \frac{\mathrm{pk}}{\mathrm{n}}$ to:

$$\max_{\phi} \left[\mu_{\mathrm{a}} \phi + \nu_{\mathrm{a}} \right] \tag{20}$$

subject to the same constraints as before, where $\mu_{\mathrm{a}} = (1-\upsilon)\tilde{\Lambda}(\mathbf{S}') \left[\mathrm{R}^{\mathrm{T}}(\mathbf{s}', \mathbf{S}') - \mathrm{k}^{\beta-1}\bar{\mathrm{R}}(\mathbf{s}, \mathbf{S}) \right]$ is the excess return on risky investments, and $\nu_{\mathrm{a}} = (1-\upsilon)\tilde{\Lambda}(\mathbf{S}')\bar{\mathrm{R}}(\mathbf{s}, \mathbf{S})$ is the cost of liabilities. In both instances, $\tilde{\Lambda}(\mathbf{S}') = \Lambda(\mathbf{S}') \left(1 - \sigma + \sigma \mathrm{V}(\mathbf{S}')\right)$ is the *augmented* stochastic discount factor. Note how the financial intermediary stochastic discount factor depends the household's discount factor (because all bank equity ultimately belongs to the consumer) and on the exogenous exit rate σ.

The leverage ratio, when internalizing the credit demand system, can be written:

$$\phi = \frac{\mathrm{pk}}{\mathrm{n}} = \left(\frac{\mathrm{k}}{\mathrm{K}(\mathbf{S})} \right)^{-\frac{1}{\theta}} \frac{\mathrm{P}(\mathbf{S})\mathrm{k}}{\mathrm{n}} \tag{21}$$

From equation 21 we see how leverage depends on relative prices p(j), which in turn are set conditional on the elasticity of substitution θ. Varying the degree of banking competition feeds directly to the bank's preference for risk-taking.

For illustration, we now contrast our ϕ with the standard definitions of market leverage from GK models. To get to the GK version precisely, we need to set $\theta \to \infty$ and $\beta = 1$, as well as $\sigma_{\xi} = 0$. Bank heterogeneity then collapses to the case of a representative, homogeneous intermediary. In that case, bank leverage becomes $\phi_{\mathrm{GK}} = \frac{\mathrm{P}(\mathbf{S})\mathrm{k}}{\mathrm{n}}$, which can be seen cleanly from Equation 21. Because all relative prices collapse to unity, the market power channel is shut off.

Additionally, it can be seen from the expression for μ_{a} above how the value function depends on bank characteristics when $\beta > 1$. Notice how μ_{a} depends explicitly on the choice of k. With decreasing returns, the bank internalizes the impact of balance sheet choices on excess returns which in turn feed into the value of the franchise. With constant returns, bank-level characteristics do not matter and the state vector includes only aggregate variables. We discuss the role of bank size heterogeneity more formally in the next section.

3.7 Bank Size Heterogeneity and Scale Variance

An important departure of our framework from Gertler and Kiyotaki (2010) and Gertler and Karadi (2011) is that the value function of the incumbent is *not* linear in bank-level net worth. That is, the solution is not invariant to the scale of the intermediary. The role of bank size heterogeneity can be best visualized from the following proposition, where we re-introduce the (j) notation for additional emphasis on heterogeneity:

Proposition 2. *The solution to the incumbent banker's problem for each j, conditional on* $\beta > 1$*, the aggregate state vector* $\boldsymbol{S}$*, initial net worth* $\mathrm{n(j)}$ *and idiosyncratic state* $\xi\mathrm{(j)}$ *is*

$$\mathrm{V}\Big(\mathrm{n(j)}, \xi\mathrm{(j)}; \boldsymbol{S}\Big) = \zeta\Big(\mathrm{n(j)}, \xi\mathrm{(j)}; \boldsymbol{S}\Big)\mathrm{n(j)}$$

where the marginal value of net worth is:

$$\zeta\Big(\mathrm{n(j)}, \xi\mathrm{(j)}; \boldsymbol{S}\Big) = \frac{\Big(1 - \upsilon\mathrm{(j)}\Big)\mathbb{E}\left(\Lambda(\boldsymbol{S}')\Big[1 - \sigma + \sigma\zeta\Big(\mathrm{n'(j)}, \xi'\mathrm{(j)}; \boldsymbol{S}'\Big)\Big]\mathrm{k(j)}^{\beta-1}\bar{\mathrm{R}}\mathrm{(j)}\right)}{1 - \varphi\Big(\mathrm{n(j)}, \xi\mathrm{(j)}; \boldsymbol{S}\Big)} \tag{22}$$

and the multiplier on the moral hazard leverage constraint is

$$\varphi\Big(\mathrm{n(j)}, \xi\mathrm{(j)}; \boldsymbol{S}\Big) = \max\left[1 - \frac{\Big(1 - \upsilon\mathrm{(j)}\Big)\mathbb{E}\left(\Lambda(\boldsymbol{S}')\Big[1 - \sigma + \sigma\zeta\Big(\mathrm{n'(j)}, \xi'\mathrm{(j)}; \boldsymbol{S}\Big)\Big]\mathrm{k(j)}^{\beta-1}\bar{\mathrm{R}}\mathrm{(j)}\right)\mathrm{n(j)}}{\lambda \mathrm{k(j)}^{1-\frac{1}{\theta}}\Big(\mathrm{K}(\boldsymbol{S})\Big)^{\frac{1}{\theta}}\mathrm{P}(\boldsymbol{S})}, 0\right] \tag{23}$$

Proof: Appendix B

Proposition 2 shows that bank-level characteristics matter for aggregation. In particular, the distributions of both bank net worth *and* idiosyncratic return shocks $\xi\mathrm{(j)}$ are state variables. The former is due to $\beta > 1$ and $\theta > 1$, whereas the latter is due to the persistence of the shock process and the incomplete markets assumption.

3.8 Cyclical Entry and Exit

There is a large number of new, potential financial varieties which are managed by financiers. Before entry, the financier pays a fixed equity issuance cost e in units of aggregate capital. After paying the sunk cost, the financier receives an idiosyncratic return profitability draw $\xi_0 \in \Xi$ from the ergodic distribution $\mathrm{G}_0(\xi)$ that governs ξ. The new entrant is also bestowed with a starting level of net worth $\mathrm{n}_0(\mathbf{S}) = \iota \mathrm{N}_t$ where $0 < \iota < 1$. This assumption is motivated by pro-cyclical bank entry in the data. Immediately afterwards, the entrant decides whether to operate or to exit. Conditional on staying, the financier becomes a new financial variety and adds to J_t. Conditional on its startup state vector $\{\mathrm{n}_0(\mathbf{S}), \xi_0\}$, the entrant operates if its expected discounted franchise value exceeds e. The value function for the entering financial variety is therefore:

$$\mathrm{V}^e(\mathrm{n}_0, \xi_0; \mathbf{S}) \equiv \max\left[\mathbb{E}\mathrm{V}(\mathbf{s}; \mathbf{S}) - e, 0\right] \tag{24}$$

In equilibrium, either V^e is equal to 0, the number of entrants is 0, or both. The incumbents are subject to two sources of exit risk: the involuntary exogenous exit rate σ and the endogenous probability of default $\nu(\mathrm{j})$ which is specific to every local market. If a financial variety exits, its market niche would never be taken over by any of the incumbents. Importantly, the extensive margin dynamic depends explicitly on the state of the business cycle. The two main factors that impact financiers' entry decisions are bank franchise values $\mathrm{V}(\mathbf{S})$ and startup equity injections $\mathrm{n}_0(\mathbf{S})$. Both are procyclical since aggregate demand for credit (and the stock of capital) are positively related to ψ. That immediately implies that equilibrium entry is also procyclical, in line with the data.

3.9 Dynamics of the Cross-Sectional Distribution of Banks

Define π^e as a mass of financial varieties that exits either due to endogenous default or because of the exogenous exit shock. The cross-sectional distribution evolves according to:

$$\mu'(\mathrm{n}, \xi_i) = (1 - \pi'^{e}) \int_{\{(\mathrm{n},\xi_i)|\mathrm{K}(\mathrm{n},\xi_i;\mathbf{S})\in\mathbf{B}\}} \mathrm{G}_{ji}\mu(d\mathrm{n}, d\xi_i) + \mathrm{M}' \int_{\{(\mathrm{n}_0,\xi_i)|\mathrm{K}(\mathrm{n},\xi_i;\mathbf{S})\in\mathbf{B}\}} \mathrm{G}_0(\xi_i), \forall \xi_i \in \Xi \tag{25}$$

Recall that $\mathrm{G}_0(.)$ is the CDF of ξ for new entrants and $\mathrm{G}_{x'x}$ is the Markov chain for ξ of the incumbent.

3.10 Households

For simplicity, assume inelastic labor supply normalized to 1. The representative household chooses the supply of deposits to each financial variety, $\mathrm{b}_t(\mathrm{j})$, and consumption C_t, subject to the budget constraint:

$$\max_{C_t, b_t(j)} \left[\mathbb{E}_t \sum_{t=1}^{\infty} \beta_h^t u(C_t) \right]$$

subject to

$$C_t + \int_0^{J_t} b_t(j)dj \leq W_t + \int_0^{J_t} \bar{R}_t(j) b_{t-1}(j)dj + \int_0^{J_t} \pi_t(j)dj$$

where W_t is the equilibrium wage and π are profits (net of any transfers) from bank ownership redistributed back to the household lump sum. The first order condition for deposits yields the following equation:

$$\bar{R}_t(j) = \frac{1 - \upsilon_t(j) x_t(j) \mathbb{E}\left(R^T_{t+1}(j)\Lambda_{t+1}\right)}{\left(1 - \upsilon_t(j)\right)\mathbb{E}\left(\Lambda_{t+1}\right)} \tag{26}$$

where $\Lambda_{t+1} = \beta_h \frac{u'(c_{t+1})}{u'(c_t)}$ is the stochastic discount factor. Deposits are risky because there is no deposit insurance. The consumer prices bank default risk into the distribution of variety-specific deposit rates, which depend on the deposit *recovery rate* $x_t(j)$:

$$x_t(j) = \min\left[\frac{\phi_t(j)}{\phi_t(j) k_t(j)^{\beta-1} - 1}, 1\right]$$

with $\phi_t(j)$ being the market leverage ratio, defined as before.

3.11 Recursive Industry Equilibrium

Credit market clearing requires:

$$K(\mu, \mathbf{S}) = \int_{\mathbf{B}} \left(k(\mathbf{s}; \mathbf{S})\right)\mu(dn, d\xi) + M(\mathbf{S}) \int_{\mathbf{B}} \left(k(n_0, \xi_0; \mathbf{S})\right) dG(\xi_0) + M(\mathbf{S})e \tag{27}$$

where the first term on the right hand side is total demand by incumbents, the second term is total demand by entrants, and the final term is the entry cost paid by the entrants. Similarly, equilibrium in the market for bank deposits requires:

$$\int_0^J b(j)dj = \int_{\mathbf{B}} \left(d(\mathbf{s}; \mathbf{S})\right)\mu(dn, d\xi) \tag{28}$$

Goods market clearing requires production goods to be used either for household consumption or firm investment. The latter includes investment demand that is intermediated both by the incumbent and the new entrants.

$$Y(\mathbf{S}) = C(\mathbf{S}) + I(\mathbf{S})$$

A *recursive industry equilibrium* is defined as a set of functions that include the value function of the banker $V(\mathbf{s};\mathbf{S})$, optimal policies for bank capital investment $k(\mathbf{s};\mathbf{S})$ and deposit demand $d(\mathbf{s};\mathbf{S})$, household consumption $c(b_{-1};\mathbf{S})$ and deposit supply $b(b_{-1}(j);\mathbf{S})$, the mass of new bank entrants $M(\mathbf{S})$, competitive wage $W(\mathbf{S})$ and capital $R_k(\mathbf{S})$ pricing functions, the aggregate price index of financial varieties $P(\mathbf{S})$, a marginal utility process $\Lambda(\mathbf{S})$, and the menu of market-clearing deposit rates $\bar{R}(\mathbf{s};\mathbf{S})$ such that:

1. The household's choices $\{C(\mathbf{S}), b(j)\}$ are optimal conditional on $\{W(\mathbf{S}), \bar{R}(\mathbf{s};\mathbf{S})\}$.
2. The banker's choices $\{k(\mathbf{s};\mathbf{S}), p(\mathbf{s};\mathbf{S}), d(\mathbf{s};\mathbf{S})\}$ are optimal conditional on $\{\psi, \Lambda(\mathbf{S}), K(\mathbf{S}), P(\mathbf{S}), \bar{R}(\mathbf{s};\mathbf{S}), \mu(\mathbf{S})\}$
3. The returns on factors of production are: $R^k(\mathbf{S}') = \psi'\left(\frac{\alpha A K(\mathbf{S}')^{\alpha-1}+(1-\delta)P(\mathbf{S}')}{P(\mathbf{S})}\right)$, $W(\mathbf{S}) = (1-\alpha)K(\mathbf{S})^{\alpha}$.
4. $\{K(\mathbf{S}), D(\mathbf{S}), N(\mathbf{S})\}$ are consistent with the cross-sectional distribution and the monopolistic credit demand system in (4)-(7).
5. The free-entry condition (24) is satisfied and is consistent with individual choices.
6. The credit market clears as in (27). The deposit market clears for each variety j as in (28).
7. The cross-sectional distribution evolves according to (25) and is consistent with bank-level demand functions.

3.12 Symmetric, Non-Stochastic Steady State

To shed more light on the equilibrium properties of the model, it is useful to inspect features of the symmetric, non-stochastic steady state. In doing so, and following the result in Proposition 1, we pin down the aggregate interest rate on deposits. It is determined in a static equilibrium where aggregate price levels and interest rates are obtained simultaneously for a given level of aggregate demand for credit. To that goal, we require additional assumptions, for tractability. First, we set $\xi(j)$ to the ergodic mean for all j. Second, we consider symmetric equilibria only, i.e., when $p(j) = P\ \forall j$. By extension, this implies homogeneous probabilities of default $\upsilon(j)$ and deposit interest rates. Note that this corresponds to analyzing a representative (average) banker from the distribution, rather than shutting down the default risk or idiosyncratic returns channels completely. The aggregate cost of funds is determined in the following proposition.

Proposition 3 (Aggregate Rate Rule)**.** *The aggregate rate-setting rule in the banking sector is*

$$\bar{R} = \left[\frac{\beta\theta-1}{\theta-1}\frac{K^{\beta-1}}{R^T(1-\nu)}\right]^{-1} \tag{29}$$

Proof. Follows immediately from Proposition 1 after assuming symmetry in the price-setting rule, i.e. in equation 17 set $p(j) = P\ \forall j$ and $R(j) = R\ \forall j$.

□

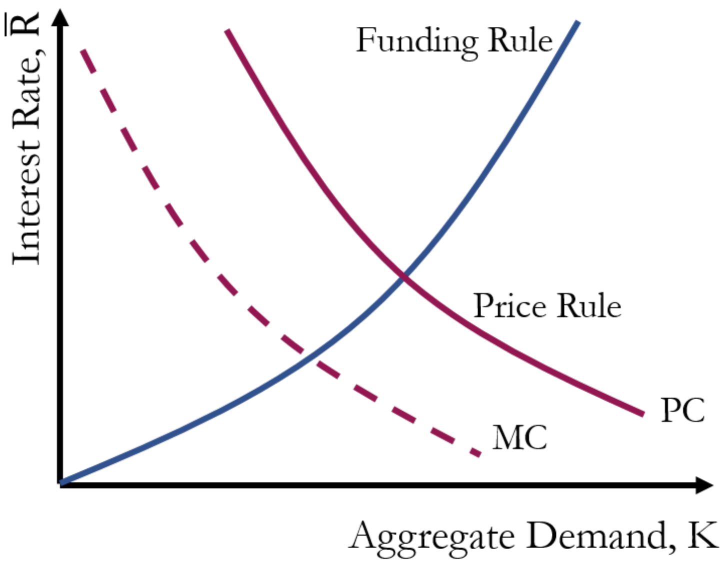
The aggregate rate-setting schedule is a downward-sloped demand curve for bank lending. Note that the slope of the line is independent of the elasticity of substitution, which acts as a horizontal curve shifter. In the limiting case of θ approaching infinity, the special case of perfect competition in the credit market is achieved. A symmetric equilibrium is defined by the intersection with the upward-sloping funding cost rule, equation 26 from the household's problem. It is straightforward to see that the rate is increasing in K because (a) $\bar{R}(j)$ is decreasing in the deposit recovery rate $x(j)$ and (b) $x(j)$ is decreasing in the leverage ratio $\phi(j)$ and $k(j)$. Everything else equal, as debt-financed capital grows, the recovery rate falls and the deposit rate goes up. In the symmetric equilibrium, the average rate $\bar{R}$ is increasing in the aggregate stock of capital K.

A static, symmetric equilibrium generally exists when $\beta > 1$. We can visualize the equilibrium graphically, for some given values of ν and $R^T > 0$ as well as $\beta > 1$ and $\theta > 1$. Figure 4 portrays graphically symmetric equilibria under the alternative scenarios of perfect (PC) and monopolistic (MC) competition in lending. Monopolistic competition is our baseline case when $\theta > 1$ but is finite. Perfect competition is the theoretical limit of $\theta \to \infty$. The downward-sloping curve is the aggregate rate-setting rule (Equation 29). On the horizontal axis we have quantities - the aggregate demand for credit K. On the vertical axis are aggregate interest rates.

As θ falls, greater credit market power shifts the pricing rule leftward with no immediate effects on the funding schedule. The wedge between the MC and PC curves grows and the social deadweight loss from credit markups increases. This yields a decline in both aggregate demand and the cost of bank funds. Credit market power leads to an *aggregate under-lending externality*. In the monopolistic competition (MC) equilibrium, aggregate demand is strictly lower than in the perfect competition (PC) counterfactual.[6] Our result parallels the canonical under-utilization of labor resources in Blanchard and Kiyotaki (1987). Note that the only resource used in the production of financial varieties is capital k(j), thus the under-utilization of risky capital in equilibrium. In a more sophisticated setup, if the banker also employed labor in order to supply each additional unit of k(j), labor would be also underutilized in equilibrium. Assuming no market power on the liability side of the balance sheet, the credit supply externality would only strengthen in this case.

[6]This result stems from our assumption that all non-financial firms depend on bank funding. In practice, the relevance of the externality scales with the share of bank-dependent firms in the distribution. Clearly, if firms can finance investment with retained earnings or equity issuance, the credit supply externality would have little to no bite. This observation is true not only for our environment but also for the general class of GK models.

Figure 4: **Monopolistic Banking Equilibrium - Visualization**



Notes: This figure visualizes the static, symmetric equilibrium with monopolistic competition in bank lending. The downward-sloping curve is the aggregate pricing rule from Equation 29. The upward-sloping line is the deposit supply schedule from the household's problem. The graph highlights equilibria under monopolistic (MC) and perfect competition (PC) in the bank credit market. The MC case reflects the baseline scenario of a finite $\theta > 1$. The PC equilibrium is approximated by $\theta \to \infty$. On the horizontal axis is aggregate demand for credit K. On the vertical axis is the aggregate rate $\bar{R}$.

3.13 Numerical Algorithm

There are several considerable computational challenges in solving our model. First, the financial sector needs to construct forecasts for the return to aggregate capital $R^{k'}$, which depends on the forecasts for K' and P'. Both aggregates, in turn, potentially depend on the whole distribution of bank assets k(j) - an infinitely dimensional object. We solve this problem with a variant of the stochastic simulation and partial aggregation method of Maliar et al. (2010). This method builds on the general algorithm for solving models with heterogeneous agents and aggregate uncertainty that was originally proposed in Krusell and Smith (1998). In the algorithm, linear forecasts are cast on a number of *moments* of the banking distribution. Specifically, we assume a linear projection of K' and P' on m moments of the k(j) distribution. In the baseline case, we will track only the mean. In Section 6.4 we will also track the second moment. Second, the deposit market must clear all points of the state space. Third, the bank adopts and augments the households' stochastic discount factor $\Lambda(S)$, which is an endogenous state variable that must be kept track of. Fourth, the dynamic distribution of bank net worth must be consistent with both endogenous entry and exit. Finally the bank faces an occasionally binding leverage constraint which could bind on any part of the idiosyncratic or aggregate state space. Below we describe all steps of our algorithm in more detail.

0. Solve a simpler version of the model without aggregate uncertainty. Use the solution as an initial condition for the full model with aggregate risk. Construct grids for aggregate capital and the aggregate shock. For individual bank net worth and household wealth, we use an unequally spaced grid with more points on lower values of n(j) and b_{-1}.

1. Solve the problem of the representative household given equilibrium wages and deposit rates. We use the endogenous gridpoints method for speed (Carroll, 2006). Construct $\Lambda(\mathbf{S})$.

2. Solve the financial intermediation problem in three steps

   (a) Conjecture a starting law of motion for aggregate capital $\log(K') = \eta_k\Big(\log(m_k)\Big)$ and the projection for prices $\log(P') = \eta_p\Big(\log(m_k)\Big)$. Construct $R^{k'}$.

   (b) Given η_k, η_p, Λ and initial guesses for V(j), ν(j), and $\bar{R}$(j) solve the bankers' problem using value function iteration. To handle the occasionally binding leverage constraint, first guess that the constraint is binding on a point in the grid space. Compute the implied Lagrange multiplier. If the constraint is slack, re-solve the problem using global maximization routines while ignoring the constraint.

   (c) Compute the implied policy function for deposit demand. Update the interest rates on deposits, compare with the initial guess, and iterate upon convergence is achieved on each point of the state space.

3. Using the newly computed policy functions, run a large simulation of N varieties over T quarters where incumbents are subject to both idiosyncratic and aggregate shocks. Compute the implied time-varying distribution of bank net worth n(j), the supply of investment into firm claims k(j), and the prices of claims p(j).

4. Solve the optimal entry problem in 24 and determine the time-varying mass of entrants M for each quarter of T.

5. Construct the sequence of the aggregate demand for capital K that tracks demand of the incumbent and of the new entrants. Run linear regressions $K' = \hat{\eta}_k(m_k)$ and $P' = \hat{\eta}_k(m_k)$. Compare regression coefficients with the original η_k and η_p. If convergence is not achieved, return to step 2(a) and continue until convergence of the inner loop.

6. When the inner loop converges, construct updated versions of competitive wages and deposit rates using optimal ν(j) and ϕ(j). Return to step 1, re-solve the household problem and construct a new $\Lambda(\mathbf{S})$. Continue with the inner loop. Keep iterating the program until

convergence of the outer loop is achieved. Accuracy of the algorithm is discussed in Appendix C.

4 Quantifying the Model

We begin a quantitative illustration of the model by first reporting the details of our calibration.

4.1 Calibration

Table 2 displays the parameter values chosen for the model calibration. The unit of time in the model is a quarter. We start by describing standard macro parameters. The share of aggregate capital in production is set to $\alpha = 0.36$. The discount rate β_{h} =0.996 targets the steady-state annual risk-free rate of 1.56%. Aggregate capital $\mathrm{K_t}$ depreciates at the quarterly rate $\delta = 0.025$. We assume log-utility in consumption ($\sigma_{\mathrm{h}} = 1$).

For parameters in the banking block, the dividend payout ratio is set to $\sigma = 0.9$. This number is broadly consistent with the exit rate of financial intermediaries in the U.S. According to the Federal Deposit Insurance Corporation, there were about 11000 commercial banks in the U.S. at the start of 1980. This number has fallen below 5000 by 2020. This implies an approximate annual exit rate of 3% and a life expectancy of a banker of about 8.25 years. In the model, that number is 10 conditional on zero default risk. The fraction of bank assets that are divertible by the manager is $\lambda = 0.12$. This number targets a steady state bank leverage ratio of 8. Endowment of new entrants and the fixed cost of entry are calibrated in order to keep the net entry rate relatively stable over the cycle at around 5%.

The monopolistic banking block requires calibration of two main parameters. First, the elasticity of substitution $\theta = 2$ implies an average stationary credit margin of 1.48 over the cost of funds, which is broadly in line with the existing evidence on loan margins in the financial sector (De Loecker et al., 2020; Diez et al., 2018). Jamilov (2020) estimates the average nationwide elasticity of substitution across U.S. commercial banks to be roughly 1.2, and we pursue a more conservative calibration approach. The returns to scale parameter of $\beta = 1.01$ suggests minor dis-economies of scale and almost-constant returns. The calibrated value is consistent with empirical evidence on some dis-economies of scale in European banking (Anolli et al., 2015).

Calibration of the idiosyncratic return process follows closely the recent work by Galaasen et al. (2020). Galaasen et al. (2020) estimate the pass-through of idiosyncratic firm shocks on bank-level returns using matched bank-firm data from Norway. They estimate the shock process and find annual persistence and standard deviation parameters of $\rho_{\xi} = 0.11$ $\sigma_{\xi} = 0.25$. That is, the idiosyncratic rate of return shock in banking is (a) volatile and (b) not very persistent. These

Table 2: **Model Parameters**

Parameter	Description	Value
	Standard Macro	
α	Share of capital in production	0.36
β_h	Discount factor	0.996
σ_h	Household risk aversion	1
δ	Capital depreciation rate	0.025
	Banking	
σ	Dividend payout ratio	0.9
λ	Share of divertible assets	0.12
ι	Entry starting endowment	30% of N_t
e	Entry fixed cost	0.11
	Monopolistic Competition	
θ	Credit demand elasticity	2
β	Local returns to scale	1.01
	Idiosyncratic and Aggregate Risk	
κ	Fraction of returns exposed to idiosyncratic risk	0.5
ρ_ξ	Serial correlation of idiosyncratic risk	0.5
σ_ξ	SD of idiosyncratic risk	0.1
ρ_ψ	Serial correlation of aggregate risk	0.914
σ_ψ	SD of aggregate risk	0.015

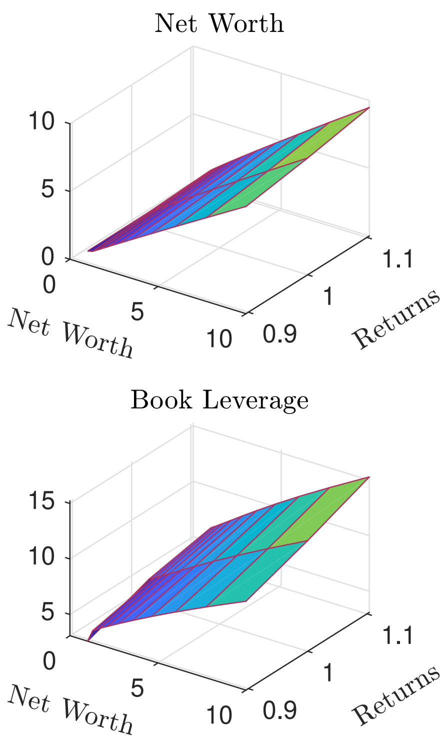
Notes: Parameters that are exogenously fixed in the model calibration.

values are in line with our chosen quarterly parameters in the Table. The fraction of financial wealth exposed to idiosyncratic risk $\kappa = 0.5$ is in line but slightly on the upper side of the pre-financial crisis share of the shadow banking business in overall U.S. banking (Gorton and Metrick, 2010). For the aggregate shock process, we pick σ_ψ and ρ_ψ such that output volatility in the model corresponds roughly to that of the pre-Crisis period. Both idiosyncratic and aggregate processes, in order to keep the state space manageable, are discretized with the Tauchen (1986) method.

4.2 Model Policy Functions

Figures 5 and 6 plot two-dimensional policy functions for key financial-sector quantities, prices, and measures of risk. The two dimensions are the idiosyncratic states: initial net worth (size) $n_t(j)$ and returns $\xi_t(j)$. Figure 5 reports quantities. (Book) leverage is defined as total assets over total net worth. Overall, banks with larger $n_t(j)$ and $\xi_t(j)$ tend to choose greater quantities of future net

Figure 5: **Model Policy Functions - Quantities**



Notes: Optimal choices represented as two-dimensional surfaces. The two dimensions are idiosyncratic state variables in the banking sector: initial net worth n(j) and idiosyncratic return ξ(j). Leverage is in book values.

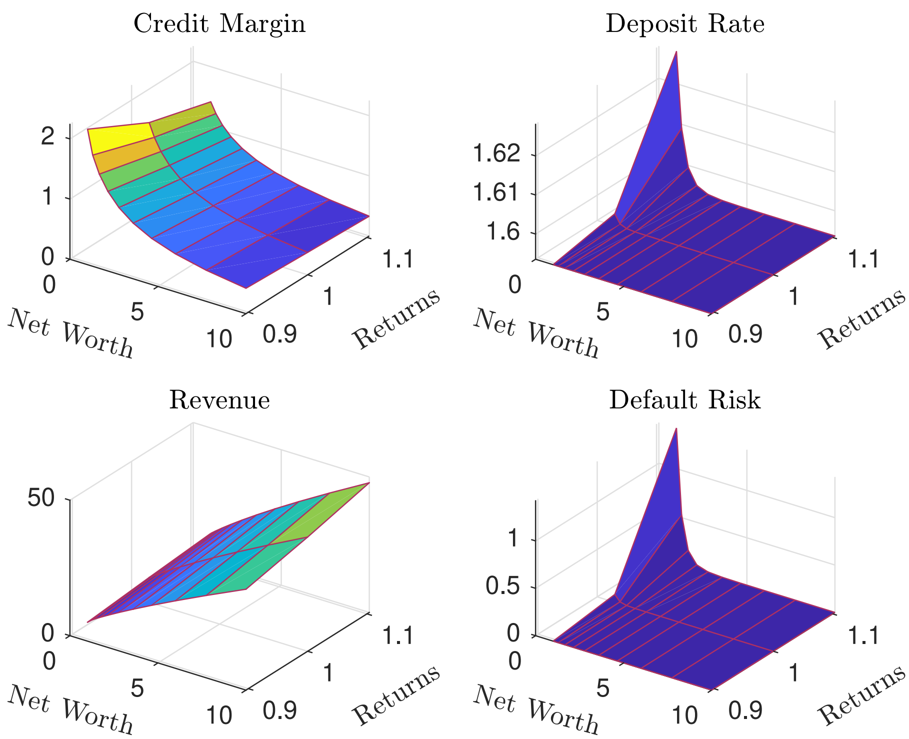
worth $n_{t+1}(j)$ and assets $k_{t+1}(j)$. Such banks are also more (book) levered, which is consistent with the empirical facts from Coimbra and Rey (2019) and Gopinath et al. (2017) on both financial and non-financial firms. This is also in line with our empirical analysis from Section 2.2. Finally, larger and more profitable banks have a greater franchise value $V_{t+1}(j)$.

Figure 6 reports policy functions of prices and measures of risk. Generally, credit margins, χ(j), decline with bank size and profitability in the cross-section. This feature of the model is a consequence of us working with CES aggregation. Noticeably it is consistent with the empirical evidence in Section 2.2 documenting that banks with higher net worth charge lower margins. Larger, more profitable banks also earn more in total revenues $p_t(j)k_t(j)$. Small institutions face elevated risks of insolvency $\nu_t(j)$, which is priced in the interest rate on deposits $\bar{R}(j)$.

4.3 Ergodic Cross-Sectional Distributions

Figures 7 and 8 present two-dimensional histograms for bank book leverage, default risk, relative prices, and deposit rates. These cross-sectional distributions are ergodic, i.e., obtained

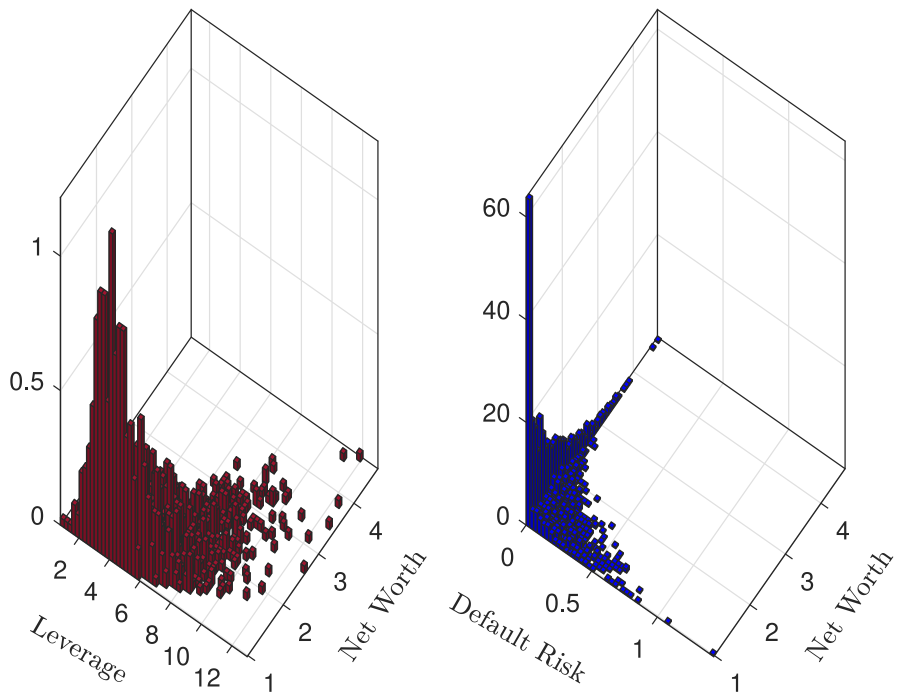
Figure 6: **Model Policy Functions - Prices**



Notes: Optimal choices represented as two-dimensional surfaces. The two dimensions are idiosyncratic state variables in the banking sector: initial net worth n(j) and idiosyncratic return ξ(j). Deposit rates are in percent, annualized. Default risk is in percent, annualized.

from the recursive general equilibrium with aggregate uncertainty. One of the two axes represents net worth (size) n(j). The other axis is the plot-specific variable of interest. The distribution of leverage is centered around 8, our steady state target. It is right-skewed, i.e., there is a small number of banks with high leverage ratios. This is consistent with the cross-sectional facts reported in Section 2.2 above. Leverage also grows with initial net worth, which was also visible from the model policy functions. The distribution of margins χ(j) is centered around 1.5, consistent with our target. Credit margins decline with the size of the intermediary, as mentioned before. Finally, the distribution of default (insolvency) risk is concentrated in the left tail of the bank net worth distribution. Small institutions face the probability of insolvency that is as high as 1% on a yearly basis. This distribution is priced into the cross-section of deposit rates $\bar{\mathrm{R}}$(j): smaller, riskier institutions must compensate the risk-averse household for "investing" into a risky bank. In equilibrium, the premium is reflected in higher interest rates on deposits.

Figure 7: **Ergodic Distributions - Risk-Taking**



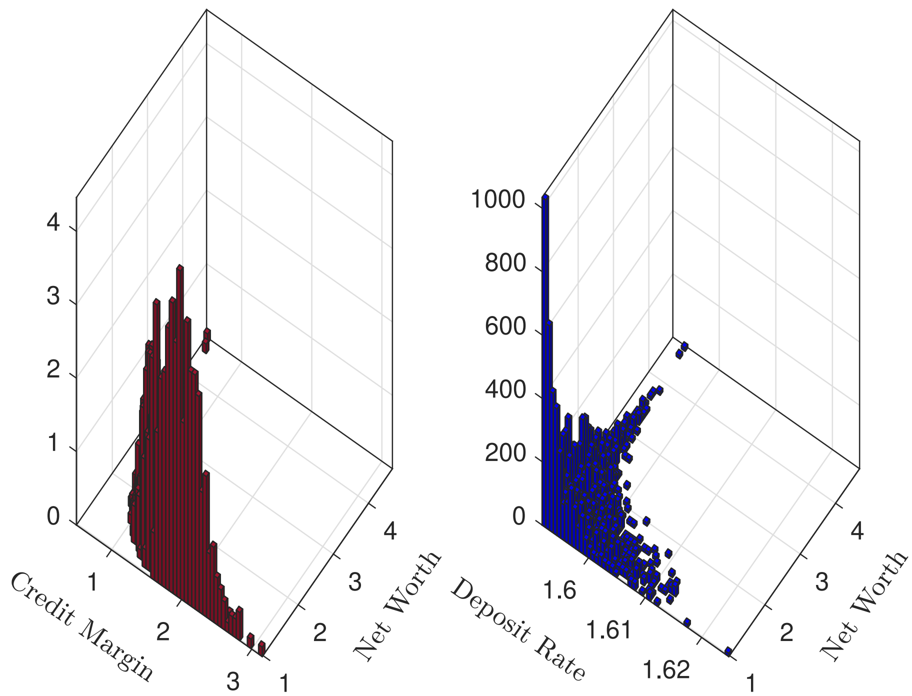
Notes: ergodic cross-sectional distributions obtained from the recursive competitive equilibrium. Default risk is in percent, annualized. Leverage is in book values. Figures plot the pdf normalization.

4.4 Credit Cycle Statistics

Table 3 reports unconditional standard deviations and correlations with output from our model economy. We focus on the financial cycle and the first three moments of the distributions of bank assets, net worth, credit margins, default risk, and book leverage. In order to obtain these moments, we simulate the model for 10,000 quarters. We see that model-based standard deviations are smaller than in the data. This is due to the fact our statistics are based on a model with a single aggregate shock ψ. We correctly match the regularity that higher-order moments of the distributions (especially concentration) are typically more volatile than the mean.

The table also reports model-implied credit cycle correlations. In terms of correlations, our model can match the data well. We match the pro-cyclicality of the first two moments of financial assets and net worth. We also obtain - correctly - negative correlations for the concentration of both assets and net worth. In the model credit margins are counter-cyclical, which is in line with the data. The model predicts pro-cyclical second and third moments of the margins distribution, while in the data those moments are negative and positive, respectively. All three moments of the distribution of

Figure 8: **Ergodic Distributions - Prices**



Notes: ergodic cross-sectional distributions obtained from the recursive competitive equilibrium. Deposit rates are in percent, annualized. Figures plot the pdf normalization.

default risk have correct cyclical properties. Book leverage in the model is pro-cyclical in all three moments. In the data, concentration of leverage is counter-cyclical. Finally, both the total number of incumbent intermediaries and the mass of new entrants is correctly pro-cyclical in the model. Overall, the model is able to replicate 15 out of the 17 free, untargeted correlations, including the entry rate and the number of active intermediaries, which are both pro-cyclical as also reported in Corbae and D'Erasmo (2019).

4.5 Financial Recessions

Next we characterize equilibrium dynamics in the model. We study the model behavior in response to a (one standard deviation) negative ψ (capital quality) shock. This shock is representative of a worsening of banks' balance sheets. Our numerical approach consists of two general steps. First, we run a "benchmark" simulation of our model economy for 10,000 periods. Second, in some quarter T*, the economy is hit with a counterfactually low realization of the aggregate capital quality shock ψ_{T*}. It is then allowed to revert back to the benchmark path with its normal autocor-

Table 3: **Credit Cycle Statistics - Data and Model**

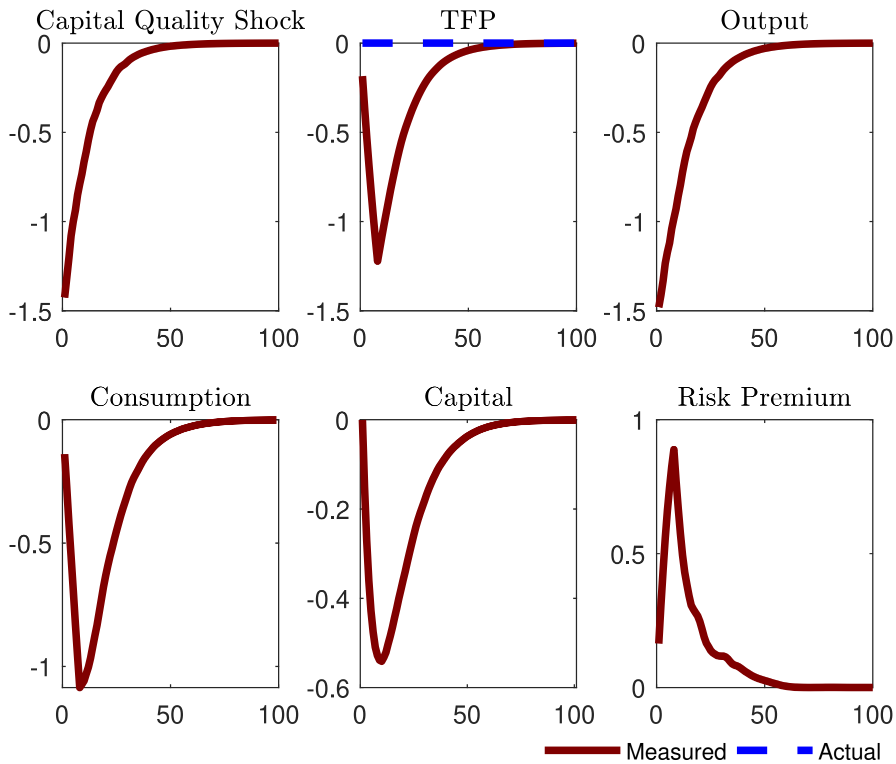
Variable	Data		Model	
	Standard Deviation	Correlation with Y_t	Standard Deviation	Correlation with Y_t
Assets Mean	13.383	0.498	2.316	0.798
Assets Dispersion	19.371	0.642	4.276	0.541
Assets HHI	18.281	-0.568	13.897	-0.103
Net Worth Mean	11.268	0.211	1.920	0.842
Net Worth Dispersion	18.076	0.544	3.837	0.683
Net Worth Concentration	16.64	-0.472	8.553	-0.238
Margins Mean	31.046	-0.563	0.765	-0.305
Margins Dispersion	42.404	-0.370	1.829	0.437
Margins Concentration	56.595	0.725	10.996	0.476
Default Mean	57.751	-0.325	6.887	-0.740
Default Dispersion	58.498	-0.309	6.040	-0.493
Default Concentration	32.021	0.033	21.058	0.278
Book Leverage Mean	6.036	0.701	0.679	0.197
Book Leverage Dispersion	6.855	0.043	1.817	0.097
Book Leverage Concentration	20.157	-0.641	16.937	0.043
Bank Entry Mass		0.700	0.195	0.810
Number of Banks			0.102	0.811

Notes: Table reports standard deviations and correlations with output of key financial aggregates. Columns (2-3) report moments in the data. Columns (4-5) report moments from the model. To obtain model-based moments we solve and simulate the model for 10,000 periods. All simulations are conditional on capital quality shocks only.

relation coefficient while being subjected to normal shocks as in the benchmark. The differential between the benchmark and the counterfactual simulations identifies the impact of the aggregate shock on each variable relative to its stochastic steady state value.

Figure 9 displays the responses of key macroeconomic aggregates. The crisis episode is associated with a contraction in output, consumption, and aggregate capital. The risk premium, defined in the model as $R^T(\mathbf{S}) - \bar{R}$, spikes upwards due to the deterioration of conditions in the financial sector: the value of productive capital, and by extension, bank assets has declined. This effect gets amplified through the tightening of bank leverage constraints and a further fall in asset values. The impact on the real economy runs directly through firms' reliance on bank financing and the collapse of non-financial investment. There is no aggregate uncertainty in the total factor productivity (TFP) in the model, and so the actual A_t is unaffected. Measured TFP ($\tilde{A}_t$), however, defined as the simple Solow residual of the production function, declines in line with the rest of the

Figure 9: **Financial Recession - Macro Outcomes**



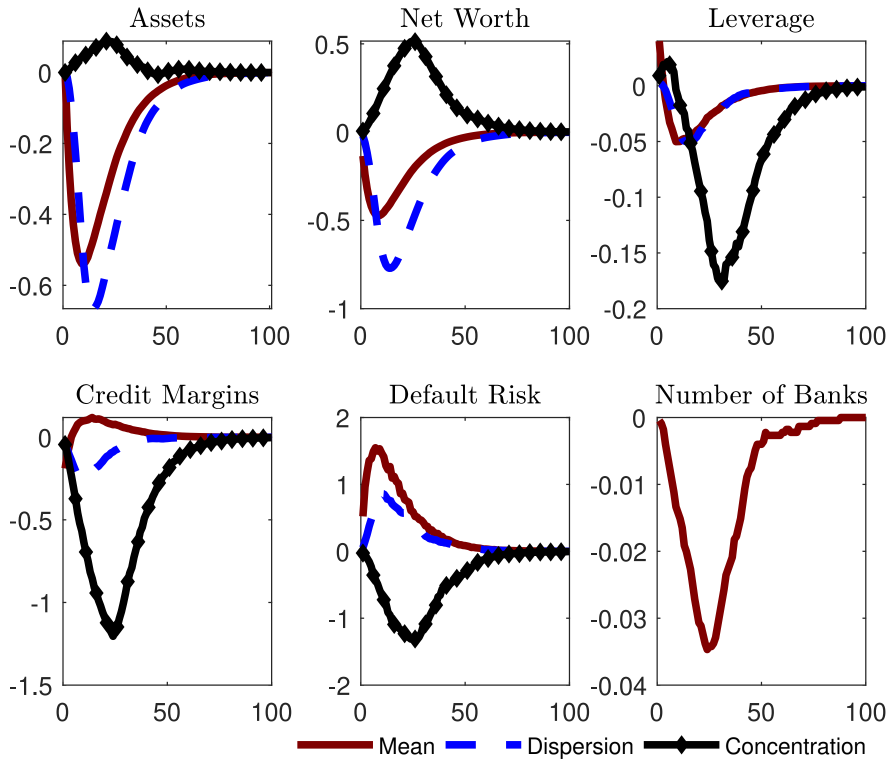
Notes: Response to a one standard deviation negative ψ_t shock. The benchmark economy is in the stochastic steady state. All figures plot percentage point differences. See main text for details.

real aggregates[7].

Figure 10 presents the response of the financial sector. For each financial characteristic, we plot responses of the mean, dispersion, and concentration of the corresponding cross-sectional distribution. First, the mean and dispersion of bank assets and net worth fall, while concentration levels of both rise. The negative aggregate shock shifts the distribution of bank size *leftward*. Concentration rises due to the uneven distributional effects in the financial sector. On one side, bank exit due to default accelerates and bank size gets clustered around zero due to limited liability. On the other hand, institutions with ex-ante high net worth and returns are affected marginally less and become larger in relative terms than prior to the shock. Effectively, a form of *reallocation of credit provision* takes place that favors banks with ex-ante strong balance sheets and high

[7]Formally, we define $\tilde{A}_t$ as the difference between the time-series of output and capital. However, Y_t is the actual, realized series while capital is from the stochastic steady state: $\tilde{A}_t = Y_t - K_t^{ss}$. That is, in response to unexpected ψ_t shocks, the econometrician can always measure and observe that Y_t is falling but cannot correctly attribute it to the "true" capital stock that is subject to financial frictions and imperfect competition. With measured capital not showing any response to the shocks, the econometrician attributes the unexplained component of the drop in Y_t to the measure of our ignorance.

Figure 10: **Financial Recession - Banking Outcomes**



Notes: Response to a one standard deviation negative ψ_t shock. The benchmark economy is in the stochastic steady state. All figures plot percentage point differences. See main text for details.

profitability. Additionally, because startup equity of new entrants is tied up to the average level of net worth, all entrants begin with a lower level of net worth, which further boosts the size concentration.

Mean leverage (in book values) falls, and so do its dispersion and skewness. In the model book leverage is positively correlated with net worth. As the distribution of net worth shifts to the left, smaller intermediaries become less risky in terms of book leverage. New entrants begin with lower levels of initial equity and, conditional on operation, immediately choose very low levels of book leverage. This gets reflected in a reduced skewness of leverage.

Furthermore, average credit margins rise, while their dispersion and skewness both fall. As seen from our model policy functions, credit margins decline with bank net worth. As the average intermediary becomes smaller, the average loan margin increases. However, the presence of a small fraction of very large intermediaries is reflected in the decline of the skewness of margins, in line with the data showing that the concentration of credit margins is pro-cyclical. A similar response pattern is observed for default risk: smaller institutions are considerably more likely to be driven to insolvency due to negative realizations of idiosyncratic return shocks. Large intermediaries with a sufficient buffer stock of net worth are immune to this risk, which makes the default risk distribution left-skewed.

Finally, the total number of banks (financial varieties) falls. This is the result of two drivers of the extensive margin. First, bank entry stalls because the cutoff level of initial profitability has increased. This is due to the fact that startup equity injections, which are tied to aggregate net worth, are now lower. Second, the bank exit rate, driven by endogenous default, rises. As a result, the number of active incumbent intermediaries shrinks.

Overall, and conditional on capital quality shocks only, the model is capable of reproducing the cyclical properties of all higher-order moments except for two (skewness of leverage and dispersion of margins).

5 Inspecting the Model Mechanisms

In this section we build on the "financial recession" experiment of the previous section and isolate the contributing roles of the three key model mechanisms: (i) bank market power, (ii) idiosyncratic risk, and (iii) endogenous entry.

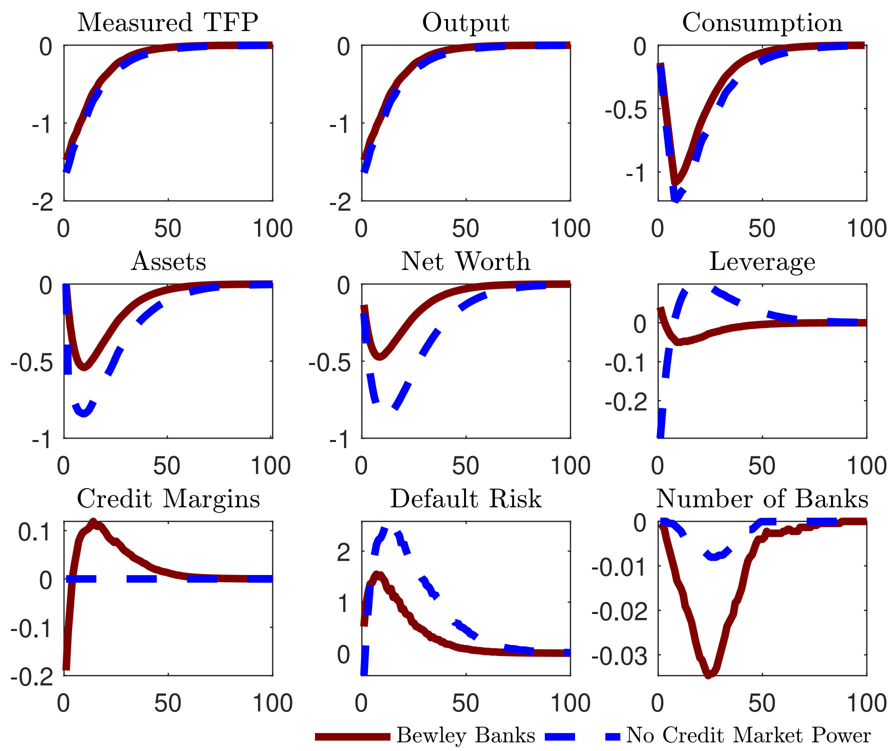
5.1 Shutting Down Bank Market Power

In order to analyze the impact of bank market power on business cycle fluctuations, we shut down the credit margins channel. Specifically, we set θ to a very large number. This turns off the bank's ability to charge market-specific loan margins over the cost of funds. However, because the idiosyncratic risk channel and scale-dependency are still active, the distribution of banks still plays a role. Numerically, we run the same exercise as with the baseline crisis experiment but separately for the two economies.

Figure 11 plots the response functions to a negative capital quality shock. We observe that the presence of credit market power *dampens* the responsiveness of both output and consumption. In cumulative terms (not shown on the figure), in response to the same shock, the economy with no credit market power suffers a 10% greater decline in output by the 40th period. This result arises from the fact that variable loan margins are an additional margin of adjustment for banks to “insure” against adverse shocks *ex post*. In response to negative aggregate shocks, credit margins rise as the distribution of net worth shifts leftward. The aggregate price of capital *increases*, which acts as an endogenous stabilization mechanism. The market value of bank assets falls by less, and the real economy goes through a more benign contraction.

Note that in the economy with no credit market power, bank book leverage falls on impact albeit recovering and growing over time. Leverage falls initially because book assets fall by more than

Figure 11: **The Role of Credit Market Power**



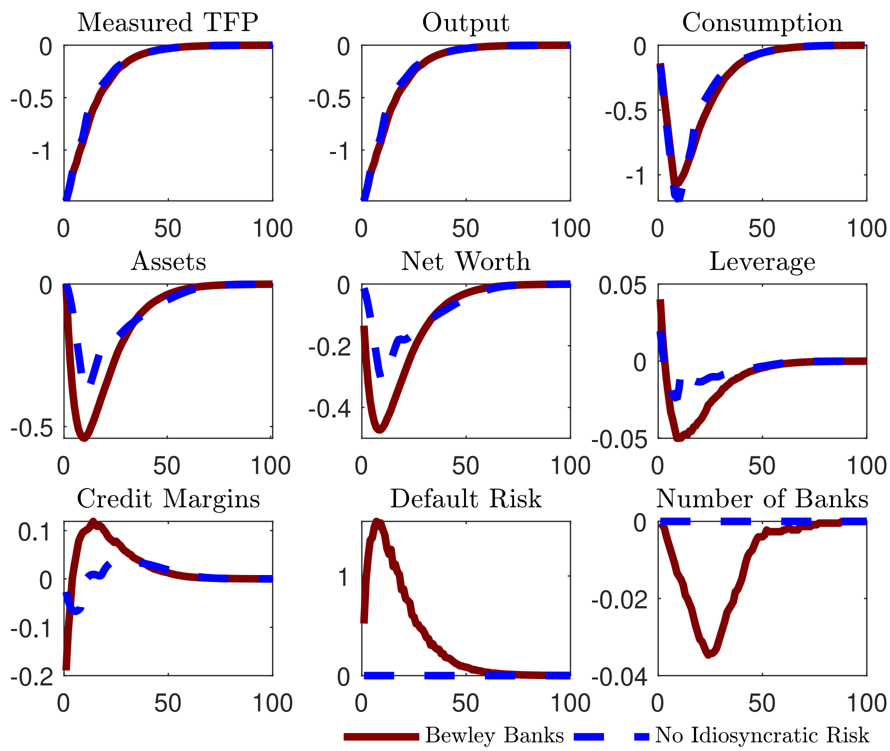
Notes: Response to a one standard deviation negative ψ_t shock. The "Bewley Banks" economy is the baseline. Dashed blue lines represent the economy where θ is set to 10^{10}. All figures plot percentage point differences.

net worth when credit market power is low. Default risk is also considerably higher in the perfect competition economy. This result resembles the often discussed competition-stability trade-off, but applied to the case of business cycle analysis. Finally, the number of banks declines by more in the baseline Bewley economy. This observation comes as a result of the impact of market power on franchise values. The average bank franchise value is lower when the margins are high. This implies that the participation threshold is higher in the baseline economy with monopolistic banks than in the one with perfectly competitive banks, especially in crisis episodes.

5.2 Shutting Down Idiosyncratic Risk

Next we proceed by shutting down the idiosyncratic risk channel. Specifically, we set σ_ξ equal to 0. This has two immediate effects in our economy. First, the economy essentially reduces to the case of a representative financial variety as there is no ex-post heterogeneity in returns. Second, the extensive margin is shut down as well, as a result. Note that in this no-risk economy the bank can still charge credit margins that are market-specific. The details of the numerical experiment

Figure 12: **The Role of Idiosyncratic Risk**



Notes: Response to a one standard deviation negative ψ_t shock. The "Bewley Banks" economy is our baseline. Dashed blue lines represent the economy where σ_ξ is set to 0. All figures plot percentage point differences.

are the same as before.

Figure 12 plots the responses for the two economies, with and without idiosyncratic risk. Idiosyncratic risk acts as a source of *amplification* in the financial sector with limited effects on output and consumption. This is in contrast to the strong dampening impact of credit market power. The cumulative effect on consumption is around 5-10% greater (in absolute and cumulative terms) in the Bewley economy relative to the no-risk counterfactual. In the Bewley economy, when markets are incomplete, bank assets and net worth decline by more in reaction to aggregate shocks. This effect arises due to ex-post heterogeneity in returns and net worth that the idiosyncratic risk and decreasing returns channels grant us: the distribution of banks in the stochastic steady state features a positive mass of fragile intermediaries - those with low initial net worth and a history of low $\xi(j)$. In crises episodes, fragile intermediaries experience heavier balance sheet losses, which is the intensive margin effect. In addition, the fraction of those fragile intermediaries goes up, i.e. the extensive margin adjustment reinforces the original shock. As a result, both aggregate net worth and capital decline by far more and take longer to replenish.

Note that in the no-risk economy, book leverage falls on impact albeit by a small amount. This is due to the fact that book leverage and net worth are positively correlated and there is no bank net worth heterogeneity in the no-risk case. By similar logic, credit margins fall in the no-risk economy but rise in the Bewley economy. As the distribution of bank net worth shifts to the left, and because private credit margins decline in size, the average margin increases. In the no-risk economy, the margin of the representative banker declines due to the first-order effect of the aggregate shock. Finally, there is no default risk in the no-risk economy by construction.

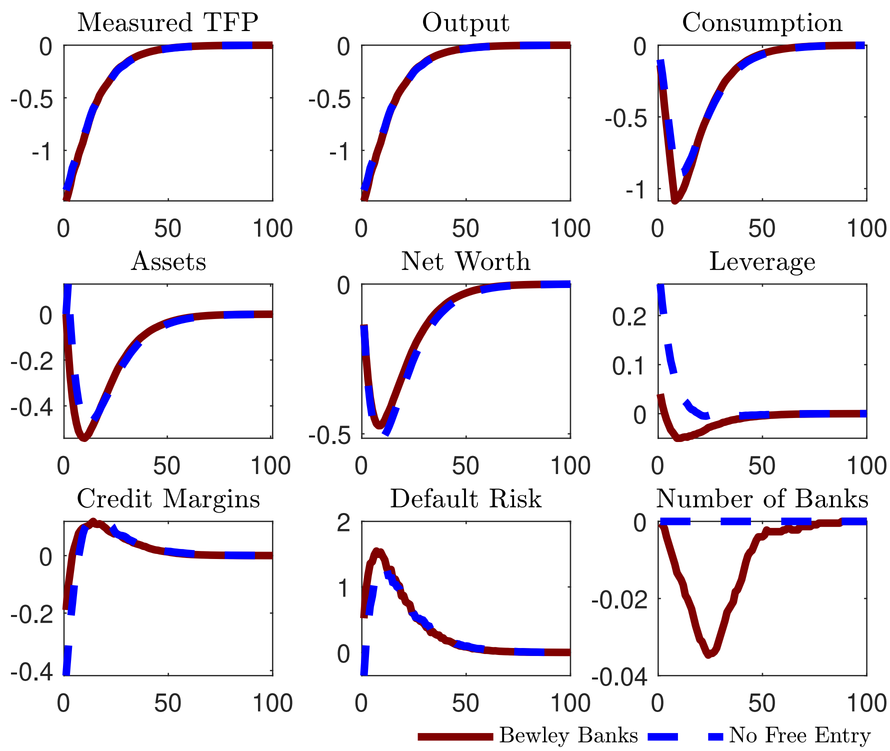
5.3 Shutting Down Endogenous Bank Entry

Finally we explore the implications of shutting down the bank entry channel. In the baseline economy, recall that financiers must solve the dynamic problem in (24) when deciding whether to enter and to operate. We replace this condition with the assumption that the number of banks remains invariant over the business cycle. In other words, the exact mass of varieties that exits due to the σ shock or endogenous default gets replaced regardless of the state of the business cycle such that $\mathrm{J_t}$ is time-invariant. In the Bewley economy, the quantity of capital that new entrants intermediate upon entry is endogenous: it depends on the cutoff level of the idiosyncratic return process, which determines the entry decision of the marginal financier, as well as the startup equity injection. In the no-entry economy, we assume that the mass of entrants intermediates 30% of the aggregate capital stock, which is roughly the ratio that arises endogenously in the Bewley economy. Because entry is now exogenous, the cost of entry becomes a redundant parameter. As a result, this exercise isolates the contribution of the mass of new entrants, and not the capital that they intermediate. Details of the exact numerical experiment are the same as before.

Figure 13 plots the responses for the two economies. Endogenous bank entry acts as a minor source of *amplification*: aggregate consumption falls by about 10-15% more (in cumulative terms) in the Bewley economy than in the no-entry version. With exogenous entry, output and the measured TFP fall by slightly less. In the financial sector, exogenous entry is associated with a dampened effect on bank assets, leverage, and margins, primarily due to changes on impact. In the no-entry economy, the volume of capital contributed by new entrants is time-invariant while it fluctuates together with the mass of new entrants in the Bewley economy. Following an adverse aggregate shock, there is immediately less entry and, by extension, lower capital in the Bewley economy. Over time, the effect vanishes slowly because the banking sector manages to re-grow back the stock of net worth.

The differential effect on capital explains the noticeable differences in the responses of leverage, credit margins, and default risk. With exogenous entry, leverage spikes up on impact because assets initially increase slightly and then fall by less than bank net worth. Credit margins first fall considerably and then rise, with a cumulative growth of about 25% less than in the baseline.

Figure 13: **The Role of Endogenous Entry**



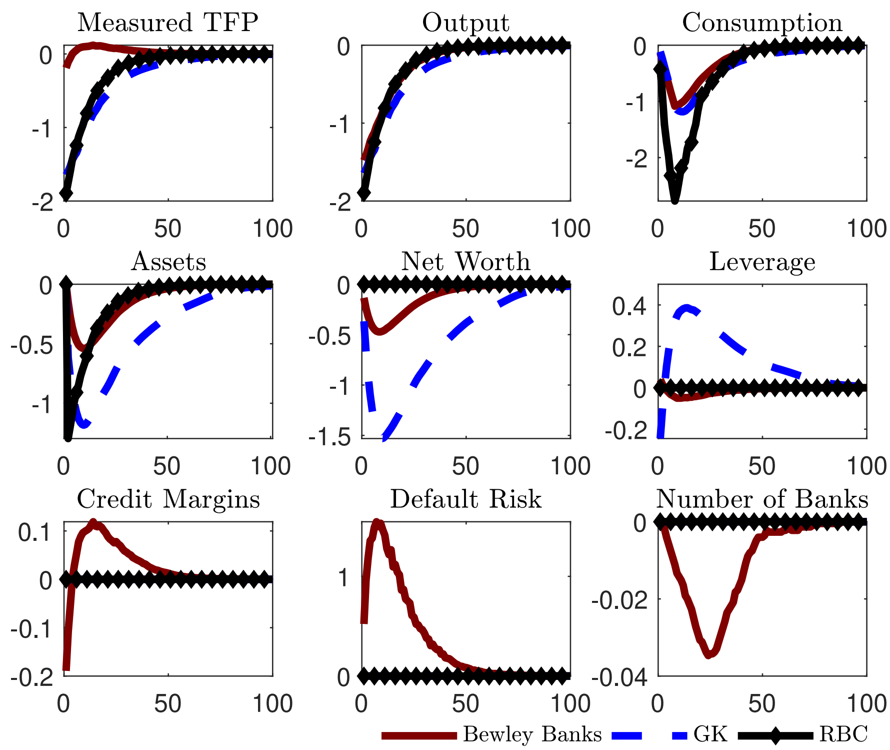
Notes: Response to a one standard deviation negative ψ_t shock. The "Bewley Banks" economy is our baseline. Dashed blue lines represent the economy with exogenous bank entry.

This takes place because private margins $\frac{p_t(j)}{\bar{R}(j)}$ are decreasing in aggregate demand for capital K_t. Collectively, this pushes up the average relative price P_t when capital falls by more. Similar logic applies to the response of default risk: $\nu_t(j)$ increases with ψ_t and declines with K_t. As initial aggregate capital changes by less in the no-entry economy, so does default risk.

6 Main Results and Experiments

In this section we report our main quantitative results. First, we show how our model tractably nests the canonical GK and RBC environments. Recessions in our environment can be either dampened or amplified, depending on the cyclicality of idiosyncratic risk. Second, we demonstrate a specific feature of our environment, i.e., the aggregate state-dependency on the endogenous, dynamic distribution of bank net worth. Third, we use our model to simulate a persistent rise in banking concentration that has been documented for the US. We conclude by identifying and characterizing banking crises episodes in our model economy using an event study approach.

Figure 14: **Financial Recession - Bewley Banks, GK Banks, and RBC**



Notes: Response to a one-standard deviation negative shock to ψ_t. The red straight, blue dashed, and black diamond lines represent, respectively, the baseline model, the GK counterfactual with no idiosyncratic risk and no monopolistic competition in banking, and the RBC counterfactual with no leverage constraints in addition to all the assumptions from GK. All figures plot percentage point differences.

6.1 Nesting GK and RBC Models

We now show how our framework nests the RBC and GK models. The GK environment can be achieved in several simple steps. First, we eliminate credit market power by setting θ to a very large number. The distribution of relative bank-level prices p(j) collapses to unity. Second, we set $\beta = 1$ which brings back scale invariance. Finally, we set $\sigma_\xi = 0$ which shuts down the idiosyncratic risk channel. The resulting financial intermediary sector collapses one-to-one to the representative bank in GK. To go from GK to RBC, we set the leverage constraint parameter λ to a very low number. That is, financial frictions are absent and the leverage constraint is always slack on all points of the idiosyncratic and aggregate state spaces.[8] We then compare the response to a one-standard-deviation aggregate ψ shock in the Bewley Banks framework to the GK and RBC models.

[8] Technically, our "RBC" economy is the true RBC model up to the augmented stochastic discount factor $\bar{\Lambda}$. Recall that the representative investor in our economy is the financial intermediary and not the household.

Figure 14 presents the results. Several observations are worthy of note. First, the GK economy goes through the most contractionary recession due to the financial accelerator embedded into the banking sector. Second, notice that the change in bank leverage is almost an order of magnitude greater in the GK case. Moreover, qualitatively the change is in the opposite direction - leverage falls on impact and then grows over time in the same way as in the full-competition scenario displayed in Figure 11. In contrast, market leverage (not shown) increases in both economies but again by almost an order of magnitude more in GK. Last but not least, macroeconomic responses in the Bewley economy are *dampened* relative to GK. This result emerges for the following two reasons. Recall that credit market power acts as a significant dampener of exogenous aggregate shocks, a fact that we established in Section 5.1. The dampening nature of market power dominates the idiosyncratic risk channel, which amplifies the no-risk counterfactual.

But even in the limiting case with idiosyncratic risk but no market power there is still dampening relative to GK. This occurs because private capital and net worth are both increasing in σ_ξ due to the precautionary lending motive. The only way for the intermediary to hedge idiosyncratic risk is to lend ("save") more. The net effect on market leverage, a sufficient statistic for probability of default, is negative. As a result, an economy with more exogenous environmental risk features less *endogenous* riskiness due to lower leverage ratios and more equity. The inverse relation between risk and leverage is consistent with the conventional wisdom that has been stressed by a number of other papers (Fostel and Geanakoplos, 2008; Gertler et al., 2012). Lower leverage in the stochastic steady state leaves the economy in a less fragile initial condition.

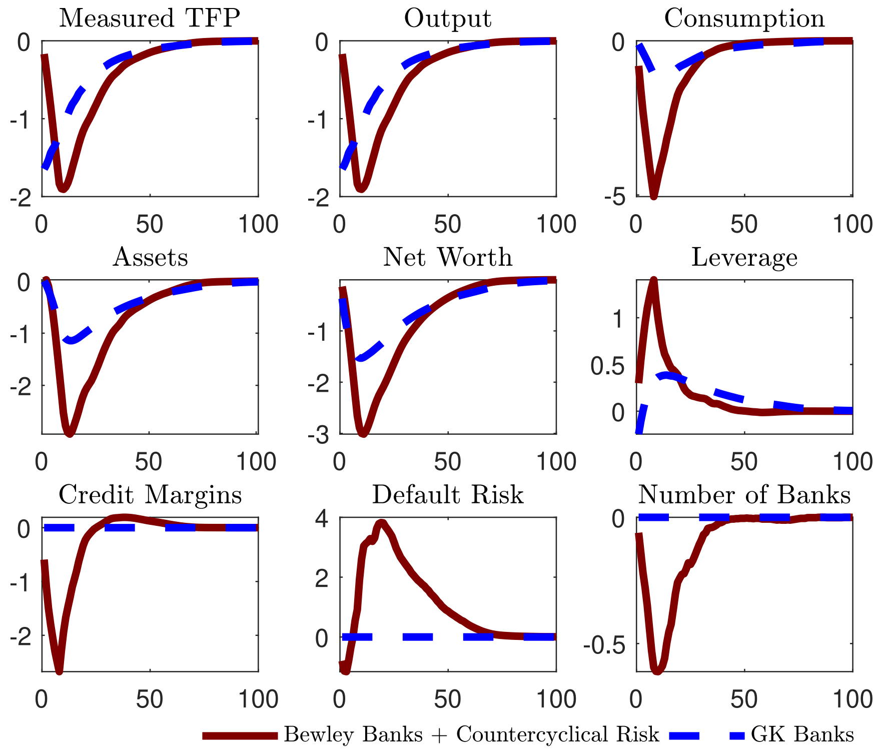
6.2 Counter-cyclical Return Risk and Endogenous Amplification

One way to counteract the power of the precautionary lending motive and to enhance amplification is to introduce a *counter-cyclical* rate of return risk.[9] In our model, microeconomic risk and volatility in the non-financial industry map straightforwardly to $\xi_t(j)$. We now allow idiosyncratic risk to be state-dependent. Specifically, we assume that $\mu_\xi(\mathbf{S})$ - the unconditional mean of idiosyncratic rate of return risk in Equation 10 - is now state-dependent and deterministic: it falls by 1 percentage point in the low aggregate state. As a result, when the aggregate state of nature is low, the bank faces a higher probability of experiencing a bad idiosyncratic return draw. This puts additional downward pressure on bank balance sheets precisely when the marginal value of net worth is the highest.[10]

[9] Bloom et al. (2018) have shown that microeconomic volatility rises sharply in recessions. Among many others, Challe and Ragot (2015) and Sterk and Ravn (2020) develop quantitative frameworks with counter-cyclical income and unemployment risk.

[10] Making σ_ξ counter-cyclical would further strengthen the outcome. However, as also noted in Bloom et al. (2018), we would still require a negative shock to the mean in addition to the second-moment disturbance.

Figure 15: **Bewley Banks with Counter-cyclical Return Risk σ_ξ**



Notes: Response to a one standard deviation negative ψ_t shock. The "Bewley Banks" economy is our baseline. Dashed blue lines represent the economy where $\mu_\xi(\mathbf{S})$=1 when ψ is high and $\mu_\xi(\mathbf{S})$=0.99 when ψ is low. All figures plot percentage point differences.

The results of this experiment are portrayed in Figure 15. Notice that, on impact, the recession is milder than in the GK benchmark. In fact, it's quantitatively similar to our baseline. However, once in the negative aggregate state, the counter-cyclical idiosyncratic risk kicks in. A higher fraction of local credit markets starts to draw low $\xi(j)$. This pushes leftward the distribution of bank assets and net worth. Over time, the original aggregate shock gets amplified considerably. The measured TFP, output, and consumption all fall by more than 30% relative to GK in cumulative terms by quarter 20.

The amplification effect in the financial sector is considerable. First, the quantitative effect on the number of active intermediaries is an order of magnitude greater than in the baseline Bewley Banks economy. This result follows immediately from the greater relative fall in both assets and net worth. Second, credit margins now fall both on impact and along the transition back to the original stochastic steady state. This occurs because private margins $\frac{p(j)}{\bar{R}(j)}$ are decreasing in $\mu_\xi(\mathbf{S})$. As a result of the combined effects from book assets and equity, bank leverage also rises on impact considerably.

Overall, we find that in the Bewley Banks framework, recessions could either be dampened or amplified relative to a perfectly competitive benchmark with no idiosyncratic risk. The key determinant is whether the precautionary lending motive can dominate the direct effect of business cycles on financial balance sheets. If the idiosyncratic return risk is counter-cyclical, the business cycle impact on balance sheets is too powerful for the intermediary to insure against ex-ante or ex-post. However, if idiosyncratic risk is not state-dependent, the precautionary lending motive guarantees that recessions are dampened relative to GK as the bank accumulates enough equity capital to withstand aggregate uncertainty.

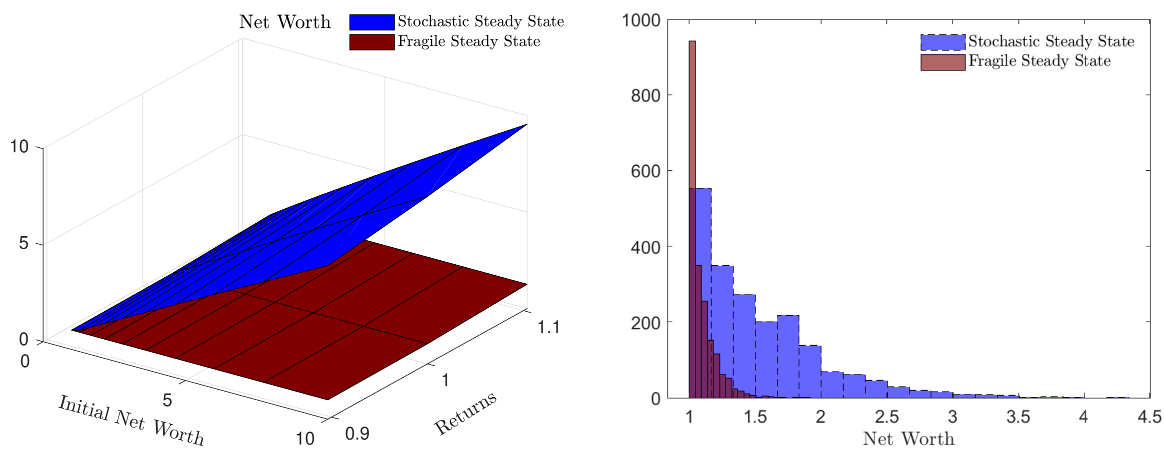
6.3 Fragile Bank Distributions and the Business Cycle

We now highlight a genuine feature of the Bewley Banks environment: namely that aggregate responses to exogenous shocks depend explicitly on the dynamic, endogenous distribution of bank net worth. Our exercise consists of the following steps. First, we solve our model with aggregate uncertainty. For simplicity, the only exogenous aggregate disturbance is the capital quality shock ψ_t. Second, we simulate the model for T=10,000 periods four times. The first simulation is our "benchmark" case - the stochastic steady state. In the second simulation, the model economy is hit with a counterfactually low ψ_t after T* periods. The negative shock is allowed to revert back to the benchmark series with its normal autocorrelation coefficient while being subjected to normal shocks as in the benchmark. The difference between the second and the first simulation is our response function to the negative ψ_t shock. This is also our baseline recession experiment in Section 4.5.

In the third simulation, after T*-1 periods we allow for a transitory exogenous change in the conditional distribution of net worth $n'(j)$. We assume that the economy temporarily moves to a "fragile" distributional state in which average net worth falls while dispersion and skewness increase[11]. We construct a new conditional distribution that is determined by a counterfactual policy function $\hat{n}'(\mathbf{s};\mathbf{S})$ equal to future net worth conditional on the *lowest* value of the initial net worth state. This shift is best visualized in the left panel of Figure 16. The blue surface is the equilibrium two-dimensional policy function which we reported in Figure 5. The red surface is the policy function consistent with the fragile state. The right panel of Figure 16 plots the conditional cross-sectional distributions of net worth from the equilibrium and the fragile states. We assume that this transitory shift lasts for 8 quarters. That is, for 8 quarters the model is being simulated with the counterfactual policy function that generates the fragile distribution. The duration of the shock is consistent with the average duration of banking crises in the data Laeven and Valencia (2012)

[11]In Section 6.4 we explore this point from a different angle: by studying model responses to direct, exogenous shocks to higher-order moments which in turn feed into the decline of the first moment endogenously.

Figure 16: **The Fragile Steady State**



Notes: the left panel shows the baseline and the perturbed two-dimensional policy functions for net worth growth $\hat{n}'(\mathbf{s};\mathbf{S})$. The right panel plots histograms of the cross-sectional distributions of net worth under the two alternative policy functions. The mean and skewness of the distribution in the stochastic steady state are 1.5151 and 1.4496, respectively. In the fragile steady state, the mean and skewness are 1.0878 and 2.2315, respectively.

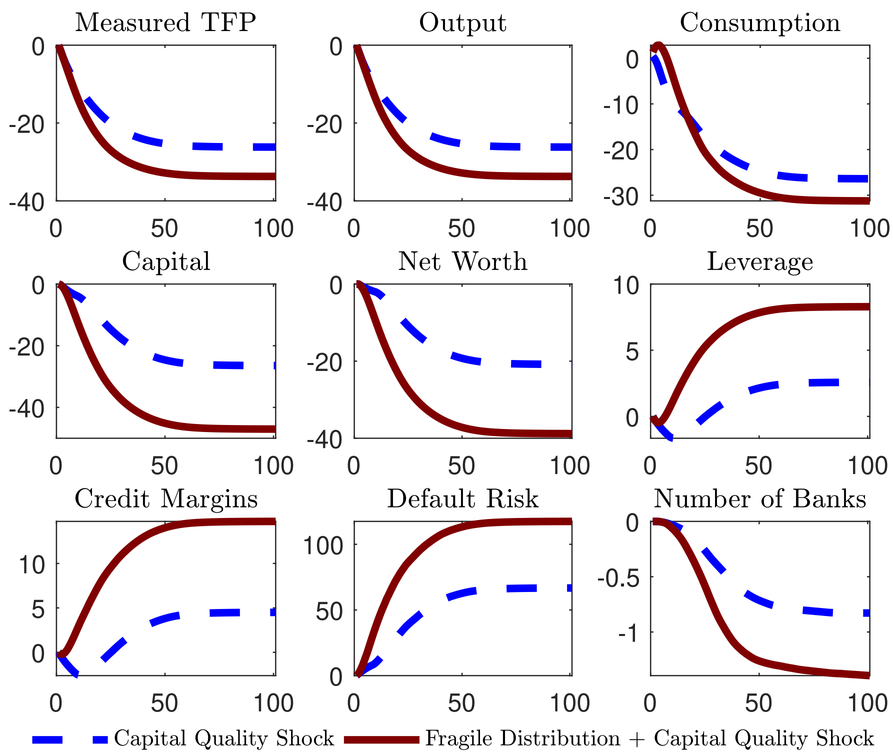
but is otherwise not materially important. After 8 quarters we revert back to the policy function consistent with the stochastic steady state.

Finally, in the fourth simulation we have the exact same distributional shock as in simulation 3. In addition, in period T* the economy is now hit with the same negative exogenous ψ_t shock as in simulation 2. The difference between the fourth and the third simulations is the model response conditional on the initial banking distribution being fragile. Finally, we compare the percentage differential between simulations 1 and 2 with the percentage differential between simulations 3 and 4. This identifies exactly the aggregate state-dependency of the economy with respect to fluctuations in the bank net worth distribution.

Figure 17 reports the results of this exercise. We plot cumulative impulse response functions for visibility. The dashed blue line is the model response to a negative capital quality shock. The red straight line is the model response to the same shock but conditional on the fragile initial bank distribution. We see from the figure that even a short, transitory negative change in the distribution of bank net worth has a permanent and considerable effect on the macroeconomy. First, the cumulative response of aggregate production is lower by 15% in the case of the fragile distribution. The response of aggregate consumption is lower by roughly 10%. Second, the distributional change has a relative contractionary effect on the size of the financial sector. The negative response of bank assets and net worth is greater in absolute terms by a factor of 2.

Third, the financial sector is more *risky* as the market leverage ratio increases by more than a factor of roughly 2.5. The probability of bank default is higher by a similar order of magnitude.

Figure 17: **Aggregate State Dependency on the Distribution of Bank Net Worth**

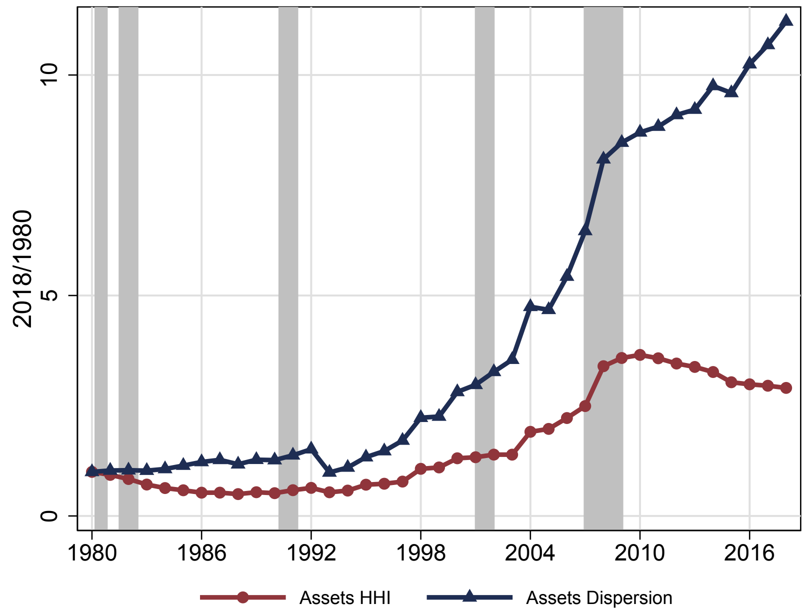


Notes: Impulse responses to a one-standard deviation negative ψ_t shock with and without a prior transitory negative shock to the conditional cross-sectional distribution of bank net worth n(j). The distributional shock lasts for 8 quarters and is depicted in Figure 16. The ψ_t shock reverts back to the stochastic steady state with the normal autocorrelation of 0.914. See main text for more details. All figures plot percentage point differences. All response functions are cumulative.

Fourth, the fragile distribution contributes to higher loan margins in the recession. Average margins increase by 10 percentage points more than in the baseline economy. Finally, the number of banks is twice as low, as potential entrants refrain from entry while internalizing low startup equity injections. All of these results arise from the fact that the initial level of equity capital is low - since bank net worth is the key state variable in the model, all endogenous responses such as margins and leverage react accordingly.

Overall, the above exercise highlights the powerful amplification mechanism that is behind the dynamic cross-section of bank net worth. When the initial distribution of net worth is fragile, aggregate responsiveness to negative exogenous shocks is considerably stronger.

Figure 18: **The Rise of Banking Concentration and Dispersion**



Notes: Time-varying dispersion and Herfindahl index of the cross-section of financial intermediary assets. Data is from Compustat. Sample includes US commercial banks only.

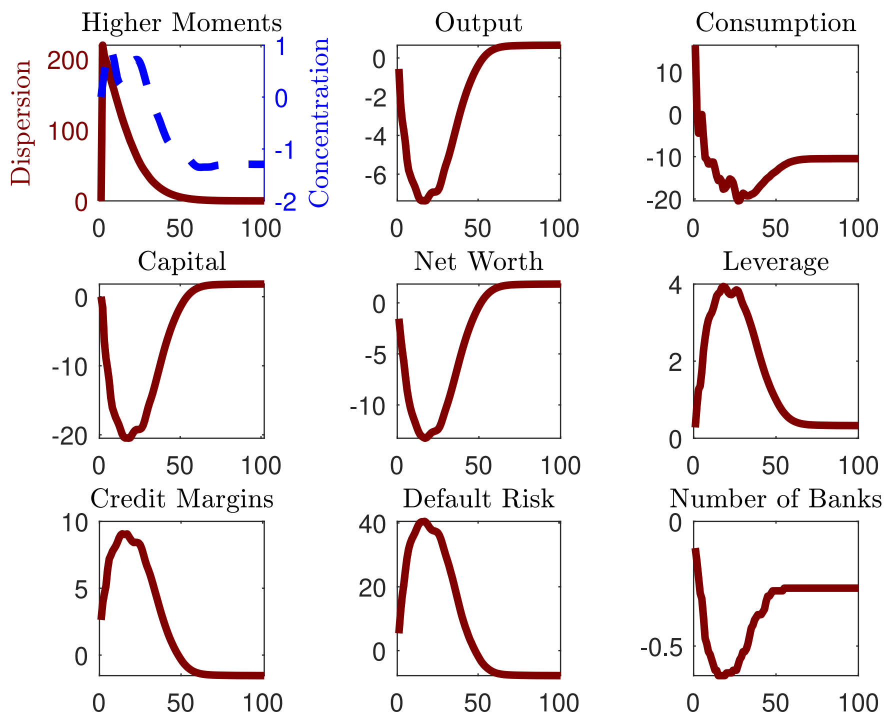
6.4 The Rise of Banking Concentration and Dispersion

Recent papers by Jamilov (2020) and Corbae and D'Erasmo (2019) have documented a considerable rise in *concentration* in the U.S. banking sector. There is a large literature in corporate finance and banking that links the rise of banking concentration to a plausibly exogenous sequence of state and federal legislations that relaxed restrictions on bank entry and geographical expansion over 1960s-1980s. (Jayaratne and Strahan, 1996; Kroszner and Strahan, 2014). Figure 18 plots the time-series of commercial bank assets concentration (measured by the HHI) and dispersion. The two measures are highly correlated and have risen by a factor of 4 and 10, respectively, since the 1980s.

In our quantitative exercise below, we explore exogenous, transitory but persistent, shocks to the higher-order moments of the bank credit distribution as a *source* of business cycle fluctuations. We operationalise this idea in the following way. Recall that according to our numerical algorithm, we track $\mathrm{m_k}$ moments of the k(j) distribution. In the baseline scenario, we only keep track of the mean. In this section, we also track the time-varying dispersion of assets, $\sigma_{\mathrm{t}}^{\mathrm{k}}$. In our numerical experiment, as we show below, concentration will rise endogenously on impact in response to the exogenous dispersion shock.

Computationally, tracking the second moment makes $\sigma_{\mathrm{t}}^{\mathrm{k}}$ a relevant *state variable*. Exogenous

Figure 19: **Response to an Exogenous Second-Moment Shock**



Notes: Response to an 8-fold positive shock to the time-varying dispersion of the cross-sectional distribution of bank assets k(j). Dispersion reverts back to the stochastic steady state with autocorrelation of 0.9. See main text for more details. Responses are cumulative. All figures plot percentage point differences.

shocks to σ_t^k will have a direct first-degree effect on the optimal responses of all agents. If accuracy of the baseline solution could be improved with the introduction of the second moment, then the mean of k(j) is not a sufficient statistic for the characterization of the dynamic cross-section. This line of reasoning is conceptually very similar to the original ideas in Krusell and Smith (1998) who found that the first moment of the distribution of household wealth was generally sufficient for the description of macroeconomic aggregates.

Our numerical exercise consists of three general steps. First, we solve the model where we allow the law of motion of the distribution Γ to have σ_t^k as an additional argument. Second, we simulate the model for 10,000 periods twice. In the first simulation, the economy remains in the stochastic steady state. In the second simulation, the economy is hit with a positive, persistent shock to σ_t^k in period T*. We assume that σ_t^k rises by a factor of 8, which is in line with empirical evidence in Figure 19. Finally, a cumulative impulse response graph plots the percentage differential of the two series. If the second moment is redundant, the response functions should be flat.

Equilibrium of the model in which σ_t^k is a state variable is characterized by the following log-linear solution for Γ:

$$\begin{aligned}\log(K') &= 0.4465 + 0.8160 \log(K) - 0.0010 \log\Big(\sigma(K)\Big); \quad \psi \text{ low} \\ \log(K') &= 0.3769 + 0.8526 \log(K) - 0.0060 \log\Big(\sigma(K)\Big); \quad \psi \text{ high}\end{aligned}$$

For projections of K'. For projections of P' we have:

$$\begin{aligned}\log(P') &= 1.3871 - 0.4406 \log(K) + 0.0562 \log\Big(\sigma(K)\Big); \quad \psi \text{ low} \\ \log(P') &= 1.4764 - 0.4907 \log(K) + 0.0748 \log\Big(\sigma(K)\Big); \quad \psi \text{ high}\end{aligned}$$

From the solution above we immediately notice that shocks to σ_t^k are contractionary - they cause quantities to fall and credit margins to rise.

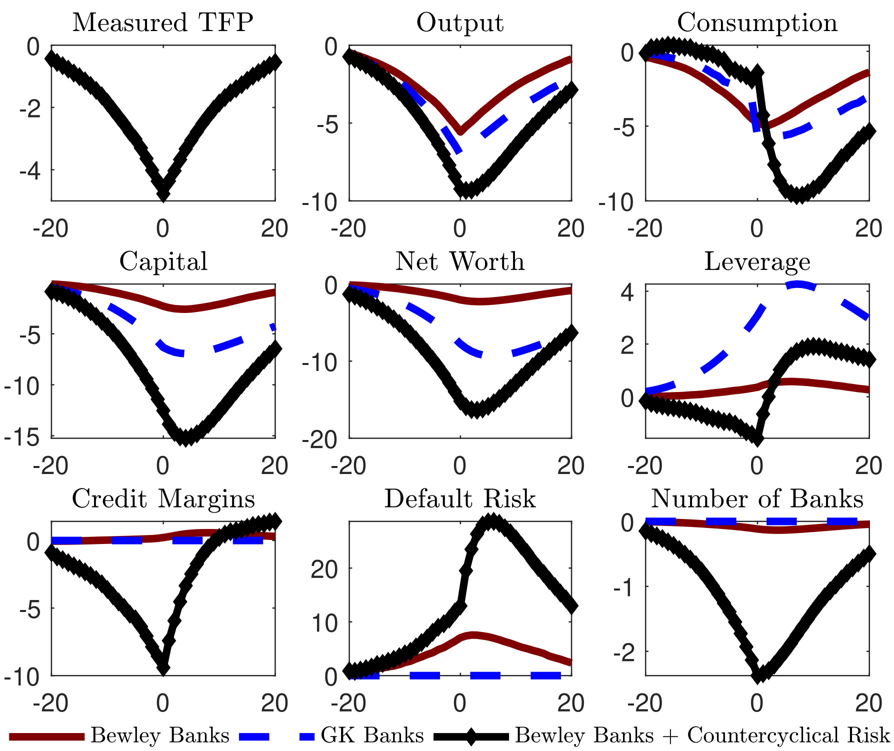
Figure 19 plots the impulse response functions. We show strong evidence that persistent shocks to higher-order moments of the banking distribution have a large impact on business cycle fluctuations. First, positive dispersion shocks cause considerable economic recessions: severe cumulative declines in aggregate output, consumption, and the measured TFP. Second, the financial sector goes through a financial crisis as bank assets, net worth, and the number of active institutions all go down significantly. Third, the financial industry accumulates more leverage when dispersion is high, i.e. the economy remains riskier for longer. Fourth, loan margins increase by a factor of 10 in cumulative terms. Overall, we can conclude that the rise of U.S. banking dispersion and concentration over 1980-2020 may have contributed to a more sluggish growth with fewer, smaller and riskier financial intermediaries that charge higher loan margins and default more often.

6.5 Banking and Economic Crises

In this section we use our model to identify and characterize systemic banking and economic crises. We employ an event study approach. Our methods follow the open-economy macroeconomics literature (Mendoza, 2010). First, we solve the model and simulate it for 10,000 periods with ψ_t as the only exogenous aggregate disturbance. Second, we define economic crises as episodes (quarters) with low measured TFP, $\tilde{A}_t$. Specifically, $\tilde{A}_t$ must be in the lower 20% of the whole simulation. This approach enables a fair comparison across different classes of models because $\tilde{A}_t$ can be readily constructed in both GK and Bewley Banks frameworks. If the bank default risk channel is active, our definition parsimoniously captures episodes of joint financial and economic distress because default risk in our framework is countercyclical.[12] Third, we store every crisis episode and look at the 20-quarter window before and after the event. For each quarter in the window,

[12]The literature on banking crises has documented the substantial negative impact of bank defaults on the real economy and consumer welfare, which in our exercise is captured in a general way with low measured TFP (Laeven and Valencia, 2012).

Figure 20: **Banking and Economic Crises - Event Study Approach**



Notes: Event study analysis of an economic crisis that is defined as a quarter with measured TFP in the bottom quintile of the 10,000 period-long simulation. Straight red, dashed blue, and black diamond lines represent, respectively, the baseline model, the GK counterfactual with no idiosyncratic risk and with perfect banking competition, and the Bewley economy with counter-cyclical idiosyncratic risk. All figures plot percentage point differences.

we calculate the unweighted average of key macroeconomic and financial aggregates. Finally, we perform the same exercise for the GK economy and for the Bewley economy with counter-cyclical idiosyncratic risk.

Figure 20 reports the results. In every picture, the red straight line corresponds to the Bewley economy, the blue dashed line corresponds to GK, and black diamond markers to the Bewley economy with counter-cyclical idiosyncratic risk. Note that the decline in $\tilde{A}_t$ across the three economies is equalized by construction. We first observe that the two baseline economies - Bewley and GK - go through the same crisis in very different ways. At the peak of the crisis, the Bewley Banks economy features a smaller contraction in output, consumption, bank assets, bank net worth. The decline in the number of active intermediaries is also relatively muted. In the Bewley economy, crises are also characterized by a build-up of loan margins in the pre-event phase. Meanwhile, at the peak of the crisis the GK economy is more risky as evidenced by higher levels of aggregate default risk and a higher bank leverage ratio. This is a variant of the canonical financial competition-stability trade-off. This observation is also consistent with our impulse-response experiment in

Figure 14.

Now, consider the behavior of the Bewley economy with counter-cyclical idiosyncratic return risk. Economic crises are associated with considerably more severe contractions in output and consumption. The deterioration in the financial sector activity is far more severe relative to the GK counterfactual. Bank assets and net worth fall by 5-10 percentage points more. Furthermore, bank default risk and leverage are significantly amplified. That is, economic recessions in the Bewley economy with counter-cyclical risk are far more likely to occur jointly with systemic banking crises and episodes of financial instability and fragility. Finally, the number of active intermediaries falls by an order of magnitude more than in the baseline Bewley economy.

It is important to highlight that our crisis episodes are still characterized by declines in aggregate risk, in all three economies. That is, our result points to the *amplification* of contractionary shocks, rather than on the endogenous build-up of risk during booms. A key reason for this distinction is the structure of our model that yields *counter-cyclical* market leverage due to the assumed specification of the equity-based leverage constraint.

7 Alternative Aggregate Shocks

So far the sole source of aggregate uncertainty in our baseline economy has been a capital quality shock ψ_t. In this section we investigate the extent to which alternative aggregate shocks can match the business cycle statistics reported in Table 1. We explore six potential "candidate" shocks. Shock by shock, we re-solve our model under the assumption that it is the only source of aggregate uncertainty in the environment. We then simulate the model economy for 10,000 quarters and report time-series correlations between Y_t and our variables and moments of interest.

The six shocks that we consider are the following. First, the baseline shock to the quality of aggregate capital ψ_t. Following Merton (1973), this shock captures fluctuations in the value of capital - its sudden obsolescence or valuation. Second, we consider a shock to Hicks-neutral total-factor productivity A_t (Kydland and Prescott, 1982). This is a standard exogenous stochastic component in the Real Business Cycle literature. Third, a shock to the banks' dividend payout ratio σ_t. This shock essentially captures changes in consumer "preferences" as was originally proposed in Krusell and Smith (1998). Our approach is more parsimonious because we do not consider shocks to β, as the authors did, for simplicity: σ_t only has a direct effect on the augmented SDF of the intermediary and an indirect effect on household behavior through general equilibrium channels.

The fourth shock we consider is a disturbance in the leverage constraint parameter λ_t. In the literature, this is sometimes labeled as a "financial shock" as in Jermann and Quadrini (2013) and Khan and Thomas (2013). The shock captures sudden changes in the ability of banks to borrow and

Table 4: **Business Cycle Statistics - Alternative Shocks**

Correlation with GDP of	Data	Capital Quality (ψ_t)	TFP (A_t)	Leverage Constraint (λ_t)	Credit Margins (θ_t)	Dividend Payout (σ_t)	Market Incompleteness (κ_t)
Assets Mean	0.498	0.798	1.000	1.000	1.000	1.000	1.000
Assets Dispersion	0.642	0.541	0.974	0.944	0.604	0.661	0.476
Assets Concentration	-0.568	-0.103	-0.040	-0.191	0.104	0.132	0.101
Net Worth Mean	0.211	0.842	0.958	0.547	0.943	0.727	0.931
Net Worth Dispersion	0.544	0.683	0.907	0.536	0.701	0.620	0.449
Net Worth Concentration	-0.472	-0.238	-0.874	-0.470	0.028	0.011	0.009
Margins Mean	-0.563	-0.305	0.035	-0.291	-0.813	-0.809	-0.347
Margins Dispersion	-0.370	0.437	0.565	-0.100	-0.019	0.185	0.010
Margins Concentration	0.725	0.476	0.735	0.614	0.023	0.002	0.065
Default Mean	-0.325	-0.740	-0.592	-0.029	-0.060	0.165	-0.036
Default Dispersion	-0.309	-0.493	-0.031	-0.003	-0.059	0.133	-0.036
Default Concentration	0.033	0.278	0.336	0.377	0.012	-0.039	0.027
Book Leverage Mean	0.701	0.197	-0.357	0.277	0.804	0.621	0.612
Book Leverage Dispersion	0.043	0.097	-0.347	0.142	0.223	0.424	0.209
Book Leverage Concentration	-0.641	0.043	0.639	0.340	0.095	0.219	0.121
Bank Entry Mass	0.700	0.810	0.717	0.119	0.841	0.665	0.858
Number of Banks		0.811	0.726	0.108	0.839	0.669	0.840

Notes: Data- and model-implied correlations with output Y_t. Each column reports correlations based on an economy with a single source of aggregate uncertainty reported in row 1. In each case, the model is solved and simulated for 10,000 quarters. Correlation between bank entry and output is taken from Corbae and D'Erasmo (2019). See main text for more details.

build up leverage. The fifth shock we consider is the "credit markup shock" θ_t in the spirit of Clarida et al. (2002) or Ball et al. (2005). This shock can capture sudden changes in the degree of financial market competition and concentration. Finally, a shock to the degree of insurability of idiosyncratic shocks κ_t. This shock best resembles the "incompleteness shock" of Davila and Philippon (2019) and captures sudden disruptions in financial market trading, for example a liquidity dry-out.

Table 4 presents correlations with output of the first three moments of model-implied time-varying distributions of bank assets $k_t(j)$, net worth $n_t(j)$, leverage $\phi_t(j)$, loan margins $\chi_t(j)$, and insolvency risk $\nu_t(j)$. We also include the mass of entering varieties M_t and the total number of banks J_t.

Consider the performance of the baseline (capital quality) ψ_t shock. As discussed before, this shock nails down all but two moments: dispersion of margins and concentration of leverage. Shocks to TFP do rather poorly. Specifically, they predict counter-cyclical book leverage mean and dispersion. They also miss the cyclicality of credit margins. Shocks to the elasticity of substitution θ_t ("credit markup shocks") and the dividend payout σ_t generate similar responses: both cannot match the counter-cyclicality of the concentration of assets. In addition, shocks to θ_t incorrectly predict that the concentration of margins is virtually acyclical. Market incompleteness shocks κ_t fail to generate counter-cyclical concentration of assets and net worth as well as counter-cyclical dispersion of loan margins. The most interesting candidate is the financial shock λ_t. It can match all but one moment - counter-cyclicality of the skewness of leverage, which is in fact the only moment that none of the candidates can replicate. A positive shock to λ_t generates a recession by tightening the limit on bank leverage, which in turn reduces the supply of credit to the economy. λ_t shocks are designed to somehow represent fluctuations in the health of the financial sector. They capture exogenous variations in the degree of moral hazard between the lender and the borrower. Overall the dynamic cross-section of financial intermediaries can be best approximated by shocks to the financial system, proxied in our case by shocks either to banks' capital quality or their leverage constraints.

8 Conclusion

We have developed a new tractable, dynamic stochastic general equilibrium framework with monopolistic competition and uninsurable idiosyncratic return risk in the financial sector. Our setup builds on the canonical macro-banking models of Gertler and Kiyotaki (2010) and Gertler and Karadi (2011) and nests them as special cases. The simultaneous assumptions of local decreasing returns to scale and idiosyncratic return risk break scale invariance. Because the marginal value of net worth and optimal leverage ratios are now both size-dependent, a time-varying distribution of bank characteristics emerges. With aggregate uncertainty, the distribution of bank net worth becomes a *time-varying* endogenous state variable.

Our framework rests upon four quantitative forces. First, the cyclicality of the idiosyncratic risk process determines whether aggregate contractions in the model economy are dampened or amplified. In the baseline scenario with acyclical risk, the precautionary lending motive dominates the first-degree impact from business cycles and dampens the effects of negative aggregate shocks. However, if idiosyncratic risk is counter-cyclical, the direct effect of business cycle fluctuations on bank balance sheets dominates. As a result, economic recessions get substantially amplified. Second, the model features an explicit aggregate state dependency on the dynamic cross-section of bank assets. Varying the initial state of the distribution - either by targeting the conditional distribution of net worth or through direct exogenous shocks to higher-order moments - has implications on business cycle fluctuations and the aggregate sensitivity to exogenous shocks.

Third, individual banks do not internalize the impact of private loan margin-setting choices on aggregate demand. This is the canonical aggregate demand externality of Blanchard and Kiyotaki (1987) applied to the case of financial intermediaries. Finally, the model generates endogenously the financial competition-stability trade-off. In the Bewley banks framework with acyclical idiosyncratic risk, it is generally the case that severe economic recessions are accompanied by relatively mild financial crises. When idiosyncratic risk is counter-cyclical, economic recessions occur jointly with significant deterioration in the financial sector and elevated financial fragility levels.

Our Bewley Banks framework is tractable and portable. It is possible to introduce nominal rigidities into our model, study unconventional credit policies such as bank-level equity injections, or to relax the closed economy assumption[13]. The tractability of our approach rests on the long and vast literature on monopolistic competition with CES aggregators a la Dixit and Stiglitz (1977). Extensions of the model to generate heterogeneous, variable equilibrium markups would achieve a three-dimensional idiosyncratic state which would include bank net worth, returns, and market power. This feature would yield a positive correlation between bank size and loan margins, potentially an interesting and powerful additional channel of transmission. We leave all these interesting and important avenues for future research to explore.

[13]In Jamilov and Monacelli (2020) we study the monetary policy transmission mechanism in a Bewley Banks environment with nominal rigidites.

References

Adrian, T. and H. S. Shin (2010): "Liquidity and Leverage," *Journal of Financial Intermediation*, 19(3), 418–437.

——— (2011): "Financial Intermediaries and Monetary Economics," *Handbook of Monetary Economics, ed. by Ben Friedman and Michael Woodford*, 3A(12).

Agarwal, S., R. Correa, B. Morais, J. Roldan, and C. Ruiz-Ortega (2020): "Owe a Bank Millions, the Bank Has a Problem: Credit Concentration in Bad Times," *International Finance Discussion Papers*, 1288.

Aiyagari, R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109(3), 659–684.

Allen, F. and D. Gale (1998): "Optimal Financial Crises," *Journal of Finance*, 53(4).

——— (2004): "Financial Intermediaries and Markets," *Econometrica*, 72(4).

Anolli, M., E. Beccalli, and G. Borello (2015): "Are European banks too big? evidence on economies of scale," *Journal of Banking and Finance*, 58.

Ball, L., G. Mankiw, and R. Reis (2005): "Monetary policy for inattentive economies," *Journal of Monetary Economics*, 52.

Begenau, J. and T. Landvoigt (2018): "Financial Regulation in a Quantitative Model of the Modern Banking System," *Working Paper*.

Benhabib, B., A. Bisin, and M. Luo (2018): "Wealth distribution and social mobility in the US: A quantitative approach," *American Economic Review*, Forthcoming.

Bewley, T. (1977): "The permanent income hypothesis: A theoretical formulation," *Journal of Economic Theory*, 16, 252 – 292.

Blanchard, O. and N. Kiyotaki (1987): "Monopolistic Competition and the Effects of Aggregate Demand," *American Economic Review*, 77(4).

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018): "Really Uncertain Business Cycles," *Econometrica*, Forthcoming.

Boissay, F., F. Collard, and F. Smets (2016): "Booms and Banking Crises," *Journal of Political Economy*, 124(2).

Capelle, D. (2019): "Competition vs. Stability: Oligopolistic Banking System with Run Risk," *Working Paper*.

Carroll, C. D. (2006): "The method of endogenous gridpoints for solving dynamic stochastic optimization problems," *Economics Letters*, 91.

Challe, D. and X. Ragot (2015): "Precautionary Saving Over the Business Cycle," *Economic Journal*, 126.

Christiano, L. and D. Ikeda (2013): "Leverage Restrictions in a Business Cycle Model," *NBER Working Paper 18688*.

Clarida, R., J. Gali, and M. Gertler (2002): "A simple framework for international monetary policy analysis," *Journal of Monetary Economics*, 49.

Coimbra, N. and H. Rey (2019): "Financial Cycles with Heterogeneous Intermediaries," *NBER Working Paper*, 23245.

Corbae, D. and P. D'Erasmo (2019): "Capital Requirements in a Quantitative Model of Banking Industry Dynamics," *NBER Working Paper*, 25424.

——— (2020): "Rising bank concentration," *Journal of Economic Dynamics and Control*, 115.

Cuciniello, V. and F. Signoretti (2015): "Large Banks, Loan Rate Markup, and Monetary

Policy," *International Journal of Central Banking*, 11(3).

Davila, E. and T. Philippon (2019): "Incompleteness Shocks," *Working Paper*.

Davydiuk, T. (2020): "Dynamic Bank Capital Requirements," *Journal of Financial Economics*, Forthcoming.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, Forthcoming.

Den Haan, W. (2010): "Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents," *Journal of Economic Dynamics and Control*, 34, 79–99.

Diez, F., D. Leigh, and S. Tambunlertchai (2018): "Global Market Power and its Macroeconomic Implications," *IMF Working Paper*, 18/137.

Dixit, A. K. and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67.

Fostel, A. and J. Geanakoplos (2008): "Leverage Cycles and the Anxious Economy," *American Economic Review*, 98.

Galaasen, S., R. Jamilov, R. Juelsrud, and H. Rey (2020): "Granular Credit Risk," *Working Paper*.

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

Gertler, M. and P. Karadi (2011): "A Model of Unconventional Monetary Policy," *Journal of Monetary Economics*, 58(1), 17–34.

Gertler, M. and N. Kiyotaki (2010): "Financial Intermediation and Credit Policy in Business Cycle Analysis," *Handbook of Monetary Economics*, 3, 547–599.

Gertler, M., N. Kiyotaki, and A. Prestipino (2016): "Wholesale Banking and Bank Runs in Macroeconomic Modelling of Financial Crises," *Handbook of Macroeconomics*, 2.

Gertler, M., N. Kiyotaki, and A. Queralto (2012): "Financial Crises, Bank Risk Exposure and Government Financial Policy," *Journal of Monetary Economics*, 59, S17–S34.

Goldstein, I., A. Kopytov, L. Shen, and H. Xiang (2020): "Bank Heterogeneity and Financial Stability," *NBER Working Paper*.

Gopinath, G., S. Kalemli-Ozcan, L. Karabarbounis, and C. Villegas-Sanchez (2017): "Capital Allocation and Productivity in South Europe," *Quarterly Journal of Economics*, 132, 1915–1967.

Gorton, G. and A. Metrick (2010): "Regulating the Shadow Banking System," *Brookings Papers on Economic Activity*, Fall.

He, Z., B. Kelly, and A. Manela (2016): "Intermediary Asset Pricing: New Evidence from Many Asset Classes," *Journal of Financial Economics*, Forthcoming.

Hellman, T., K. Murdock, and J. Stiglitz (2000): "Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?" *American Economic Review*, 90(1).

Huggett, M. (1990): "The Risk-Free Rate in Heterogeneous Agent Economies," *Manuscript, University of Minnesota*.

Imrohoglu, A. (1996): "Costs of Business Cycles with Indivisibilities and Liquidity Constraints," *Journal of Political Economy*, 1364–83.

Jamilov, R. (2020): "A Macroeconomic Model with Heterogeneous Banks," *Working Paper*.

Jamilov, R. and T. Monacelli (2020): "Monetary Policy with (H)eterogeneous (BANK)s," *Working Paper*.

Jayaratne, J. and P. Strahan (1996): "The finance-growth nexus: evidence from bank branch deregulation," *Quarterly Journal of Economics*, 111.

Jermann, U. and V. Quadrini (2013): "Macroeconomic Effects of Financial Shocks," *American Economic Review*, 102(1), 238–271.

Khan, A. and J. Thomas (2008): "Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics," *Econometrica*, 76(2), 395–436.

——— (2013): "Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity," *Journal of Political Economy*, 121(6), 1055–1107.

Korinek, A. and M. Nowak (2016): "Risk-Taking Dynamics and Financial Stability," *Manuscript*.

Kroszner, R. and P. Strahan (2014): "Regulation and Deregulation of the U.S. Banking Industry: Causes, Consequences, and Implications for the Future," *N. Rose (eds.), Economic Regulation and Its Reform: What Have We Learned?*, NBER Books.

Krusell, P. and A. Smith (1996): "Income and Wealth Heterogeneity, Portfolio Choice, and Equilibrium Asset Returns," *Macroeconomic Dynamics*, 1, 387–422.

——— (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, 867–896.

Kydland, F. E. and E. C. Prescott (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50.

Laeven, L. and F. Valencia (2012): "Systemic Banking Crises Database: An Update," *IMF Working Paper*, 12/163).

Maliar, L., S. Maliar, and F. Valli (2010): "Solving the incomplete markets model with aggregate uncertainty using the Krusellâ€"Smith algorithm," *Journal of Economic Dynamics and Control*, 34, 42–49.

Martinez-Miera, D. and R. Repullo (2010): "Does Competition Reduce the Risk of Bank Failure?" *The Review of Financial Studies*, 23.

Melitz, M. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6).

Mendoza, E. G. (2010): "Sudden Stops, Financial Crises, and Leverage," *American Economic Review*, 100.

Merton, R. C. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41.

Nakamura, E. and J. Steinson (2010): "Monetary Non-neutrality in a Multisector Menu Cost Model," *Quarterly Journal of Economics*, 125(3).

Nguyen, T. (2015): "Bank Capital Requirements: A Quantitative Analysis," *Working Paper*.

Nuno, G. and C. Thomas (2016): "Bank Leverage Cycles," *American Economic Journal: Macroeconomics*, Forthcoming.

Paravisini, D., V. Rappoport, and P. Schnabl (2020): "Specialization in Bank Lending: Evidence from Exporting Firms," *Working Paper*.

Rios Rull, V., T. Takamura, and Y. Terajima (2018): "Banking Dynamics and Capital Requirements," *Manuscript*.

Sterk, V. and M. Ravn (2020): "Macroeconomic Fluctuations with HANK and SAM: an Analytical Approach," *Journal of the European Economic Association*, Forthcoming.

Tauchen, G. (1986): "Finite state markov-chain approximations to univariate and vector autoregressions," *Economics Letters*, 20.

ONLINE APPENDIX

Bewley Banks

by Rustam Jamilov and Tommaso Monacelli

A Empirical Appendix

A.1 Data Description

We acquire financial intermediary balance sheet data for the U.S. (total assets and net worth) from Compustat North America – Fundamentals Quarterly. We include all institutions belonging to SIC sectors "Finance, Insurance and Real Estate" (all codes beginning with 6). We use the variable "ATQ – Total Assets Quarterly" for total assets, "CEQQ – Common Equity Quarterly" for net worth and we compute leverage at the institution-quarter level as the ratio of total assets over net worth. For robustness, we also consider an additional measure of net worth, "SEQQ – Stockholders Equity Quarterly" and results are not affected in a material way. Using the data on assets and equity, we construct the book leverage ratio as the ratio of bank assets over equity. Our baseline sample for total assets, net worth and leverage is over 1985Q1-2020Q1. We experiment with alternative sample durations in Section **??**.

In order to construct bank-level measures of the loan margin, we extract data on interest and non-interest revenues and expenses. We use the Compustat Banks – Fundamentals Quarterly database due to better coverage of income-statement data. As a result, our measure of loan margins is computed only for institutions belonging to the 2-digits SIC sector 60 (Depository institutions).

All throughout, our sample includes companies with headquarters located in the US. We exclude companies that report earnings in any other currency except the USD. All variables are deflated using the U.S. GDP implicit price deflator published in the OECD Main Economic Indicators.[14] We clean the sample from observations that are either erroneous or are extremely outliers. Specifically, we drop observations with book leverage above 100 or smaller than 1 as well as all cases of negative net worth (equity). For our main analysis we focus only on institutions appearing in the dataset for at least 80 quarters. We present robustness checks for different sample definitions in Section **??**.

We construct four proxies for loan margins. Our main measure of margins is computed as the ratio of Total Interest and Related Income (IDITQ) over Total Interest and Related Expenses (XINTQ). For robustness, we also compute the ratio between Net Interest Income (NIINTQ) and Total Interest and Related Expenses. Because of data availability, both of these measures are only available for 1993Q1-2020Q1. In addition, we construct the two following proxies: Net Current Operating Earnings (NCOEQ) divided by Total Interest and Related Expenses and Interest and Fees on Loans (IDILBCQ) over Total Interest and Related Expenses. These measures are available for the full sample 1985Q1-2020Q1. All four definitions yield quantitatively very similar business cycle correlations and volatilities. We prefer the first measure because it is the most complete and captures all relevant factors related to either income or expense.

In order to construct a bank-level measure of bank default risk, we use the Markit database on Credit Default Swaps (CDS). Our baseline measure is the 5-year CDS spread because it is the most liquid among all the maturities. If we consider 6-months, 1-year or 10-year CDS spreads, results do not change. Our sample refers to CDS contracts that are issued and traded in US Dollars. We restrict reference entities to the "Financial" sector in the US. Our sample runs from 2002Q1 until 2020Q1. For each institution in the sample, we construct quarterly aggregates from daily CDS data.

Before executing our main empirical exercises, we truncate the sample at the 1 and 99 percentiles

[14]We download this measure from the St Louis Federal Reserve, series USAGDPDEFQISMEI.

for leverage, margins and default risk. We do not perform any truncation for total assets and net worth. Generally, truncation of balance sheet items does not affect the results. Truncation of leverage, margins, and CDS spreads helps tighten up correlations of higher-order moments which are usually affected by extreme outliers.

Throughout the paper, we work with the first three moments of the time-varying distributions of intermediary assets, net worth, leverage, loan margins, and CDS spreads. Unless stated otherwise in the text, we always proxy the first moment with the unconditional mean and the second moment with the standard deviation. Our proxy for concentration depends on the variable at hand. For assets and equity we calculate the Herfindahl index (HHI) and for leverage, loan margins, and CDS spreads we compute statistical skewness[15]. Unless stated otherwise, we also always log-linearly detrend all those moments before computing any statistic. To compute business cycle correlations, we log-linearly detrend the quarterly real GDP of the U.S. which we obtain directly from the St. Louis Federal Reserve Board.

As for countries other than the U.S., we acquire financial intermediary balance sheet data from Compustat North America – Fundamentals Quarterly for Canada and from Compustat Global – Fundamentals Quarterly for all the others. We use the same procedures followed for U.S. data with two exceptions. First, we use "SEQQ – Stockholders Equity Quarterly" as our main proxy of net worth, since this is the only variable widely available for all institutions. As a result, we also compute leverage based on this variable. Second, to avoid dealing with an unbalanced number of observations across quarters, we drop from our sample all those institutions that never report for two consecutive quarters. That is, as long as a company has reported for two consecutive quarters at least once, we keep this company in our sample.

Because of data availability, our samples span different time periods according to the reference country. In particular, data for Australia cover 1997Q1-2019Q4, for Canada 1991Q1-2020Q1, for France 2001Q4-2019Q4, for Germany 2005Q4-2019Q4 and for U.K. 1996Q4-2020Q1.

For CDS we follow again the same steps illustrated for U.S. data. We do not report CDS correlations for Australia, since Markit does not cover Australian companies. Our data cover the same time period as for the U.S., that is 2002Q1-2020Q1, with the exception of Canada, for which the sample starts in 2002Q3.

For these countries, we do not construct a measure of margins, because of insufficient coverage of income-statement data.

[15]We compute HHI for variable x according to the usual formula: $HHI_t(x) = \sum_i \left(\frac{x_{it}}{x_t}\right)^2$.

A.2 Heterogeneity by Intermediary Sub-Sector

As mentioned in the main text, there is multi-modality in the distribution of financial intermediaries. It is important to complement our analysis of the aggregate financial sector with industry-level decomposition. We therefore now perform the same statistical exercise but for each of the six major sub-industries of the broader financial sector: depository credit institutions, non-depository credit institutions, broker-dealers, insurance companies, real estate companies, and holdings and investors. As before, we focus on the U.S. only. Due to data limitations, we cannot construct measures of credit margins or default risk for individual industries. Table A.1 reports the results. Immediately, we notice that there is considerable heterogeneity across sectors. We can summarize all of the notable observations in three general points. First, and perhaps most interestingly, for some sectors like depository institutions and real estate agents the mean of assets and net worth is *counter-cyclical*. Second, although leverage of the aggregate sector is pro-cyclical, it is in fact counter-cyclical for the depository institutions and non-depository credit-granting institutions. This result is consistent with the empirical evidence and mechanisms discussed in He et al. (2016). Third, as is the case with the aggregate sector, concentration of balance sheet characteristics is almost always counter-cyclical.

We now briefly elaborate more on the role of industry heterogeneity and how it relates to our structural model. Our model features a single aggregated financial sector. On the other hand, empirically, it may be more fruitful to look at the data sector-by-sector. In order to facilitate a fair comparison, we will target the aggregate financial sector as our benchmark for data-model correspondence. However, we acknowledge that business cycle fluctuations could differ noticeably across sub-sectors along the within-sector intensive margin and the extensive margin, i.e., fluctuations in the size of each subsector across time.[16]

[16]Our model could be readily extended to include multiple sectors. For example, a fraction of productive capital could be intermediated by intermediaries that face a value-at-risk constraint, i.e., "broker-dealers". The remaining capital would be managed by investors that face an equity-based constraint on risk-taking. In equilibrium, the two sectors would deliver pro-cyclical and counter-cyclical market leverage, respectively. The extensive margin, which could be endogenized, would be crucial in this setup. However, coupled with credit market power and idiosyncratic rate of return risk, this extension is currently computationally infeasible and is beyond the scope of this paper. It is, however, a very fruitful topic for future research.

Table A.1: **Business Cycle Correlations - U.S. Data by Sub-Sector**

	Mean of	Dispersion of	Concentration of	Mean of	Dispersion of	Concentration of
	Depository Institutions (SIC 60)			Insurance (SIC 63 and 64)		
Assets - GDP	-0.521	-0.093	-0.513	0.826	0.762	-0.300
Net Worth - GDP	-0.454	0.361	-0.219	0.497	0.689	-0.210
Leverage - GDP	-0.164	-0.186	-0.121	0.593	0.346	-0.159
	Non-Depository Institutions (SIC 61)			Real Estate (SIC 65)		
Assets - GDP	0.816	0.826	-0.522	-0.435	-0.429	-0.554
Net Worth - GDP	0.663	0.686	-0.441	-0.521	-0.572	-0.740
Leverage - GDP	-0.209	-0.048	0.090	0.251	0.133	0.007
	Brokers and Dealers (SIC 62)			Holdings and Investors (SIC 67)		
Assets - GDP	0.896	0.882	-0.685	0.738	-0.010	-0.836
Net Worth - GDP	0.525	0.454	-0.758	0.608	0.320	-0.703
Leverage - GDP	0.364	0.705	0.403	0.424	0.223	0.165

Notes: For every variable except CDS spreads the sample is 1985q1:2020q1. Every variable has been logged (except the skewness of leverage) and linearly detrended. Bank balance sheet data is from Compustat North America. Industry classification follows the first two digits of the Standard Industrial Classification of economic activities (SIC).

A.3 Data on Other Countries

Subject to data availability, we also report business cycle correlations of higher-order moments in other countries. We managed to build reasonably long panels for Australia, Canada, France, Germany, and the United Kingdom. Details on sample construction are in the Appendix. To the best of our knowledge, this is the first attempt to establish robust facts of this type for a set of non-US developed economies. Results are reported in Table A.2. We can summarize these statistics in three broad points. First, it's most clear that CDS spreads have counter-cyclical first and second (except for Canada) moments, and pro-cyclical skewness. Second, leverage is counter-cyclical for all countries except Australia. There is no systematic commonality for the higher-order moments of leverage. Third, the country that appears to be closest to the U.S. in terms of these business cycle patterns is Australia. Overall, there is substantial degree of heterogeneity across countries for almost every characteristic.[17]

Table A.2: **Business Cycle Correlations - Aggregate Data for Different Countries**

	Mean of	Dispersion of	Concentration of	Mean of	Dispersion of	Concentration of
		Australia			Germany	
Assets - GDP	0.207	-0.008	-0.614	-0.142	-0.039	0.012
Net Worth - GDP	0.400	0.170	-0.574	-0.044	0.032	-0.087
Leverage - GDP	0.261	0.106	-0.123	-0.189	-0.151	-0.014
CDS Spreads				-0.329	-0.063	0.112
		Canada			United Kingdom	
Assets - GDP	-0.832	-0.655	-0.109	0.042	0.234	-0.369
Net Worth - GDP	-0.772	-0.751	-0.686	-0.161	0.014	-0.779
Leverage - GDP	-0.805	-0.531	0.790	-0.085	0.124	0.528
CDS Spreads	-0.197	0.026	0.063	-0.527	-0.158	0.455
		France				
Assets - GDP	0.170	0.226	0.053			
Net Worth - GDP	0.156	0.167	-0.302			
Leverage - GDP	-0.122	0.067	0.300			
CDS Spreads	-0.355	-0.322	0.282			

Notes: Every variable has been logged linearly detrended. Bank balance sheet data is from Compustat. CDS data is from Markit. The sample for Australia is 1997q1:2019q4, for Canada is 1991q1:2020q1, for France is 2001q4:2019q4, for Germany is is 2005q4:2020q1, and for the UK is 1996q4:2020q1. See the Appendix for variable definitions and further details.

[17] Similar to our analysis of cross-industry heterogeneity, a multi-country extension of our framework is possible. The mass of differentiated local credit markets could be re-routed to represent a continuum of countries with ex-ante heterogeneity in size or magnitude of local idiosyncratic riskiness. A single financial intermediary would thus charge country-specific margins over country-specific costs of funds. This extension is on the agenda for future research.

A.4 Robustness

For robustness, we report in Tables A.3, A.4, A.5 and A.6 below the correlations using different sample definitions. Tables A.3 and A.4 show correlations both over the whole sample and by sub-industries when we include all institutions appearing at least 50 quarters (instead of 80, as in the main results) and leave the starting date unchanged to 1985Q1.

Similarly, Tables A.5 and A.6 report correlations for the sample starting in 1980Q1 and including all institutions appearing at least 80 quarters.

Table A.3: **Correlations with GDP - US Data, Less Balanced Panel**

	Mean of	St Deviation of	Concentration of
Assets - GDP	0.221	0.631	-0.329
Net Worth - GDP	-0.252	0.511	-0.256
Leverage - GDP	0.670	-0.148	-0.741
Margins - GDP	-0.571	-0.345	0.751
Default Risk - GDP	-0.325	-0.309	0.033

Notes: Aggregate business cycle correlations based on the panel of financial intermediaries that contains only institutions with at least 50 (nonconsecutive) quarters of data over the 1985q1-2020q1 sample. Balance sheet data comes from Compustat. Default risk (CDS) data is from Markit.

Table A.4: **Business Cycle Correlations - U.S. Data by Sub-Sector, Less Balanced Panel**

	Mean of	Dispersion of	Concentration of	Mean of	Dispersion of	Concentration of
	Depository Institutions (SIC 60)			Insurance (SIC 63 and 64)		
Assets - GDP	-0.762	-0.342	-0.299	0.665	0.745	-0.138
Net Worth - GDP	-0.764	0.131	-0.091	0.194	0.659	-0.063
Leverage - GDP	-0.266	-0.117	-0.378	0.055	-0.139	-0.154
	Non-Depository Institutions (SIC 61)			Real Estate (SIC 65)		
Assets - GDP	0.803	0.817	-0.294	-0.446	-0.385	-0.408
Net Worth - GDP	0.635	0.670	-0.223	-0.437	-0.346	-0.379
Leverage - GDP	0.007	0.093	-0.053	0.267	0.157	0.006
	Brokers and Dealers (SIC 62)			Holdings and Investors (SIC 67)		
Assets - GDP	0.923	0.907	-0.683	0.791	0.228	-0.771
Net Worth - GDP	0.644	0.527	-0.726	0.501	0.286	-0.649
Leverage - GDP	-0.034	0.564	0.362	0.512	0.229	-0.052

Notes: Industry-level business cycle correlations based on the panel of financial intermediaries that contains only institutions with at least 50 (nonconsecutive) quarters of data over the 1985q1-2020q1 sample. Balance sheet data comes from Compustat.

Table A.5: **Correlations with GDP - US Data, starting 1980**

	Mean of	St Deviation of	Concentration of
Assets - GDP	0.313	0.383	-0.634
Net Worth - GDP	-0.139	0.203	-0.590
Leverage - GDP	0.254	0.077	-0.313
margins - GDP	-0.485	-0.311	0.624
Default Risk - GDP	-0.281	-0.269	0.016

Notes: Aggregate business cycle correlations based on the sample of financial intermediaries that starts from 1980q1. Balance sheet data comes from Compustat. Default risk (CDS) data is from Markit.

Table A.6: **Business Cycle Correlations - U.S. Data by Sub-Sector, starting 1980**

	Mean of	Dispersion of	Concentration of	Mean of	Dispersion of	Concentration of
	Depository Institutions (SIC 60)			Insurance (SIC 63 and 64)		
Assets - GDP	-0.521	-0.270	-0.606	0.261	0.560	-0.445
Net Worth - GDP	-0.442	0.068	-0.475	-0.155	0.462	-0.412
Leverage - GDP	-0.269	-0.114	-0.103	0.438	0.489	0.242
	Non-Depository Institutions (SIC 61)			Real Estate (SIC 65)		
Assets - GDP	0.580	0.663	-0.544	-0.231	-0.154	-0.332
Net Worth - GDP	0.336	0.448	-0.514	-0.494	-0.440	-0.652
Leverage - GDP	-0.005	0.032	-0.166	0.326	0.253	0.281
	Brokers and Dealers (SIC 62)			Holdings and Investors (SIC 67)		
Assets - GDP	0.871	0.871	-0.743	0.444	0.102	-0.736
Net Worth - GDP	-0.008	0.080	-0.777	-0.027	-0.180	-0.772
Leverage - GDP	-0.005	0.032	-0.166	0.570	0.508	0.389

Notes: Industry-level business cycle correlations based on the sample of financial intermediaries that starts from 1980q1. Balance sheet data comes from Compustat.

B Model Appendix: Proofs

Proof of Proposition 1
The bank solves for each j:

$$\max_{k(j)} \Big(1-\upsilon(j)\Big)R^T(j)p(j)k(j) - \bar{R}(j)\Big(p(j)k(j)^{\beta} - n(j)\Big) \quad \text{s.t.} \quad k(j) = \left(\frac{p(j)}{P}\right)^{-\theta} K(\mathbf{S})$$

The first order condition is

$$\Big(1-\upsilon(j)\Big)R^T(j)p(j) + \Big(1-\upsilon(j)\Big)R^T(j)\frac{\partial p(j)}{\partial k(j)}k(j) - \bar{R}(j)\left(p(j)\beta k(j)^{\beta-1} + k(j)^{\beta}\frac{\partial p(j)}{\partial k(j)}\right) = 0$$

The elasticity of substitution, ignoring the influence of local credit market-level rates on the aggregate index $P(\mathbf{S})$, is

$$\frac{\partial k(j)}{\partial p(j)}\frac{p(j)}{k(j)} = -\theta$$

The price-setting rule given marginal costs is

$$p(j) = \frac{\theta}{\theta-1}MC(j)$$

where $\frac{\theta}{\theta-1}$ is the constant markup over the (endogenous) marginal cost MC(j), given by:

$$MC(j) := \frac{\beta\theta-1}{\theta}p(j)\frac{\bar{R}(j)}{\Big(1-\upsilon(j)\Big)R^T(j)}\left[\left(\frac{p(j)}{P(\mathbf{S})}\right)^{-\theta}K(\mathbf{S})\right]^{\beta-1}$$

Proof of Proposition 2
Guess that the solution to the dynamic problem 20 is a value function $V(n(j),\xi(j);\mathbf{S}) = \zeta(n(j),\xi(j);\mathbf{S})n(j)$. Define the default risk-adjusted stochastic discount factor $\tilde{\Lambda}(\mathbf{s};\mathbf{S}) = \Big[\Big(1-\upsilon(j)\Big)\Lambda(\mathbf{S})\Big(1-\sigma+\sigma\zeta(n'(j),\xi'(j);\mathbf{S}')\Big)\Big]$. The solution to the program is a system of equations:

$$\mathbb{E}\left[\tilde{\Lambda}(\mathbf{s'};\mathbf{S'})\left(R^{T'}(j) - \bar{R}(j)k(j)^{\beta-1}\right)\right] = \lambda\varphi(n(j),\xi(j);\mathbf{S})$$

$$\varphi(n(j),\xi(j);\mathbf{S})\left[\zeta(n(j),\xi(j);\mathbf{S}) - \lambda\phi(j)\right] = 0$$

Substituting the optimality conditions together with the guess into the objective function gives

$$\zeta(n(j),\xi(j);\mathbf{S}) = \varphi(n(j),\xi(j);\mathbf{S})\zeta(n(j),\xi(j);\mathbf{S}) + \mathbb{E}\Big(\tilde{\Lambda}(\mathbf{s'};\mathbf{S}')\Big)\bar{R}(j)k(j)^{\beta-1}$$

Solving for $\zeta(n(j),\xi(j);\mathbf{S})$ yields

$$\zeta(n(j),\xi(j);\mathbf{S}) = \frac{\mathbb{E}\Big(\tilde{\Lambda}(\mathbf{s}';\mathbf{S}')\Big)k(j)^{\beta-1}\bar{R}(j)}{1-\varphi(n(j),\xi(j);\mathbf{S})}$$

And the Lagrange multiplier on the leverage constraint is

$$\varphi(\mathrm{n(j)}, \xi(\mathrm{j}); \mathbf{S}) = \max \left[1 - \frac{\mathbb{E}\Big(\tilde{\Lambda}(\mathbf{s}'; \mathbf{S}')\Big)\mathrm{k(j)}^{\beta-1}\bar{\mathrm{R}}(\mathrm{j})}{\lambda\phi(\mathrm{j})}, 0 \right]$$

The result follows from (a) the fact that market leverage is $\phi(\mathrm{j}) = \mathrm{k(j)}^{1-\frac{1}{\theta}}\Big(\mathrm{K}(\mathbf{S})\Big)^{\frac{1}{\theta}}\mathrm{P}(\mathbf{S})\mathrm{n(j)}^{-1}$ (b) and the previously defined augmented stochastic discount factor $\tilde{\Lambda}(\mathbf{s}; \mathbf{S})$. The guess is verified if $\varphi(\mathrm{n(j)}, \xi(\mathrm{j}); \mathbf{S}) < 1$. Size-dependency is guaranteed by $\beta > 1$ so that each bank with different n(j) and ξ(j) chooses a different leverage ratio ϕ(j).

C Model Appendix: Accuracy

In this section we discuss the accuracy of our main numerical algorithm. Our convergence tolerance levels for the household and banking problems are $10\mathrm{e}^{-8}$ and $10\mathrm{e}^{-5}$, respectively. Deposit market clearing is achieved under the tolerance level of $10\mathrm{e}^{-5}$ for the deposit rate on each idiosynractic grid point. Finally, tolerance level for the Krusell-Smith recursion is $10\mathrm{e}^{-3}$ for both capital and prices. We perform two exercises. First, we report the R-squared from the Γ projections for K′ and P′. With a log-linear form, equilibrium of the model with aggregate uncertainty in ψ_t is characterized by the following equations for good and bad times:

$$\begin{aligned} \log(\mathrm{K}') &= 0.3380 + 0.8588\log(\mathrm{K}); \quad \mathrm{R}^2 = 0.9746; \quad \psi \text{ low} \\ \log(\mathrm{K}') &= 0.3668 + 0.8671\log(\mathrm{K}); \quad \mathrm{R}^2 = 0.9737; \quad \psi \text{ high} \end{aligned}$$

For projections of K′ and

$$\begin{aligned} \log(\mathrm{P}') &= 1.3871 - 0.4788\log(\mathrm{K}); \quad \mathrm{R}^2 = 0.9909; \quad \psi \text{ low} \\ \log(\mathrm{P}') &= 1.5854 - 0.4946\log(\mathrm{K}); \quad \mathrm{R}^2 = 0.9923; \quad \psi \text{ high} \end{aligned}$$

For projections of P′.

We also compute the accuracy measure of Den Haan (2010) for the stationary law of motion for aggregate capital and prices. Using the equilibrium law of motion Γ above, without updating this forecast function, we simulate the full-time series of capital and prices using the ψ_t as the source of stochastic exogenous fluctuations. We compare this simulation with the stochastic simulation time-series where aggregate capital and prices are built from the time-varying distribution. The average percentage error is 1.68% for capital and 0.98% for prices. These numbers are in line with existing studies that are similar to our model's degree of non-linearity and complexity (Khan and Thomas, 2008; Nakamura and Steinson, 2010; Corbae and D'Erasmo, 2019).