FEDERAL RESERVE BANK of ST. LOUIS

RESEARCH DIVISION

Working Paper Series

The Tail that Wags the Economy: Beliefs and Persistent Stagnation

Julian Kozlowski,
Laura Veldkamp
and
Venky Venkateswaran

Working Paper 2019-006A
https://doi.org/10.20955/wp.2019.006

February 2019

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

The Tail that Wags the Economy: Beliefs and Persistent Stagnation

Julian Kozlowski
FRB St Louis

Laura Veldkamp
Columbia, CEPR, NBER

Venky Venkateswaran*
FRB Minneapolis, NYU Stern, NBER

February 3, 2019

Abstract

The Great Recession was a deep downturn with long-lasting effects on credit, employment and output. While narratives about its causes abound, the persistence of GDP below pre-crisis trends remains puzzling. We propose a simple persistence mechanism that can be quantified and combined with existing models. Our key premise is that agents don't know the true distribution of shocks, but use data to estimate it non-parametrically. Then, transitory events, especially extreme ones, generate persistent changes in beliefs and macro outcomes. Embedding this mechanism in a neoclassical model, we find that it endogenously generates persistent drops in economic activity after tail events.

JEL Classifications: D84, E32

Keywords: Stagnation, tail risk, propagation, belief-driven business cycles

*kozjuli@nyu.edu; lveldkam@stern.nyu.edu; vvenkate@stern.nyu.edu. We thank our four anonymous referees and our Editor, Harald Uhlig, as well as Mark Gertler, Mike Golosov, Mike Chernov, Francois Gourio, Christian Hellwig and participants at the Paris Conference on Economic Uncertainty, IMF Secular Stagnation conference, Einaudi, SED and NBER summer institute for helpful comments and Philippe Andrade, Jennifer La'O, Franck Portier and Robert Ulbricht for their insightful discussions. We thank Adrien Avernas for providing us the data on credit spreads of bank holding companies. Veldkamp and Venkateswaran gratefully acknowledge financial support from the NYU Stern Center for Global Economy and Business.

The Great Recession was a deep downturn with long-lasting effects on credit markets, labor markets and output. Why did output remain below trend long after financial markets had calmed and uncertainty diminished? Why did the usual business cycle recovery not occur after this recession? Such a persistent, downward shift in output (Figure 1) is not unique to the 2008 crisis. Financial crises, even in advanced economies, typically fail to produce the robust GDP rebound needed to restore output to pre-crisis trends.[1]

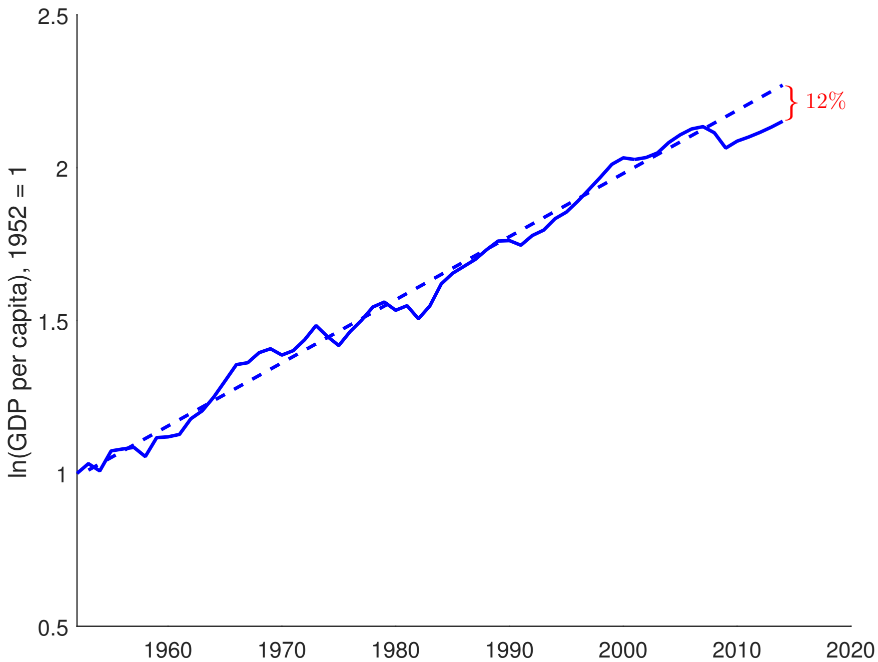


Figure 1: **Real GDP in the U.S. and its trend.**
Dashed line is a linear trend that fits data from 1950-2007. In 2014, real GDP was 0.12 *log points below trend.*

Our explanation is that crises produce persistent effects because they scar our beliefs. For example, in 2006, few people entertained the possibility of financial collapse in the U.S. Today, the possibility of another run on the financial sector is raised frequently, even though the system today is probably much safer. Such persistent changes in the assessments of risk came from observing new data. We thought the U.S. financial system was stable. Economic outcomes taught us that the risks were greater than we thought. It is this new-found knowledge that is inducing long-lived effects on economic choices.

The contribution of the paper is a simple tool to capture and quantify this scarring effect, which produces more persistent responses from extreme shocks than from ordinary business cycle shocks. We start from a simple assumption: agents do not know the true distribution of shocks in the economy, but estimate the distribution using real time data, exactly like an econometrician would. The scarcity of data on extreme events is what makes new tail observations

[1] See Reinhart and Rogoff (2009), fig 10.4.

particularly informative. Therefore, tail events trigger larger belief revisions. Furthermore, because it will take many more observations of non-tail events to convince someone that the tail event really is unlikely, changes in tail risk beliefs are particularly persistent. To explore these changes in a meaningful way, we need to use an estimation procedure that does not unduly constrain the shape of the distribution's tail. Therefore, we assume that our agents adopt a non-parametric approach to learning about the distribution of aggregate shocks. Each period, they observe one more piece of data and update their estimates using a standard kernel density estimator. Section 1 shows that this process leads to long-lived responses of beliefs to transitory events, especially extreme, unlikely ones. The mathematical foundation for persistence is the martingale property of beliefs. The logic is that once observed, the event remains in agents' data set. Long after the direct effect of the shock has passed, the knowledge of that tail event continues to affect estimated beliefs and restrains the economic recovery.

To illustrate the economic importance of these belief dynamics, Section 2 applies our belief updating tool to a well-known model used recently to analyze the Great Recession. The environment closely follows Gourio (2012, 2013) and is well-suited to our purposes because it provides a simple and quantitatively plausible link from tail risk to macro outcomes. At its core are firms subject to bankruptcy risk from aggregate capital quality shocks as well as idiosyncratic shocks to profitability. This set of economic assumptions is not our contribution. It is simply a convenient laboratory to illustrate the persistent economic effects from observing extreme events. We add one other ingredient – a financial sector, which intermediates between these firms and households. While not central to our story, this allows us to incorporate changes in beliefs about financial shocks and improve the model's ability to match the data. Section 3 describes the data we feed into the model to discipline our belief estimates. Section 4 shows that belief updating can, both qualitatively and quantitatively, explain the persistently low level of recent economic activity colloquially known as "secular stagnation." We highlight the role of our mechanism by comparing our results to those from the same economic model, but without belief revisions, i.e. with agents who have full knowledge of the distribution.

The mechanism through which tail events have persistent effects does not depend on the specific structure of the Gourio (2012) model. It requires three key ingredients. One is a shock process that can capture the extreme, unusual aspects of a tail event. During the Great Recession, these were evident mainly in real estate and capital markets. Total factor productivity shocks do not meet this criterion.[2] The capital quality shock specification is arguably the simplest one that does. Was this the first time we have ever seen such shocks? In our data set, which spans the post-WWII period in the US, yes. Of course, similar extreme events have been observed before in global history – e.g. during the Great Depression or in other countries.

[2]The fall in TFP was not particularly extreme and predates the crisis. See Appendix D.5 for more details.

Section 4.3 explores the effect of expanding the data set to include additional infrequent crises and shows that it does temper persistence, but only modestly.

The second ingredient is a belief updating process that uses new data to estimate the distribution of shocks, or more precisely, the probability of extreme events. It is not crucial that the estimation is frequentist.[3] What is important is that the learning protocol does not rule out fat tails by assumption (e.g. by imposing a normal distribution).

The third necessary ingredient is an economic model that links the risk of extreme events to real output. The model in Gourio (2012, 2013) has sufficient sources of non-linearity in policy functions to deliver sizable output responses from modest changes in disaster risk. The assumptions about preferences and debt/bankruptcy, that make Gourio's model somewhat complex, are there to deliver that curvature. They also make the economy more sensitive to disaster risk than extreme boom risk. Section 4.5 explores the role of these ingredients, by turning each on and off.

Finally, we show that recent data on asset prices and debt are also consistent with an increase in tail risk. The higher perceived risk of financial crises in the future raises credit spreads both for financial and non-financial firms. The magnitudes line up reasonably well with changes in the data. One might think a rise in tail risk should push down equity prices, when in fact, they have rebounded. Our model argues against this hypothesis – when tail risk rises, firms borrow less to avoid the risk of bankruptcy, which tends to increase the value of their equity claims. Thus, low credit spreads and a rise in equity prices are not inconsistent with tail risk. Others point to low interest rates as a potential cause of stagnation. Our story complements this narrative by demonstrating how heightened tail risk makes safe assets more attractive, depressing riskless rates in a persistent fashion. In sum, none of these patterns disproves our theory about elevated tail risk, though, in fairness, they also do not distinguish it from others.

There are other asset market variables that speak more directly to tail risk, e.g. options on the S&P 500 index. Figure 2 shows that the SKEW, an index of implied skewness constructed by the Chicago Board Options Exchange from traded option prices, has stayed persistently high. In Section 4.4, we use this series to show that the model's predictions for changes in of tail risk – specifically, the third moment of equity returns and the implied probability of large negative returns – lines up quite well with the data. Finally, other proxies for beliefs also show signs of persistently higher tail risk today. Google searches for the terms "economic crisis," "financial crisis," or "systematic risk" all rose during the crisis and never returned to their pre-crisis levels (see Appendix D.1).

[3]See Orlik and Veldkamp (2014) for an example of Bayesian estimation of tail risks.

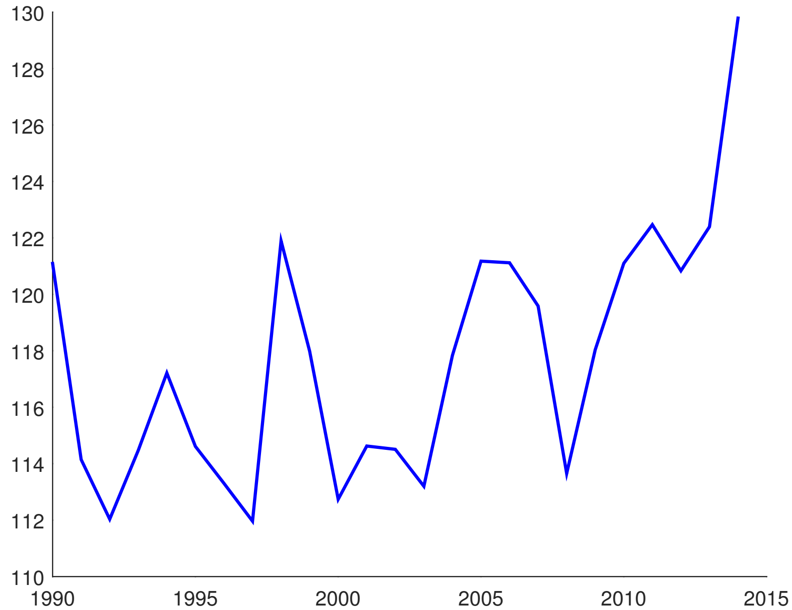


Figure 2: **The SKEW Index.**
An index of skewness in returns on the S&P 500, constructed using option prices. Source: Chicago Board Options Exchange (CBOE). 1990:2014.

Comparison to the literature There are many theories now of the financial crisis and its consequences, many of which provide a more detailed account of its mechanics (e.g., Gertler et al. (2010), Gertler and Karadi (2011), Brunnermeier and Sannikov (2014) and Gourio (2012, 2013)). Our goal is not a new explanation for why the crisis arose, or a new theory of business cycles. Rather, we offer a belief-based mechanism that complements these theories by adding endogenous persistence. It helps explain why extreme events, like the recent crisis, lead to more persistent responses than milder downturns. In the process, we also develop a new tool for tying beliefs firmly to data that is compatible with modern, quantitative macro models.

A few uncertainty-based theories of business cycles also deliver persistent effects from transitory shocks. In Straub and Ulbricht (2013) and Van Nieuwerburgh and Veldkamp (2006), a negative shock to output raises uncertainty, which feeds back to lower output, which in turn creates more uncertainty. Fajgelbaum et al. (2014) combine this mechanism with an irreversible investment cost, a combination which can generate multiple steady-states. These uncertainty-based explanations leave two questions unanswered. First, why did economic activity stay depressed long after measures of uncertainty (like the VIX) had recovered? Our theory emphasizes tail risk. Unlike measures of uncertainty, tail risk has lingered (as Figure 2 reveals), making it a better candidate for explaining the continued stagnation. Second, why were credit markets most persistently impaired after the crisis? Rises in tail risk hit credit markets because default risk is particularly sensitive to tail events.

Our belief formation process is similar to the parameter learning models by Johannes et al. (2015), Cogley and Sargent (2005) and Orlik and Veldkamp (2014) and is advocated by Hansen (2007). However, these papers focus on endowment economies and do not analyze the potential for persistent effects in production settings. Pintus and Suda (2015) embed parameter learning in a production economy, but feed in persistent leverage shocks and explore the potential for amplification when agents hold erroneous initial beliefs about persistence. In Moriera and Savov (2015), learning changes demand for shadow banking (debt) assets. But, again, agents are learning about a hidden two-state Markov process, which has a degree of persistence built in.[4] While this literature has taught us a lot about the mechanisms that triggered declines in lending and output, it often has to resort to exogenous persistence. We, on the other hand, have transitory shocks and focus on endogenous persistence. In addition, our non-parametric approach allows us to talk about tail risk.

Finally, our paper contributes to the recent literature on secular stagnation. Eggertsson and Mehrotra (2014) argue that a combination of low effective demand and the zero lower bound on nominal rates can generate a long-lived slump. In contrast, Gordon (2014), Anzoategui et al. (2015) and others attribute stagnation to a decline in productivity, education or shift in demographics. Hall (2015a) surveys these and other theories. But, while these longer-run trends may well be suppressing growth, they don't explain the level shift in output after with the financial crisis. So, while they may well be part of the explanation, our simple mechanism reconciles the recent stagnation with economic, financial and internet search evidence suggesting heightened tail risk.

The rest of the paper is organized as follows. Section 1 describes the belief-formation mechanism. Section 2 presents the economic model. Section 3 shows the measurement of shocks and calibration of the model. Section 4 analyzes the main results of the paper while Section 4.5 decomposes the key underlying economic forces. Finally, Section 5 concludes.

1 Belief Formation

A key contribution of this paper is to explain why tail risk fluctuations are persistent. Before laying out the underlying economic environment, we begin by explaining the novel part – belief revisions and their persistence. In order to do this, it is essential to depart from the assumption that agents know the true distribution of shocks to the economy. Instead, we assume that they estimate such distributions, updating beliefs as new data arrives. The first step is to choose a

[4] Other learning papers in this vein include papers on news shocks, such as, Beaudry and Portier (2004), Lorenzoni (2009), Veldkamp and Wolfers (2007), uncertainty shocks, such as Jaimovich and Rebelo (2006), Bloom et al. (2014), Nimark (2014) and higher-order belief shocks, such as Angeletos and La'O (2013) or Huo and Takayama (2015).

particular estimation procedure. A common approach is to assume that shocks follow a normal distribution and estimate its parameters (namely, mean and variance). While tractable, its thin tails make the normal distribution unsuited to thinking about tail risk changes. We could choose a distribution with more flexibility in higher moments. However, this would raise concerns about the sensitivity of results to the specific distributional form. To minimize such concerns, we take a non-parametric approach and let the data inform the shape of the distribution.

Specifically, we employ a kernel density estimation procedure, one of most common approaches in non-parametric estimation. Essentially, it approximates the true distribution function with a smoothed version of a histogram constructed from the observed data. By using the widely-used normal kernel, we impose a lot of discipline on our learning problem but also allow for considerable flexibility. We also experimented with a handful of other kernel and Bayesian specifications, which yielded similar results (see Appendix C.11).

Setup Consider a $d \times 1$ shock vector $\mathbf{x}_t$ whose true density g is unknown to agents in the economy. They do know that it is i.i.d. Their information set at time t, denoted $\mathcal{I}_t$, is the observed history of those shocks $\{\mathbf{x}_{t-s}\}_{s=0}^{n_t-1}$. They use the available data at every date to construct an estimate $\hat{g}_t$, using the following normal kernel density estimator:

$$\hat{g}_t\left(\mathbf{x}\right) = \frac{1}{n_t}\sum_{s=0}^{n_t-1}\Omega\left(\mathbf{x}-\mathbf{x}_{t-s};\boldsymbol{\Xi}_t\right)$$

where n_t is the number of available observations at t, $\Omega\left(\cdot\right)$ is the multivariate normal density function with covariance $\boldsymbol{\Xi}_t$, also referred to as the smoothing or bandwidth matrix. We use the reference rule for the optimal bandwidth, where $\boldsymbol{\Xi}_t$ is a diagonal matrix with $\hat{\sigma}_j\left(\frac{4}{(2+d)n_t}\right)^{1/(d+4)}$ as the only non-zero element in its $j-$th row ($\hat{\sigma}_j$ is the sample standard deviation of shock j).[5] As new data arrives, agents update their estimates, generating a sequence of $\{\hat{g}_t\}$.

Our mechanism rests on the persistence of belief changes induced by transitory shocks. This stems from the martingale property of beliefs: conditional on time-t information ($\mathcal{I}_t$), the estimated distribution is a martingale: on average, the agent expects her future belief to be the same as her current beliefs. This property holds exactly if the bandwidth matrix is set to zero.[6]

More generally, the smoothing embedded in the kernel induces a deviation from the martingale property. Numerically, however, these deviations are minuscule, both for the example in this section and in our full model. In other words, the kernel density estimator with the

[5]The optimal bandwidth minimizes the expected squared error when the underlying density is normal. It is widely used and is the default option in MATLAB's *ksdensity* function.

[6]In this case, the kernel puts positive probability mass only on realizations seen before. In other words, an event that isn't exactly identical to one in the observed sample is assigned zero probability, even if there are other observations arbitrarily close to it in the sample. This is obviously too extreme a specification – since events are never identical in actual macro data, every observation will have zero probability before it occurs.

optimal bandwidth is, approximately, a martingale $\mathbb{E}_t [\hat{g}_{t+j} | \mathcal{I}_t] \approx \hat{g}_t$. As a result, any changes in beliefs induced by new information are, in expectation, permanent. This property, which also arises with parametric learning (Hansen and Sargent, 1999; Johannes et al., 2015), plays a central role in generating long-lived effects from transitory shocks.

We now illustrate how this mechanism works, using an illustrative univariate example. Since our goal is to illustrate the effects of outlier realizations, we need a data series with such outliers. We will use a series of shocks to capital 'quality', estimated from post-WWII US data (plotted in the first panel of Figure 3). For now, we treat this as an arbitrary series and postpone a detailed discussion of their economic interpretation and measurement to Section 3.

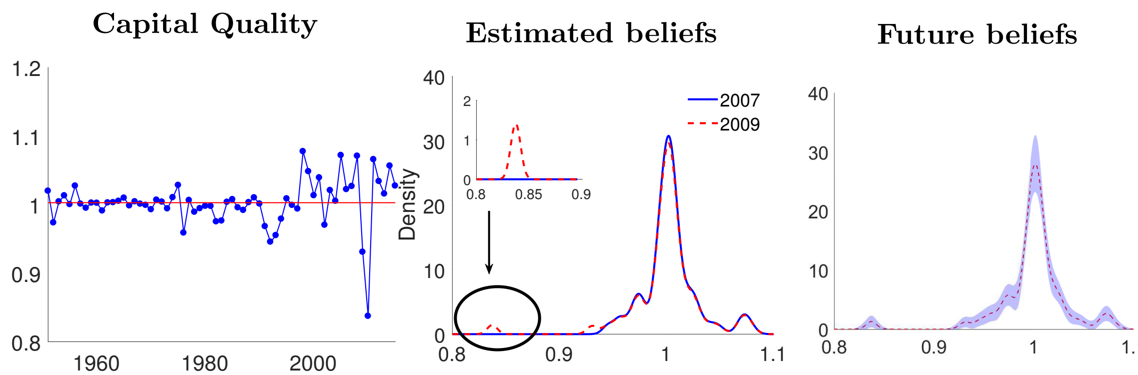


Figure 3: **Estimated beliefs.**
The first panel shows the realizations of "capital quality" shocks, defined later in the paper in (15) and measured as described in Section 3. The second panel shows the kernel density, estimated from data available up to 2007 and up to 2009. The change in the left tail represents the effect of the Great Recession. The third panel shows the average estimate of the probability density (along with a 2 standard deviation band) in 2039. This is computed by simulating data for the period 2010-2039, drawing future realizations from the estimated distribution in 2009 and estimating a kernel on each simulated series.

Estimated belief changes The second panel of Figure 3 takes all the data up to and including 2007 and shows the estimated probability distribution, based on that (pre-crisis) data.[7] Then it takes all data up to and including 2009 (post-crisis) to plot the new probability distribution estimate. The two adverse realizations in '08 and '09 lead to an increase in the assessment of tail risk: the 2009 distribution ($\hat{g}_{2009}$) shows a pronounced hump in the density around the 2008 and 2009 realizations, relative to the 2007 one ($\hat{g}_{2007}$). Crucially, even though these negative realizations were short-lived, this increase in left tail risk persists. To see how persistent beliefs are, we ask the following question: What would be the estimated probability distribution in 2039? To answer this question, we need to simulate future data. Since our best estimate of the distribution of future data in 2009 is $\hat{g}_{2009}$, we draw many 30-year sequences of

[7]From the 1950-2007 data, the optimal bandwidth for this univariate case is 0.0056.

future data from this $\hat{g}_{2009}$ distribution. After each sequence, we re-estimate the distribution g, using all available data. Obviously, each simulated path gives rise to a different estimated distribution, so we report the average across all those paths (as well as 2 standard deviation bands) in the third panel of Figure 3, which shows that the average (dashed line) is very close to the 2009 distribution. This Monte Carlo exercise illustrates how tail risk induced by financial crisis may never go away. Of course, in this simulation, we are drawing from the $\hat{g}_{2009}$ distribution, so every once in a long while, another crisis is drawn, which keeps the left tail "hump" from disappearing. If we instead drew future data from a distribution without tail risk (e.g. $\hat{g}_{2007}$), the hump would still be very persistent, but not permanent (see Section 4).

Thus, every new shock, even a transitory one, has a persistent effect on beliefs. This pattern is reminiscent of the evidence of heightened tail risk from asset markets and other proxies presented in the Introduction. In the rest of the paper, we will use a specific economic model, which maps shocks and beliefs into investment, hiring and production decisions, in order to assess the implications of these belief changes for macroeconomic outcomes. However, it is worth noting that our approach and mechanism have broader relevance as simple tools to generate endogenous persistence in many economic environments.

2 Economic Model

To explore whether our belief formation mechanism can help explain the persistence of the recent stagnation, we need to embed it in an economic environment. To have a shot at quantitatively explaining the recent episode, our model needs two key features. First, we need a shock structure that can capture extreme and unusual aspects of the 2008-'09 recession: namely, the unusually low returns to firms' (non-residential) capital and stress in the financial sector. To generate large fluctuations in returns, we use shocks to capital quality. These shocks, which scale up or down the effective capital stock, are not to be interpreted literally. A decline in quality captures the idea that a Las Vegas hotel built in 2007 may deliver less economic value after the financial crisis, e.g. because it is consistently half-empty. This would be reflected in a lower market value, a feature we will exploit later in our measurement strategy. This specification is not intended as a deep explanation of what triggered the financial crisis or the recession. Instead, it is a summary statistic that stands in for many possible explanations and allows the model to speak to both financial and macro data.[8] This agnostic approach to the causes of the crisis also puts the spotlight on our contribution – the ability of learning to generate persistent responses

[8]Capital quality shocks have been employed for a similar purpose in Gourio (2012), as well as in a number of recent papers on financial frictions, crises and the Great Recession (e.g., Gertler et al. (2010), Gertler and Karadi (2011), Brunnermeier and Sannikov (2014)). Their use in macroeconomics and finance, however, goes back at least to Merton (1973), who uses them to generate highly volatile asset returns.

to extreme events. Similarly, to capture stress in the financial sector, we adopt a tractable specification without taking a stand on the root causes – an aggregate financial shock, which directly induces default by financial intermediaries.

Second, we need a setting where economic activity is sensitive to the probability of extreme capital shocks. We use a version of the model in Gourio (2012, 2013), optimized for this purpose. Two key ingredients – namely, Epstein-Zin preferences and costly bankruptcy – combine to generate significant sensitivity to tail risk. Adding the assumption that labor is hired in advance with an uncontingent wage increases the effective leverage of firms and therefore, accentuates the sensitivity of investment and hiring decisions to tail risk. Similarly, preferences that shut down wealth effects on labor avoid a surge in hours in response to crises.

Thus, this combination of assumptions offers a laboratory to assess the quantitative potential of our belief revision mechanism. It is worth emphasizing that none of these ingredients guarantees persistence, our main focus. The capital quality shock has a direct effect on output upon impact but, absent belief revisions, does not change the long-run trajectory of the economy. Our formulation of the financial sector also rules out propagation through the financial system (other than those coming through beliefs). Finally, the non-linearity from preferences and debt influence the size of the economic response, but by themselves do not generate any internal propagation.

Persistence comes solely from our novel ingredient, belief formation and would arise even without these ingredients. We model beliefs using the non-parametric estimation described in the previous section and show how to discipline this procedure with observable macro data.

2.1 Setup

Preferences and technology: An infinite horizon, discrete time economy has a representative household, with preferences over consumption (C_t) and labor supply (L_t):

$$U_t = \left[(1-\beta)\left(C_t - \frac{L_t^{1+\gamma}}{1+\gamma}\right)^{1-\psi} + \beta E_t\left(U_{t+1}^{1-\eta}\right)^{\frac{1-\psi}{1-\eta}}\right]^{\frac{1}{1-\psi}} \tag{1}$$

where ψ is the inverse of the intertemporal elasticity of substitution, η indexes risk-aversion, γ is inversely related to the elasticity of labor supply, and β represents time preference.[9]

The economy is also populated by a unit measure of firms, indexed by i and owned by the representative household. Firms produce output with capital and labor, according to a standard Cobb-Douglas production function $k_{it}^{\alpha} l_{it}^{1-\alpha}$. Firms are subject to an aggregate shock to capital

[9]This utility function rules out wealth effects on labor, as in Greenwood et al. (1988). Appendix C.7 relaxes this assumption.

quality ϕ_t. A firm that enters the period t with capital $\hat{k}_{it}$ has effective capital $k_{it} = \phi_t \hat{k}_{it}$. These capital quality shocks are i.i.d. over time. The i.i.d. assumption is made in order to avoid an additional, exogenous, source of persistence.[10]

Firms are also subject to an idiosyncratic shock v_{it}. These shocks scale up and down the total resources available to each firm (before paying debt, equity or labor). Formally,

$$\Pi_{it} = v_{it} \left[k_{it}^{\alpha} l_{it}^{1-\alpha} + (1-\delta) k_{it} \right] \tag{2}$$

where δ is the rate of capital depreciation. The shocks v_{it} are i.i.d. across time and firms and are drawn from a known distribution, F.[11] The mean of the idiosyncratic shock is normalized to be one: $\int v_{it} \quad di = 1$. The primary role of these shocks is to induce an interior default rate in equilibrium, allowing a more realistic calibration, particularly of credit spreads.

Labor: We make two additional assumptions about labor markets. First, firms hire labor in advance, i.e. before observing the realizations of aggregate and idiosyncratic shocks. Second, wages are non-contingent – in other words, workers are promised a non-contingent payment and face default risk. These assumptions create an additional source of leverage.

Credit and default: Firms have access to a competitive non-contingent debt market, where lenders offer bond price (or equivalently, interest rate) schedules as a function of aggregate and idiosyncratic states, in the spirit of Eaton and Gersovitz (1981). A firm enters period $t+1$ with an obligation, b_{it+1} to bondholders and a promise of $w_{it+1} l_{it+1}$ to its workers. After workers exert labor effort, shocks are realized and the firm's shareholders decide whether to repay their obligations or default. Default is optimal for shareholders if, and only if,

$$\Pi_{it+1} - b_{it+1} - w_{it+1} l_{it+1} + \Gamma_{t+1} < 0$$

where Γ_{t+1} is the present value of continued operations. Thus, the default decision is a function of the resources available to the firm (Π_{it+1}) and the *total* obligations of the firm ($b_{it+1} \ + w_{it+1} l_{it+1} \ \equiv B_{it+1}$). Let $r_{it+1} \in \{0, 1\}$ denote the repayment policy of the firm.

In the event of default, equity holders get nothing. The productive resources of a defaulting firm are sold to a new firm at a discounted price, equal to a fraction $\theta < 1$ of the value of the

[10]The i.i.d. assumption also has empirical support. In the next section, we use macro data to construct a time series for ϕ_t. We estimate an autocorrelation of 0.15, statistically insignificant. In Appendix C.9, we show that this generates almost no persistence in the economic response.

[11]This is a natural assumption - with a continuum of firms and a stationary shock process, firms can learn the complete distribution of any idiosyncratic shocks after one period.

defaulting firm. The proceeds are distributed *pro-rata* among the bondholders and workers.[12]

Let q_{it} denote the bond price schedule faced by firm i in period t. The lenders pay q_{it} at time t in exchange for a promise of one unit of output at $t+1$. Debt is assumed to carry a tax advantage. A firm which issues b_{it+1} of debt at price q_{it}, receives a date-t payment of $\chi q_{it} b_{it+1}$, where $\chi > 1$. This effective subsidy to debt issuance, along with the cost of default, introduces a trade-off in the firm's capital structure decision, breaking the Modigliani-Miller theorem.[13]

For a firm that does not default, the dividend payout is its total available resources times output shock, minus its payments to debt and labor, minus the cost of building next period's capital stock (the undepreciated current capital stock is included in Π_{it}), plus the proceeds from issuing new debt, including its tax subsidy

$$d_{it} = \Pi_{it} - B_{it} - \hat{k}_{it+1} + \chi q_{it} b_{it+1}. \tag{3}$$

Importantly, we do not restrict dividends to be positive, with negative dividends interpreted as (costless) equity issuance. Thus, firms are not financially constrained, ruling out another potential channel of persistence.

Intermediaries: Credit is extended to firms by a continuum of competitive intermediaries, who live for 2 periods and have no resources of their own and so compete to raise money from households by issuing debt and equity claims. We model intermediary default with a simple formulation: with probability π_t, an intermediary fails to repay both its debt- and equity-holders. This can be interpreted in different ways, e.g. as stemming from shocks to loan portfolios or losses from other activities (e.g. derivatives or mortgages) or the possibility diversion of funds by the intermediaries. From our perspective, the exact micro-foundation is not crucial and so we directly treat the default probability π_t as a primitive financial shock. In the next section, we use data on bank failures to construct a time series for this variable. As with firm default, we assume that default by intermediaries does not destroy resources, so the money lost ultimately flows to the representative household. Appendix B.2 formally presents the problem of intermediaries.

The two aggregate shocks – the capital quality shock, ϕ_t and financial shock, π_t – are assumed to be iid over time, but correlated with each other in an arbitrary fashion. Formally, in each period, (ϕ, π) is an iid draw from a joint distribution $g(\cdot)$.

[12]Default does not destroy resources - the penalty is purely private. This is not crucial - it is easy to relax this assumption and assume that all or part of the penalty represents physical destruction of resources.

[13]The subsidy is assumed to be paid by a government that finances it through lump-sum taxes.

Timing and value functions:

1. Firms enter t with capital $\hat{k}_{it}$, labor l_{it}, outstanding debt b_{it}, and a wage obligation $w_{it}l_{it}$.

2. All shocks – the aggregate shocks (ϕ_t, π_t) and the firm-specific profit shock v_{it} – are realized. Production takes place.

3. The firm decides whether to default ($r_{it} = 0$) or repay ($r_{it} = 1$) its bond and labor claims. The debt- and equity-holders of each intermediary are repaid with probability π_t.

4. The firm makes capital $\hat{k}_{it+1}$, debt b_{it+1} and employment l_{it+1} choices for the following period, along with a wage contract w_{it+1}. Workers commit to next-period labor supply l_{it+1}. Note that all these choices are made concurrently.

In recursive form, the problem of the firm is

$$V\left(\Pi_{it}, B_{it}, S_t\right) = \max\left[0, \max_{d_{it}, \hat{k}_{it+1}, b_{it+1}, w_{it+1}, l_{it+1}} d_{it} + \mathbb{E}_t M_{t+1} V\left(\Pi_{it+1}, B_{it+1}, S_{t+1}\right)\right] \tag{4}$$

subject to

$$\begin{aligned}
\text{Dividends:} \quad & d_{it} \;\leq\; \Pi_{it} - B_{it} - \hat{k}_{it+1} + \chi q_{it} b_{it+1} && (5)\\
\text{Discounted wages:} \quad & \mathcal{W}_t \;\leq\; w_{it+1} q_{it} && (6)\\
\text{Future obligations:} \quad & B_{it+1} = b_{it+1} + w_{it+1} l_{it+1} && (7)\\
\text{Resources:} \quad & \Pi_{it+1} = v_{it+1}\left[(\phi_{t+1}\hat{k}_{it+1})^{\alpha} l_{it+1}^{1-\alpha} + (1-\delta)\phi_{t+1}\hat{k}_{it+1}\right] && (8)\\
\text{Bond price:} \quad & q_{it} = \mathbb{E}_t M_{t+1}(1-\pi_{t+1})\left[r_{it+1} + (1-r_{it+1})\frac{\theta V\left(\Pi_{it+1}, 0, S_{t+1}\right)}{B_{it+1}}\right] && (9)
\end{aligned}$$

The first max operator in (4) captures the firm's option to default. The expectation $\mathbb{E}_t$ is taken over the idiosyncratic and aggregate shocks, given beliefs about the aggregate shock distribution.

In (6), the firm's wage promise w_{it+1} is multiplied by the bond price q_{it}, since workers are effectively paid in bonds and are subject to the risk of default.[14] Equation (6) requires the value of this promise be at least as large as $\mathcal{W}_t$, the representative household's marginal rate of substitution. Both $\mathcal{W}_t$ and the stochastic discount factor M_{t+1} are defined using the household's

[14] Note that this implies that workers' claims are also subject to the risk of intermediary default. For example, under the diversion interpretation, workers also stand to lose if the intermediary manages to successfully divert funds. This assumption is made only to simplify the algebra and does not have a material effect on our results.

utility function:

$$\mathcal{W}_t = \left(\frac{dU_t}{dC_t}\right)^{-1} \frac{dU_t}{dL_{t+1}} \qquad M_{t+1} = \left(\frac{dU_t}{dC_t}\right)^{-1} \frac{dU_t}{dC_{t+1}} \tag{10}$$

Equation (9), derived in Appendix B.2, shows that the equilibrium bond price is a function of the expected repayment (the term inside the square brackets) as well as the risk of intermediary default π_{t+1}. The term $V(\Pi_{it+1}, 0, S_{t+1})$ denotes the value of a defaulting firm's assets.

The aggregate state S_t consists of $(\Pi_t, L_t, \mathcal{I}_t)$ where $\Pi_t \equiv AK_t^{\alpha} L_t^{1-\alpha} + (1-\delta)K_t$ is the aggregate resources available, L_t is aggregate labor input (chosen in $t-1$) and $\mathcal{I}_t$ is the economy-wide information set. Equation (9) reveals that bond prices are a function of the firm's capital $\hat{k}_{it+1}$, labor l_{it+1} and debt B_{it+1}, as well as the aggregate state S_t. The firm takes the aggregate state and the function $q_{it} = q\left(\hat{k}_{it+1}, l_{it+1}, B_{it+1}, S_t\right)$ as given, recognizing that its capital, labor and leverage choices affect its bond price.

Information and beliefs The set $\mathcal{I}_t$ includes the history of all shocks (ϕ_t, π_t) observed up to and including time-t. For now, we specify a general function, Ψ, which maps $\mathcal{I}_t$ into an appropriate probability space. The expectation operator $\mathbb{E}_t$ is defined with respect to this space. In the next section, we use the kernel density estimation procedure from section 1 to fully characterize Ψ.

Equilibrium Definition. For a given belief function Ψ, a recursive equilibrium is a set of functions for (i) aggregate consumption and labor that maximize (1) subject to a budget constraint, (ii) firm value and policies that solve $(4-8)$, taking as given the bond price function (9) and the stochastic discount factor and aggregate MRS functions in (10) and are such that (iii) aggregate consumption and labor are consistent with individual choices.

2.2 Solving the Model

Here, we show the key equations characterizing the equilibrium, relegating detailed derivations to Appendix B.1. First, use the binding dividend and wage constraints (5) and (6) to substitute out for d_{it} and w_{it} in (4). This leaves 3 choice variables $(\hat{k}_{it+1}, l_{it+1}, b_{it+1})$ and a repayment decision. The latter is characterized by a threshold rule in the idiosyncratic shock v_{it}:

$$r_{it} = \begin{cases} 0 & \text{if } v_{it} < \underline{v}_{it} \\ 1 & \text{if } v_{it} \geq \underline{v}_{it} \end{cases}$$

It turns out to be more convenient to recast the problem as a choice of $\hat{k}_{it+1}$, leverage, $lev_{it+1} \equiv \frac{B_{it+1}}{\hat{k}_{it+1}}$, and the labor-capital ratio, $\frac{l_{it+1}}{\hat{k}_{it+1}}$. Since all firms make symmetric choices, we can suppress the i subscript: $\hat{k}_{it+1} = \hat{K}_{t+1}$, $l_{it+1} = L_{t+1}$, $lev_{it+1} = lev_{t+1}$, $\underline{v}_{it+1} = \underline{v}_{t+1}$. The optimality condition for $\hat{K}_{t+1}$ can be written as:

$$\begin{aligned} 1 + \chi \mathcal{W}_t \frac{L_{t+1}}{\hat{K}_{t+1}} = \mathbb{E}[M_{t+1} R^k_{t+1}] + (\chi - 1) lev_{t+1} q_t - (1-\theta)\mathbb{E}[M_{t+1} R^k_{t+1} h(\underline{v}_{t+1})] \\ - \mathbb{E}[M_{t+1} R^k_{t+1} \pi_{t+1} (\underline{v}_{t+1}(1 - F(\underline{v}_{t+1})) + \theta h(\underline{v}_{t+1}))] \end{aligned} \tag{11}$$

$$\text{where} \qquad R^k_{t+1} = \frac{\phi^\alpha_{t+1} \hat{K}^\alpha_{t+1} L^{1-\alpha}_{t+1} + (1-\delta)\,\phi_{t+1} \hat{K}_{t+1}}{\hat{K}_{t+1}} \tag{12}$$

The term R^k_{t+1} is the average *ex-post* per-unit, pre-wage return on capital, while $h(\underline{v}) \equiv \int_{-\infty}^{\underline{v}} v f(v) dv$ is the expected value of the idiosyncratic shock in the default states.

The first term on the right hand side of (11) is the usual expected direct return from investing, weighted by the stochastic discount factor. The other terms are all related to debt. The second term reflects the tax advantage of debt – the firm raises $lev_{t+1} q_t$ (per unit of capital) from the bond market, on which it earns a subsidy of $\chi - 1$. The third term captures default-related costs, equal to a fraction $1 - \theta$ of available resources. The final term reflects the effect of intermediary default (it disappears if $\pi_{t+1} = 0$ w.p. 1).

The optimal labor choice equates the expected marginal cost of labor, $\mathcal{W}_t$, with its expected marginal product, adjusted for the effect of additional wage promises on the cost of default:

$$\chi \mathcal{W}_t = \mathbb{E}_t \left[M_{t+1} (1-\alpha)\, \phi^\alpha_{t+1} \left(\frac{\hat{K}_{t+1}}{L_{t+1}} \right)^\alpha (J^l(\underline{v}_{t+1})(1 - \pi_{t+1}) + \pi_{t+1}(1 - h(\underline{v}_{t+1}))) \right] \tag{13}$$

where $J^l(\underline{v}) = 1 + h(\underline{v})(\theta\chi - 1) - \underline{v}^2 f(\underline{v})\, \chi (\theta - 1)$ adjusts the marginal product of labor for the fact that labor is chosen in advance in exchange for a debt-like promise. Finally, the choice of leverage is governed by:

$$\begin{aligned} \mathbb{E}_t M_{t+1} \left[(1-\theta) \underline{v}_{t+1} f(\underline{v}_{t+1}) \right] = \left(\frac{\chi - 1}{\chi} \right) \mathbb{E}_t \left[M_{t+1} \left(1 - F(\underline{v}_{t+1}) \right) \right] \\ - \mathbb{E}_t \left[M_{t+1} \pi_{t+1} (1 - F(\underline{v}_{t+1}) - (1-\theta) \underline{v}_{t+1} f(\underline{v}_{t+1})) \right]. \end{aligned} \tag{14}$$

The left hand side is the marginal cost of increasing leverage. Higher leverage shifts the default threshold $\underline{v}$, raising the expected losses from the default penalty (a fraction $1 - \theta$ of the firm's value). The right hand side is the net marginal benefit – higher leverage brings in more subsidy (the tax benefit times the value of debt issued) but entails paying intermediary default premium.

The three firm optimality conditions, (11), (13), and (14), along with those from the household side (10) and the economy-wide resource constraint, characterize the equilibrium.

3 Measurement, Calibration and Solution Method

This section describes how we use macro data to estimate beliefs and parameterize the model, as well as our computational approach. One of the key strengths of our theory is that we can use observable data to estimate beliefs at each date.

Measuring capital quality shocks Recall from Section 1 that the Great Recession saw unusually low returns to non-residential capital, stemming from unusually large declines in the market value of capital. To capture this, we need to map the model's aggregate shock, namely the capital quality shock, into market value changes. A helpful feature of capital quality shocks is that their mapping to available data is straightforward. A unit of capital installed in period $t-1$ (i.e. as part of $\hat{K}_t$) is, in effective terms, worth ϕ_t units of consumption goods in period t. Thus, the change in its market value from $t-1$ to t is simply ϕ_t.

We apply this measurement strategy to annual data on non-residential capital held by US corporates. Specifically, we use two time series Non-residential assets from the Flow of Funds, one evaluated at market value and the second, at historical cost.[15] We denote the two series by NFA_t^{MV} and NFA_t^{HC} respectively. To see how these two series yield a time series for ϕ_t, note that, in line with the reasoning above, NFA_t^{MV} maps directly to effective capital in the model. Formally, letting P_t^k the nominal price of capital goods in t, we have $P_t^k K_t = NFA_t^{MV}$. Investment X_t can be recovered from the historical series, $P_{t-1}^k X_t = NFA_t^{HC} - (1-\delta)\, NFA_{t-1}^{HC}$. Combining, we can construct a series for $P_{t-1}^k \hat{K}_t$:

$$\begin{aligned} P_{t-1}^k \hat{K}_t &= (1-\delta) P_{t-1}^k K_{t-1} + P_{t-1}^k X_t \\ &= (1-\delta) NFA_{t-1}^{MV} + NFA_t^{HC} - (1-\delta)\, NFA_{t-1}^{HC} \end{aligned}$$

Finally, in order to obtain $\phi_t = \frac{K_t}{\hat{K}_t}$, we need to control for nominal price changes. To do this, we proxy changes in P_t^k using the price index for non-residential investment from the National

[15] These are series FL102010005 and FL102010115 from Flow of Funds. See Appendix D.3.

Income and Product Accounts (denoted $PINDX_t$).[16] This yields:

$$\begin{aligned}
\phi_t = \frac{K_t}{\hat{K}_t} &= \left(\frac{P_t^k K_t}{P_{t-1}^k \hat{K}_t}\right)\left(\frac{PINDX_{t-1}^k}{PINDX_t^k}\right) \\
&= \left[\frac{NFA_t^{MV}}{(1-\delta)NFA_{t-1}^{MV} + NFA_t^{HC} - (1-\delta)\,NFA_{t-1}^{HC}}\right]\left(\frac{PINDX_{t-1}^k}{PINDX_t^k}\right) \qquad (15)
\end{aligned}$$

Using the measurement equation (15), we construct an annual time series for capital quality shocks for the US economy since 1950. The left panel of Figure 3 plots the resulting series. The mean and standard deviation of the series over the entire sample are 1 and 0.03 respectively. The autocorrelation is statistically insignificant at 0.15.

As Figure 3 shows, for most of the sample period, the shock realizations are in a relatively tight range around 1. However, we saw two large adverse realizations during the Great Recession: 0.93 in 2008 and 0.84 in 2009. These reflect the large drops in the market value of non-residential capital stock – in 2009, for example, the aggregate value of that stock fell by about 16%. What underlies these large fluctuations? The main contributor was a fall in the value of commercial real estate (which is the largest component of non-residential assets).[17] Through the lens of the model, these movements are mapped to a change in the economic value of capital – in the spirit of the hypothetical example of the Las Vegas hotel at the beginning of Section 2 whose market value falls.

Measuring financial shocks Recall that the financial shock, π_t, denotes the fraction of intermediary assets diverted or otherwise lost. To construct a proxy for the financial shock, π_t, we use data on bank failures from the Federal Deposit Insurance Corporation and compute the fraction of total bank assets held by institutions which were either taken over by or otherwise obtained assistance from the FDIC. Applying an average loss rate[18] of 30%, yields our proxy for π_t, which is plotted in the right panel of Figure 3. It shows an unusually large spike during 2008-'09, reflecting the extreme nature of the recent financial crisis.

Belief Estimation We then apply our kernel density estimation procedure to these two time series and construct a sequence of beliefs. In other words, for each t, we construct $\{\hat{g}_t\}$ using

[16]Our results are robust to alternative measures of nominal price changes, e.g. computed from the price index for GDP or Personal Consumption Expenditure, see Appendix C.1.

[17]One potential concern is that the fluctuations in the value of real estate stem mostly from land price movements. While the data in the Flow of Funds do not allow us to directly control for changes in the market value of land, they do suggest a limited role for land. Measured at historical cost, land accounts for less than 5% of total non-residential capital. The observed fluctuations in the value of these assets during 2008-09 are simply too large to be accounted for by land price movements, even if they are sizable.

[18]This is consistent with the estimates in James (1991) and Bennett and Unal (2015).

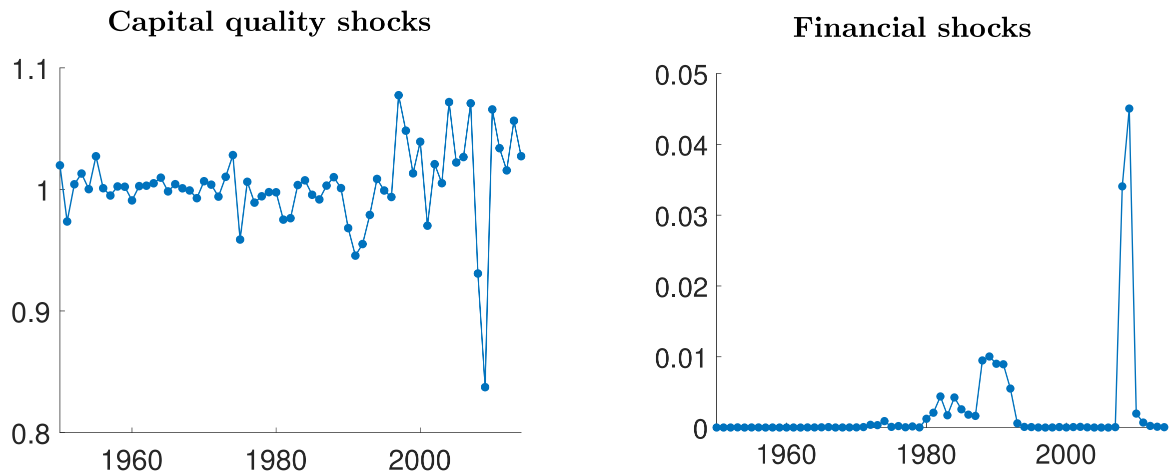


Figure 4: **Data: Capital quality and financial shocks.**

Note: The first panel shows the realizations of "capital quality" shocks and the second panel shows the realizations of financial shocks. In 2008 and 2009 we observe tail realization in both series.

the available time series until that point. Figure 5 reveals the effect of the extreme realizations in 2008 and 2009. The first panel plots the marginal probability distribution of ϕ_t for two dates – 2007 and 2009. They show that the Great Recession significantly increased perceived tail risk. The estimated probabilities for 2007 implies almost zero mass below 0.90, while the one for 2009 attaches a non-trivial (approximately 2.5%) likelihood to this region of the state space. The second panel shows a similar pattern for the financial shock, π_t – the likelihood of large losses is much higher under $\hat{g}_{2009}$. Finally, since these extreme realizations were correlated, the increases appear as concentrated spikes in the joint distribution. This is reflected in third panel, which plots the difference between the probabilities implied by $\hat{g}_{2009}$ and $\hat{g}_{2007}$.

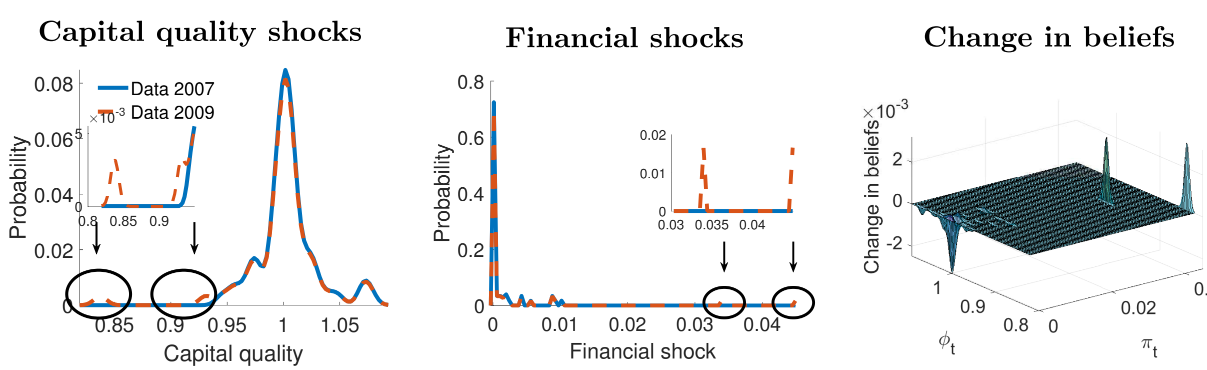


Figure 5: **Change in beliefs due to the Great Recession.**

Note: The first panel shows the probability distribution for capital quality shock, ϕ_t under $\hat{g}_{2007}$ and $\hat{g}_{2009}$ and the second panel for the financial shock, π_t. The third panel plots the change in the joint distribution, $\hat{g}_{2009} - \hat{g}_{2007}$.

Calibration A period is interpreted as a year. We choose the discount factor β and depreciation δ to target a steady state capital-output ratio of 3.5 (this is taken from Cooley and Prescott (1995)) and an investment-output ratio of 0.12 (this is the average ratio of non-residential investment to output during 1950-2007 from NIPA accounts).[19] The share of capital in the production, α, is 0.40, which is also taken from Cooley and Prescott (1995). The recovery rate upon default, θ, is set to 0.70, following Gourio (2013). The distribution for the idiosyncratic shocks, v_{it} is assumed to be lognormal, i.e. $\ln v_{it} \sim N\left(-\frac{\hat{\sigma}^2}{2}, \hat{\sigma}^2\right)$ with $\hat{\sigma}^2$ chosen to target a default rate of 0.02.[20] The labor supply parameter, γ, is set to 0.5, in line with Midrigan and Philippon (2011), corresponding to a Frisch elasticity of 2.

For the parameters governing risk aversion and intertemporal elasticity of substitution, we use standard values from the asset pricing literature and set $\psi = 0.5$ (or equivalently, an IES of 2) and $\eta = 10$.[21] The tax advantage parameter χ is chosen to match a leverage target of 0.70, which is obtained by adding the wage bill (approximately 0.2 of the steady state capital stock) to financial leverage (the ratio of external debt to capital, about 0.5 in US data - from Gourio (2013)). Table 1 summarizes the resulting parameter choices.

Parameter	Value	Description
Preferences:		
β	0.91	Discount factor
η	10	Risk aversion
ψ	0.50	1/Intertemporal elasticity of substitution
γ	0.50	1/Frisch elasticity
Technology:		
α	0.40	Capital share
δ	0.03	Depreciation rate
$\hat{\sigma}$	0.25	Idiosyncratic volatility
Debt:		
χ	1.06	Tax advantage of debt
θ	0.70	Recovery rate

Table 1: **Parameters**

Numerical solution method Given the importance of curvature in policy functions for our results, we solve the non-linear system of equations (11) – (14) using collocation methods.

[19] This yields $\beta = 0.91$ and $\delta = 0.03$, which are lower than other estimates in the literature. However, an alternative calibration strategy with $\delta = 0.06$ (consistent with reported depreciation rates in the Flow of Funds data) and $\beta = 0.95$ (which leads to the same capital-output ratio) generates almost identical results.

[20] This is in line with the target in Khan et al. (2014), though a bit higher than the one in Gourio (2013). We verified that our quantitative results are not sensitive to this target.

[21] Appendix C.6 examines the robustness of our main results to these parameter choices. See also the discussion in Gourio (2013).

Appendix A describes the iterative procedure. In order to maintain tractability, we need to make one approximation. Policy functions at date-t depend both on the current estimated distribution, $\hat{g}_t(\phi, \pi)$, and the distribution $\mathcal{H}$ over next-period estimates, $\hat{g}_{t+1}(\phi, \pi)$. Keeping track of $\mathcal{H}(\hat{g}_{t+1}(\phi, \pi))$, (a distribution over a distribution, i.e. a compound lottery) as a state variable would render the analysis intractable. However, the approximate martingale property of $\hat{g}_t$ discussed in Section 1 offers an accurate and computationally efficient approximation to this problem. The martingale property implies that the average of the compound lottery is $\mathbb{E}_t[\hat{g}_{t+1}(\phi, \pi)] \approx \hat{g}_t(\phi, \pi)$, $\forall(\phi, \pi)$. Therefore, when computing policy functions, we approximate $\mathcal{H}(\hat{g}_{t+1}(\phi, \pi))$ with its mean $\hat{g}_t(\phi, \pi)$, the current estimate of the distribution. Appendix C.2 uses a numerical experiment to show that this approximation is quite accurate. Intuitively, future estimates $\hat{g}_{t+1}$ are tightly centered around $\hat{g}_t$, i.e. $\mathcal{H}(\hat{g}_{t+1})$ has a relatively small variance. This can also be seen from the illustrative example in Section 1: as Figure 3 shows, even 30 years out, beliefs are tightly clustered around the mean belief. For 1-10 quarters ahead, where most of the utility weight is, this error is even smaller.

4 Main Results

In this section, we evaluate, quantitatively, the ability of the model of generate persistent responses from tail events and confront its predictions with data. The key model feature behind persistence is the learning mechanism. To isolate its role, we compare results from our model to those from the same model where the distribution of shocks is assumed to be known with certainty. In this "no learning" economy, agents know the true probability of the tail event and so, observing such a realization does not change their beliefs. Next, we demonstrate how this mechanism makes large, unusual recessions different from smaller, more normal ones by comparing the model's predictions for the response to the Great Recession to a counterfactual, much less extreme shock. Then, we explore an economy where agents have learned from earlier episodes such as the Great Depression. It shows that beliefs about tail risk are particularly persistent, not because tail events were never seen before, but because relevant data on tail events is observed infrequently. Finally, we show that incorporating learning delivers more realistic equity, bond and option price predictions.

4.1 Belief Updating and Persistence

Our first set of results compare the predictions of the learning and no-learning models for macro aggregates (GDP, investment and labor) since 2008-'09. They show that the model with learning does significantly better in terms of matching the observed, persistent behavior of macro variables. Then, to rule out the possibility that persistence comes primarily from the occurrence of future crises, we show that the economic responses are extremely persistent, even if no future crises occur.

To compute our benchmark results, we begin by estimating $\hat{g}_{2007}$ using the data on (ϕ_t, π_t) from 1950-2007. We then compute the stochastic steady state by simulating the model for 1000 periods drawing from the estimated $\hat{g}_{2007}$. We discard the first 500 observations and time-average aggregate variables across the remaining periods. This corresponds to the long-run average value of the variable under the assumption that the true data generating process is $\hat{g}_{2007}$[22]. This steady state is the starting point for our analysis: we are interested in changes relative to this level. We then feed in the measured shocks from 2008 through 2014 and re-estimate the distribution to obtain $\hat{g}_{2014}$. To see how persistent the responses are, we need to simulate future time paths, which in turn requires an assumption about the distribution from which future shocks will be drawn. Given all the data available to us, our best estimate for this distribution is $\hat{g}_{2014}$. Therefore, we simulate future paths by drawing shock sequences from $\hat{g}_{2014}$ and compute the mean for various aggregate variables across these paths.

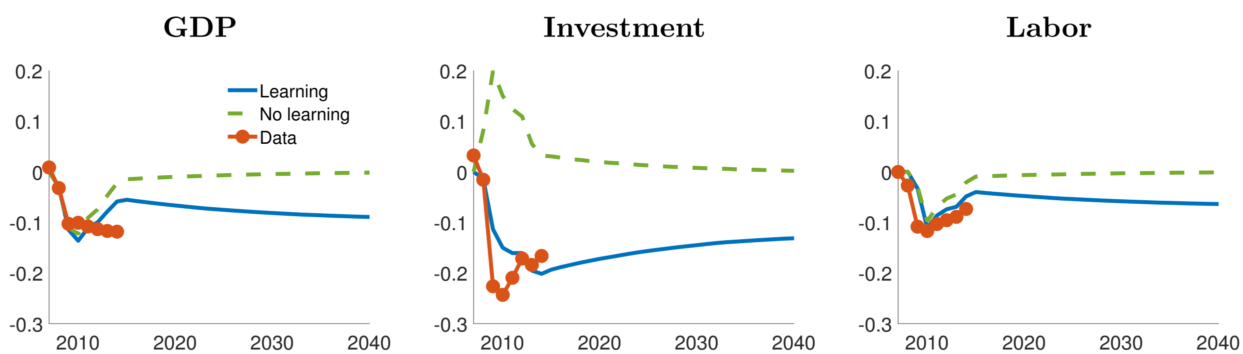


Figure 6: **Learning leads to persistent effects on output, investment and labor.** *Solid line shows the change in aggregates (relative to the stochastic steady state associated with $\hat{g}_{2007}$). The circles show de-trended US data for the period 2008-2014. For the dashed line (no learning), agents believe that shocks are drawn from $\hat{g}_{2014}$ and never revise those beliefs.*

The blue solid line in Figure 6 plots these averages for output, investment and employment as log deviations from their steady state values under $\hat{g}_{2007}$. It shows a pattern of prolonged stagnation, where the economy (on average) never fully recovers from the negative shocks in 2008-'09. and instead moves towards a new, lower (stochastic) steady state. These results do not imply that stagnation will necessarily continue forever: they show that, from the perspective of an agent with the current information set, recovery is not expected.[23]

[22]Under this assumption, the long-run behavior of the economy is described by an ergodic distribution.

[23]Appendix C.12 shows that a large fraction of this persistent response is due to changes in beliefs rather than the direct impact of the ϕ_t shock on capital.

The solid red line with circles in Figure 6 plots the actual data (in deviations from their respective 1950-2007 trends) for the US economy.[24] As the graph shows, the model's predictions line up reasonably well with the data, even though none of the series were used in the calibration or measurement. The predicted path for employment lags and slightly under-predicts the actual changes, largely due to the assumption that labor is chosen in advance. The predicted drop in investment is also slightly lower than what was observed.[25]

Table 2 summarizes the long-run effects of the belief changes, by comparing stochastic steady states under $\hat{g}_{2007}$ and $\hat{g}_{2014}$. As mentioned earlier, these are the average levels that the economy ultimately converges to, under the assumption that the data-generating process (and therefore, long-run beliefs) is $\hat{g}_{2007}$ or $\hat{g}_{2014}$. Capital and labor are, on average, 15% and 7.5% lower under the post-crisis beliefs. This translates into 11% lower output and 12% fall in investment.[26] Thus, even though the shocks experienced during the Great Recession were transitory, the resulting changes in beliefs persistently reduce economic activity.

	Stochastic steady state levels		**Change**
	$\hat{g}_{2007}$	$\hat{g}_{2014}$	
Output	6.2	5.5	-10.6%
Capital	26.3	22.3	-15.0%
Investment	0.7	0.6	-12.0%
Labor	2.3	2.2	-7.5%
Consumption	5.5	4.9	-10.4%

Table 2: **Belief changes lead to significant reductions in economic activity.**
Columns marked $\hat{g}_{2007}$ and $\hat{g}_{2014}$ represent average levels in the stochastic steady state of a model where shocks are drawn from $\hat{g}_{2007}$ or $\hat{g}_{2014}$ distributions respectively.

Turning off belief updating To demonstrate the role of learning, Figure 6 also plots simulated outcomes (dashed green line) from an otherwise identical economy where agents know the final distribution $\hat{g}_{2014}$ from the very beginning. This corresponds to a standard rational expectations econometrics approach, where agents are assumed to know the true distribution of shocks and the econometrician estimates this distribution using all the available data. Now, by assumption, agents do not revise their beliefs after the Great Recession. The post-2009 paths in this case are simulated as follows: the economy is assumed to be at its stochastic steady state

[24]Data on output and labor input are from Fernald (2014). The investment series is non-residential investment from the NIPA published by the Bureau of Economic Analysis, adjusted for population and price changes. Each series is de-trended using a log-linear trend estimated using data from 1950-2007, see Appendix D.4.

[25]Additional outcomes are reported in Appendix C.4.

[26]The investment drop is slightly smaller than that of capital because the shock distribution has also changed. For example, under $\hat{g}_{2014}$, the mean shock is slightly lower (relative to $\hat{g}_{2007}$). Intuitively, this acts like a slightly higher depreciation rate – so, even though capital is lower by 15%, the drop in investment is only 12%.

in 2007 and is subjected to the same sequence of shocks – the actual realizations in 2008-14 and subsequently, sequences of shocks drawn from $\hat{g}_{2014}$.

Without belief revisions, the negative capital quality shocks during 2008 and '09 spark an investment boom, as the economy replenishes the lost capital. While the curvature in utility moderates the speed of this transition, the overall pattern of a steady recovery back to the original steady state is clear.[27] In contrast, with learning, agents revise higher the perceived risk of intermediary default, which raises firm borrowing costs and dampens investment. This effect is amplified because this risk covaries with the risk of adverse capital quality shocks. This shows that learning is key to generating persistent reductions in economic activity.

What if shocks are persistent? An alternative explanation for the prolonged stagnation is that the shocks themselves were persistent. In Appendix C.9, we show that allowing for a realistic amount of persistence in the ϕ_t shocks does not materially change the dynamics of aggregate variables. This is because the observed autocorrelation of the ϕ_t process is too low to generate any meaningful persistence.

What if there are no more crises? In the results presented above, we put ourselves on the same footing as the agents in our model and draw future time paths of shocks using the updated beliefs $\hat{g}_{2014}$. One potential concern is that persistent stagnation comes not from belief changes *per se* but from the fact that along future paths, crises occur with non-trivial probability. This concern is not without merit. If future shocks were drawn instead from $\hat{g}_{2007}$, where the probability of a crisis is near zero, beliefs are no longer martingales: they change by the same amount on impact, but then converge back to their pre-crisis levels. Without the permanent effect on beliefs, persistence should fall.

However, as Figure 7 shows, persistence in aggregate variables over a 30 year horizon turns out to be almost the same with and without future crises (solid and dashed lines respectively). This is because beliefs about tail probabilities tend to be extremely persistent because tail-relevant data arrives infrequently. In other words, it takes many, many no-crisis draws to undo the initial upward shift in tail probability.

The fact that most data is not relevant for inferring tail probabilities is a consequence of our non-parametric approach. If instead, we imposed a parametric form like a normal distribution, then probabilities (including those for tail events) would depend only on the mean and variance of the distribution. Since mean and variance are informed by all data, tail probability revisions are frequent and small. As a result, the effects of observing the '08 and '09 shocks are more

[27]Since the no-learning economy has the same end-of-sample beliefs as the learning model, they both ultimately converge to the same *levels*, even though they start at different points (normalized to 0 for each series).

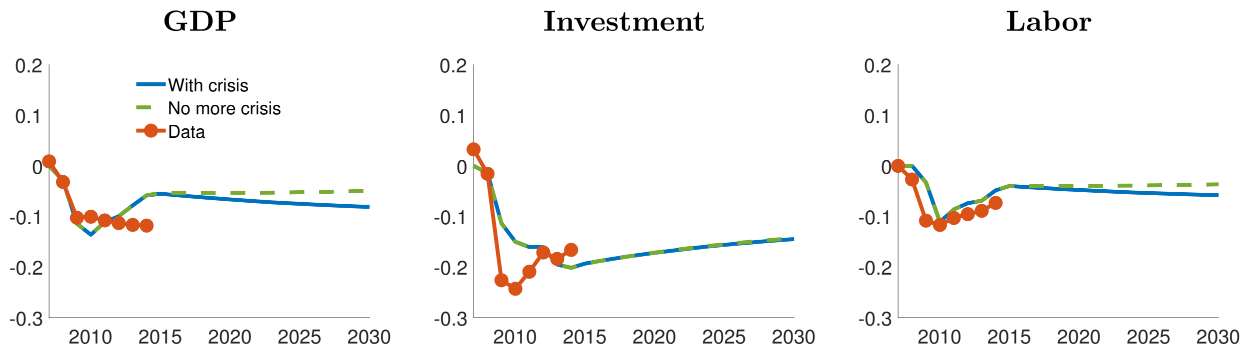


Figure 7: **What if there are no more crises?**
Solid (With crisis) line shows the change in aggregates when the data generating process is $\hat{g}_{2014}$ and agent updates beliefs. Dashed line (No more crisis) is an identical model in which future shocks are drawn from $\hat{g}_{2007}$. The circles show de-trended US data for the period 2008-2014.

transitory. See Appendix C.10 for more details.

4.2 Shock Size and Persistence

The secular stagnation puzzle is about why this recession had more persistent effects than others. Assuming exogenously persistent shocks does not answer this question, since that would imply that all downturns are equally persistent. Our model provides an explanation for long-lived responses to unusually large adverse shocks. Of course, in our setting, every negative shock to capital quality has both a transitory direct effect (in our setting, it lowers effective capital) and a persistent effect through beliefs. The extent to which a shock generates persistent outcomes depends on the relative size of these two effects. Observing a relatively unexpected tail event changes beliefs a lot and therefore, generates a large persistent effect. A small shock, on the other hand, has a negligible effect on beliefs and therefore, generates little persistence. This finding – that learning does not matter when 'normal' shocks hit – is also why we focus on the Great Recession. With other, more normal cycles, versions of the model with and without learning would be almost observationally equivalent, yielding little insight into the role of learning.

Figure 8 compares the effects on beliefs and output for small and large adverse shocks to capital quality (1 and 5 standard deviations below the mean respectively[28]), again starting from the stochastic steady state associated with $\hat{g}_{2007}$. Obviously, the output effects are smaller for the smaller shock (the bottom left panel), but more importantly, they are also more transitory and nearly the same with or without learning. Persistence comes almost entirely from the gradual replenishment of capital, which is present even without learning. Belief changes are

[28]These correspond to the realizations observed in 2001 and 2009 respectively.

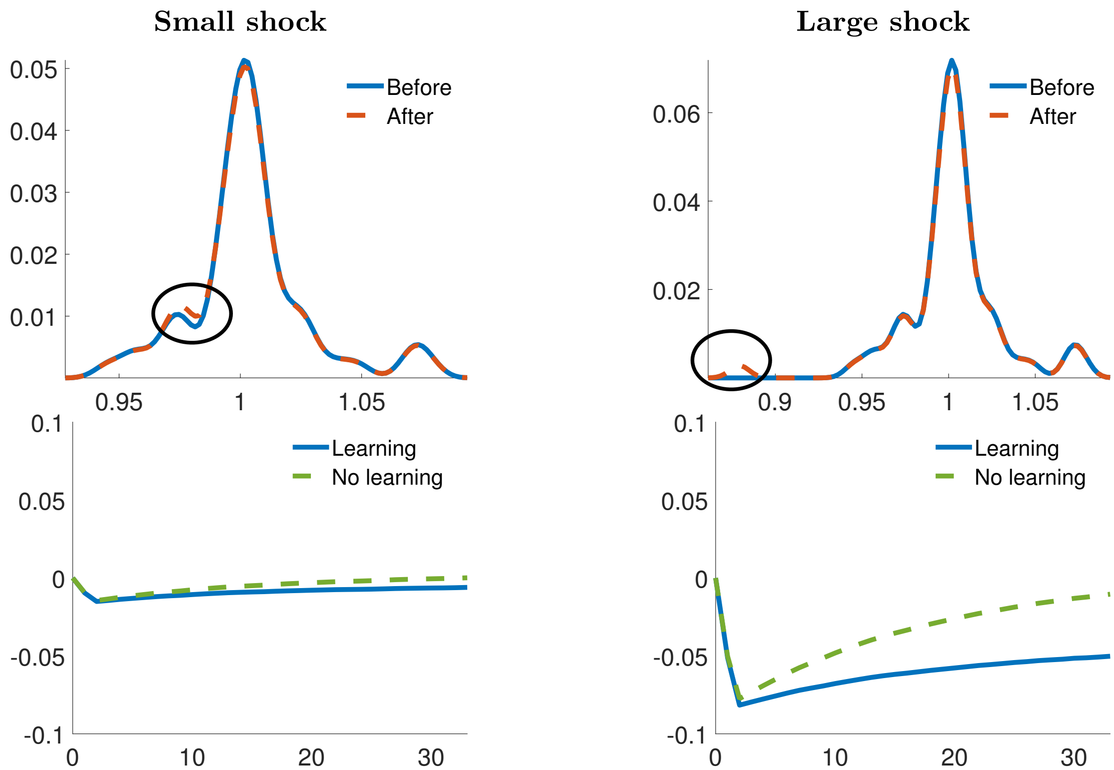


Figure 8: **Small shocks create negligible persistence.**
The top figures show the estimated probability for capital quality shocks before the shock (solid blue) and after a shock (dashed red). The bottom figures show the response of output to the shock under learning and no learning. The left (right) figure consider a small (large) shock of one (five) standard deviation.

still long-lived, but quantitatively, their effects on economic activity are small. In contrast, a large shock induces an economic response that is both more clearly distinguishable from the no-learning model and is more persistent (the bottom right panel).

This is because beliefs do not change much after a 1 standard deviation shock, as the top left panel of Figure 8 shows (there is a small increase in the distribution around 0.97). The increase after a larger shock (the top right panel) is much more pronounced. This differential behavior of beliefs has more to do with the *likelihood* than about the size of the shock *per se*. If, on the other hand, large shocks were generally more frequent, then observing a small shocks would be relatively more surprising and lead to larger belief changes. In other words, learning offers a novel explanation for why fluctuations triggered by rare events are particularly persistent.

4.3 Longer data sample and the Great Depression

Since our data sample starts in 1950, the Great Depression is not in our agents' information set. This raises the question: How would access to more data, with large adverse shocks in it, affect the response to the recent financial crisis? In the limit, as data accumulates, agents know the true distribution; new data ceases to affect beliefs. However, this convergence occurs more slowly for tail events converge than elsewhere in the distribution, precisely because they are infrequent. In this section, we perform an experiment to show that, even if agents had access to twice as much data, the 2008-09 experience continues to exert a large, persistent effect on economic activity.

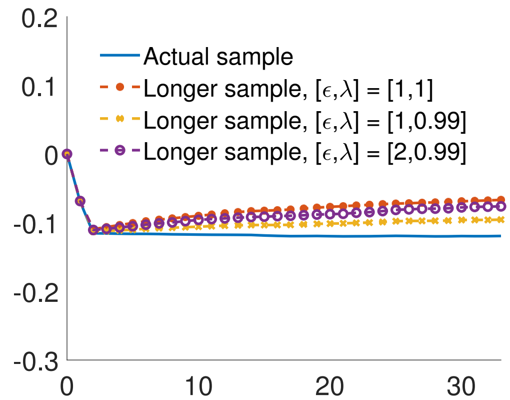


Figure 9: **Longer data sample tempers persistence only slightly.**
Each line shows the response of GDP to a tail event (shock realizations equal to those in 2009) under different information sets. The solid line uses the actual data from 1950-2009, while the dashed lines use a hypothetical series extended through 1890, where the actual time series from 1950-2009 is used as a proxy for the period 1890-1949, with one modification: $\{\phi_{1929}, \phi_{1930}\} = \{\phi^{\varepsilon}_{2008}, \phi^{\varepsilon}_{2009}\}$. *The parameter* ε *indexes the severity of the Great Depression relative to the Great Recession, while* $\lambda \in (0, 1]$ *indexes the extent to which older observations are discounted where* $\lambda = 1$ *represents no discounting.*

The key challenge to extending our analysis to earlier episodes is data availability – the non-financial asset series used to measure ϕ_t is available only from 1950. Other macro and financial series turn out to be unreliable proxies.[29] But, our goal here is not to explain the Great Depression, but to understand how more data, especially with previous crises, affects learning today. Therefore, we pursue an alternative approach and use the post-WW II sample to construct hypothetical scenarios for the pre-WW II period. Specifically, we assume that (ϕ_t, π_t) realizations for the period from 1890-1949 were identical to those in 1950-2009, with one adjustment: the shocks during the Great Depression period (i.e. for 1929 and 1930) were as bad as the Great Recession, i.e. we set $\{\phi_{1929}, \phi_{1930}\} = \{\phi_{2008}, \phi_{2009}\}$.

We then study the aggregate effects of a tail event – specifically, shock realizations that are equal to those observed in 2009 – both under the benchmark information set (i.e. from 1950-2007) as well as the expanded one. Under the latter, the effect of the event on beliefs is

[29] We projected the post-1950 ϕ_t series on a number of variables and used the estimated coefficients to impute values for ϕ_t pre-1950. However, this did not produce accurate estimates – specifically, it missed crises both in and out of sample. We explored a wide range of macro and asset pricing variables – including GDP, unemployment, S&P returns and the Case-Shiller home price index. We also experimented with lead-lag structures. Across specifications, the projections for 1929-1930 showed only modestly adverse realizations.

moderated by the larger size of data sample and by the presence of other tail events of similar (or worse) magnitude in it.

Of course, the further back the data sample extends, the assumption that old and new data are treated as equally relevant becomes less realistic. We therefore consider the possibility that agents discount older observations. This could reflect the potential for unobserved regime shifts or experiential learning with overlapping generations (Malmendier and Nagel, 2011).[30] To capture such discounting, we modify our kernel estimation procedure: observations from s periods earlier are assigned a weight $\lambda^s, \quad \lambda \in (0, 1]$. When $\lambda = 1$, there is no discounting.

Figure 9 reveals that, even without discounting ($\lambda = 1$, the dashed line), the difference between the benchmark and expanded information set is modest: the drop on impact is identical, by construction, but the additional data attenuates the longer term effects slightly. When older data is discounted by 1% ($\lambda = 0.99$, the center panel), this attenuation almost completely disappears and the response looks very similar to the benchmark.[31]

It is also possible that the shocks during the Great Depression were larger than the 2008-'09 ones. To allow for this, we also consider a case where $(\phi_{1929}, \phi_{1930}) = \{\phi^2_{2008}, \phi^2_{2009}\} = (0.86, 0.70)$. Note that these are very large shocks – 5 and 10 standard deviations below the mean, together eroding almost 50% of the effective capital stock. Figure 9 shows that, even in this case, the 2009 shock generates considerable persistence (the dotted line shows responses under 1% discounting, i.e. $\lambda = 0.99$).

In sum, expanding the information set by adding more data does not drastically alter our main conclusions, especially once we assume that agents discount older data.

4.4 Evidence from Asset Markets

Our framework stays close to a standard neoclassical macro paradigm and therefore, inherits many of its limitations when it comes to asset prices. Our goal in this section is not to resolve these shortcomings but to show that the predictions of the model are broadly in line with the patterns in asset markets. As with macro aggregates, effects of learning are detectable only after tail events, so we focus on the period since the Great Recession. In Table 3, we compare the predictions of the model for various asset market variables, both pre- and post-crisis, with their empirical counterparts, constructed using averages over 1990-07 and 2010-15 respectively.[32] The model does not quite match levels, but our focus is on the changes induced

[30] This is also similar to Sargent (2001), Cho et al. (2002) and Evans and Honkapohja (2001).

[31] Stronger discounting leads to a *larger* decline in GDP. Intuitively, this raises the weight of recent observations. For example, with $\lambda = 0.98$, the persistent drop in GDP is about 16%.

[32] The overall performance of the model is not sensitive to the time periods chosen. In an earlier version of this paper, we used shorter samples for both pre- and post-crisis and reached similar conclusions. We exclude 2008 and 2009 in order to avoid picking up outsized fluctuations in asset markets at the height of the crisis.

by the 2008-09 experience. We find that while the model's predictions for the change in credit spreads and equity prices are broadly consistent with the data. More interestingly, the model's predictions for tail risk, a much more direct indicator, lines up quite well with the observed changes in the probability of extreme events priced into traded options.

	Model			Data		
	$\hat{g}_{2007}$	$\hat{g}_{2014}$	Change	1990-2007	2010-2015	Change
Asset prices and debt						
Credit Spreads (non-financial)	0.88%	1.18%	0.30%	1.89%	2.23%	0.34%
Credit Spreads (financial)	0.15%	0.47%	0.33%	0.93%	1.51%	0.58%
Equity Premium	0.95%	2.43%	1.48%	3.29%	7.50%	4.21%
Equity/Assets	46.88%	47.93%	1.05%	55.28%	56.87%	1.59%
Risk free rate	9.60%	9.09%	-0.50%	1.42%	-1.32%	-2.74%
Debt			-17%	3%	-23%	-26%
Tail risk for equity						
Third moment ($\times 10^2$)	0.00	-0.24	-0.24	-1.34	-1.59	-0.25
Tail risk	0.0%	1.6%	1.6%	9.3%	11.2%	1.9%

Table 3: **Changes in financial market variables, Model vs data.**
Model: shows average values in the stochastic steady state under $\hat{g}_{2007}$ and $\hat{g}_{2014}$. The equity/assets is the ratio of the market value of equity claims to capital K_t. Third moment is $E\left[\left(R^e - \bar{R}^e\right)^3\right]$, where R^e is the return on equity and Tail risk is $Prob\left(R^e - \bar{R}^e \leq -0.3\right)$, where both are computed under the risk-neutral measure. Without learning, all changes are zero.
Data: For non-financial credit spreads, we use the average spread on senior unsecured bonds issued by non-financial firms computed as in Gilchrist and Zakrajek (2012). Financial credit spreads are option-adjusted spreads on bank holding company bonds calculated by Bank of America Merrill Lynch, as in Atkeson et al. (2018). For the equity premium, we follow Cochrane (2011) and Hall (2015b) and estimate the one-year ahead forecast for real returns on the S&P 500 from a regression using Price-Dividend and aggregate Consumption-GDP ratios. See footnote 39 for details. Equity/assets is the ratio of the market value of equities to value of non-financial assets from Table B.103 in the Flow of Funds. The risk-free real rate is computed as the difference between nominal yield of 1-year US treasuries and inflation. Debt is measured as total liabilities of nonfinancial corporate business from the Flow of Funds (FL104190005, Table B.103), adjusted for population growth and inflation. The numbers reported are deviations from a log-linear 1952-2007 trend. The third moment and tail risk are computed from the VIX and SKEW indices published by CBOE. See footnote 41 and Appendix C.5 for details.

The model predicts a increase in credit spreads, both for financial intermediaries and firms.[33] The former falls short of the observed change – 0.33% in the model versus 0.58% in the data. The fact that financial credit spreads have remained persistently high since the financial crisis has also been noted by Atkeson et al. (2018): our analysis shows that belief revisions in response

[33]The model under-predicts the levels in both cases. Fixing this requires adding more sources of risk – either aggregate or idiosyncratic – or additional curvature.

to the observed shocks during 2008-'09 can help explain a significant portion of this increase.[34]

Spreads for non-financial firms are predicted to rise by less than that of financials. This is due, in part, to equilibrium effects: an increase in bankruptcy risk induces firms to issue less debt, which is 17% lower in the new steady state.[35] In the data, total liabilities of non-financial corporations (relative to trend)[36] show a similar change – a drop of about 26%. This reduction in debt offsets the rise in default risk and therefore, credit spreads: the net effect is a modest increase (0.30%), which is quite close to its empirical counterpart (0.34%). [37]

One might think that higher tail risk should imply lower equity prices. The fact that equity prices have surged recently and are higher than their pre-crisis levels thus would appear inconsistent with a rise in tail risk. But again, this logic is incomplete – while higher tail risk does increase the risk premium, it also induces firms to cut debt, which mitigates the increase in risk (Modigliani and Miller, 1958). The net effect in the model is to slightly raise the market value of a dividend claim associated with a unit of capital under the post-crisis beliefs relative to the pre-crisis ones. In other words, the combined effect of the changes in tail risk and debt is mildly positive.[38] In the data, the ratio of the market capitalization of the non-financial corporate sector to their (non-financial) asset positions also shows an increase. While the magnitudes differ – we don't claim to solve all equity-related puzzles here – our point is simply that rising equity valuations are not evidence against tail risk.

Furthermore, changes in equity premia (the difference between expected return on equity and the riskless rate) are in the right ballpark, even though the model under-predicts the level relative to the data (reflecting its limitations as an asset pricing model). The higher tail risk under the post-crisis beliefs implies an rise of 1.48% in the equity premium, relative to that under the pre-crisis ones. The analogous object in the data is computed following the methodology in Cochrane (2011) and Hall (2015b)[39] and shows that equity premia in 2010-15 were about 4.21% higher than the pre-crisis average. In other words, tail risk can account for

[34] Sarin and Summers (2016) look at a broad range of asset prices and also conclude "To our surprise, we find that financial market information provides little support for the view that major institutions are significantly safer than they were before the crisis and some support for the notion that risks have actually increased."

[35] The leverage ratio (debt and wage obligation divided by total assets) is also slightly lower, by about 0.5%.

[36] Total liabilities of nonfinancial corporate business is taken from series FL104190005 from Table B.103 in the Flow of Funds. As with the other macro series, we adjust for inflation and population growth and then detrend using a simple log-linear trendline. The numbers reported in the table are the (averages of the) deviations from a log-linear trend, computed from 1952-2007.

[37] Belief changes about financial shocks play an important role here. Without them, the rise in tail risk with respect to capital quality shocks alone would have a negligible effect on spreads.

[38] Aggregate market capitalization in the model is the value of this claim times the capital stock.

[39] We estimate one-year ahead forecast from a regression where the left-hand variable is the one-year real return on the S&P and the right hand variables are a constant, the log of the ratio of the S&P at the beginning of the period to its dividends averaged over the prior year, and the log of the ratio of real consumption to disposable income in the month prior to the beginning of the period.

about a third of the recent rise in equity premia. Of course, this measure, like all others, is noisy and volatile. We are not claiming that the model can explain all the fluctuations – no model can – but it doesn't seem to be at odds with recent trends in equity market variables.

The table also shows the model's predictions for riskless rates. Again, as with the equity premium, the model does not quite match the level[40], but does a better job predicting the change since the Great Recession. Higher tail risk increases the premium for safe assets, reducing the riskless rate. Under our calibration, the change in beliefs induced by the 2008-09 realizations leads to a 50 bp drop in the riskless rate. In the data, the real rate (computed as the difference between 1-year nominal Treasury yield and inflation) averaged -0.81 % between 2013-15, as against 0.61% during 2005-07, a drop of about 1.4%. Thus, the model under-predicts the drop. In Kozlowski et al. (2018), we show how liquidity considerations can amplify the drop in government bond yields from higher tail risk.

In sum, none of these trends in asset markets is at odds with the tail risk story we are advancing. If credit spreads and equity premia are not clear indicators of tail risk, what is? For that, we need to turn to option prices, in particular options on the S&P 500, which can be used to isolate changes in perceived tail risk. A natural metric is the third moment of the distribution of equity returns. It is straightforward to derive this from the SKEW and VIX indices, calculated from options on the S&P 500 traded on the CBOE.[41] As Table 3 shows, the market-implied distribution has became more negatively skewed after the Great Recession. We compute the same risk-neutral third moment in the model (using the distribution for stock returns under the 2014 and 2007 beliefs). The model under-predicts the skewness in levels[42], but predicts a change (-0.0024) that lines up almost exactly with the data. To show how this maps into probabilities of tail events, we also report the implied (risk-neutral) odds of a return realization 30% less than the mean.[43] Again, the model-implied level is too low, but the predicted change (1.6%) is quite close to the corresponding object in the data (1.9%).

4.5 Understanding the Economic Response to Belief Changes

What model ingredients are needed for belief revisions to have substantial aggregate effects? To answer this, we perform a series of experiments, turning off each ingredient one-by-one in order

[40]This is partly due to our choice of a relatively low value for the discount factor, β. As we discussed in footnote 19, an alternative calibration strategy with a higher value for β yields very similar results.

[41] Formally, the third central moment under the risk-neutral measure is given by

$$\mathbb{E}\left(R^e - \bar{R}^e\right)^3 = \frac{100 - SKEW_t}{10} \cdot VIX_t^3 \; .$$

For more information, see http://www.cboe.com/micro/skew/introduction.aspx.

[42]Fixing this would require additional shocks and/or amplification mechanisms.

[43]For details of the computation, see Appendix C.5.

to isolate its contribution. Table 4 presents the results – columns (A), (B), (C) and (D) analyze versions without intermediation, debt, capital quality shocks and mean changes respectively.

	Data	Models				
		Baseline	**(A)**	**(B)**	**(C)**	**(D)**
Real economy						
Output	-12.0%	-10.6%	-6.6%	-4.4%	-1.6%	-7.7%
Labor	-7.0%	-7.5%	-4.7%	-3.1%	-1.1%	-5.5%
Investment	-17.0%	-12.0%	-6.2%	-3.0%	-2.3%	-10.6%
Credit spreads						
Firms	0.34%	0.30%	0.02%		0.12%	0.21%
Intermediaries	0.58%	0.33%	0.00%		0.13%	0.22%

Table 4: **The contribution of various model elements.**
Note: The model numbers report changes in the long-run under $\hat{g}_{2014}$, relative to $\hat{g}_{2007}$. Model (A) no intermediaries; (B) no debt; (C) no capital quality shocks; and (D) constant mean.

Role of financial shocks Column (A) in Table 4 shows that the tail realizations of the capital quality shock in 2008 and 09 induce a long-run drop in GDP of 6.6%, about two-thirds of the baseline decline. The fall in investment sees a larger attenuation (6% vs 12%). Abstracting from financial shocks also has a much more significant effect on credit spreads of firms: this version produces a very modest increase in spreads (0.02%). Thus, changes in beliefs about financial risk plays a significant role in the model's ability to match investment and credit spreads, but large, persistent effects on economic activity obtain even without them.

Role of debt To abstract from debt, we set the tax advantage parameter χ to 1, which implies all-equity financing, i.e. debt and leverage are 0. Note that this also makes financial shocks irrelevant. Column (B) in Table 4 shows that belief revisions about capital quality shocks trigger a 4.5% long-run reduction in output without debt, compared to 10.6% in the full blown version. Thus, debt plays a central role in generating large economic responses, particularly in combination with financial shocks: it accounts for about 2.2% without financial intermediation (A-B) and an additional 4% (Baseline - A) once financial tail risk is introduced.

Debt also helps explain why some shocks generate more persistent responses than others, a central question of the paper. As we discussed in Section 4.2, the belief changes induced by larger shocks are not only larger but also occur further out in the tail. As a result, they are amplified by debt, further increasing the persistent component. See analysis in Appendix C.8.

Role of curvature in utility Curvature in utility also plays an important role. Appendix C.6 shows that the effect of tail risk on aggregate outcomes is increasing in both risk aversion and the intertemporal elasticity of substitution (IES). Higher risk aversion raises risk premia, while a higher IES dampens a precautionary motive for accumulating capital in response to higher risk. The analysis also highlights the role of Epstein-Zin preferences – with CRRA preferences, for example, high risk aversion implies a low IES (and vice versa), moderating the drop in long-run economic activity.

Role of mean vs higher moments. Observing a tail event changes the mean as well as higher moments of estimated beliefs. The changes in the mean are relatively modest – $\mathbb{E}(\phi_t)$ is only 0.0009 higher under $\hat{g}_{2014}$ compared to $\hat{g}_{2007}$, while $\mathbb{E}(\pi_t)$ rises by 0.0011. To quantify their effect, we simulated the long-run effects under the assumption that the mean belief remains unchanged pre- and post-2007. The results, in column marked (D) in Table 4, suggest that most of our effects on aggregate economic activity (almost 75%) stems from changes in higher moments.[44]

4.6 Open Questions

The analysis in the preceding sections demonstrates the quantitative potential of belief revisions in explaining macro aggregates and financial market variables. It also raises new questions – both conceptual and empirical – where future research might be fruitful.

One direction is to move towards an Bayesian approach. This would allow us to incorporate the risk of future belief changes into agents' current decisions, a channel we abstract from in our classical non-parametric approach. (Collin-Dufresne et al., 2016), for example, use a model of Bayesian parameter learning to improve asset pricing predictions in an endowment setting. Embedding this in a model with production could bring to light new implications for macro phenomena as well. We conjecture that tail events will have large, persistent effects even in a Bayesian setting, provided two conditions are satisfied. First, the specification is sufficiently flexible, e.g. has one or more parameters governing tail risk. Second, the priors about these parameters reflects substantial uncertainty. Otherwise, there is not much scope for learning.

[44] Economic activity is, in fact, quite sensitive to the mean capital quality shock. In Appendix C.3, we use a deterministic version of the model without debt to derive, in closed form, the elasticity of the steady-state capital to the mean capital quality:

$$\frac{d\ln k_{ss}}{d\ln \phi_{ss}} = \left(\frac{1+\gamma}{\gamma}\frac{\alpha}{1-\alpha}\right) + \left(\frac{1}{1-\alpha}\frac{\alpha+\gamma}{\gamma}\right)\frac{(1-\delta)}{1/\beta-(1-\delta)\phi_{ss}} = 2 + 3(7.5) = 24.5.$$

Capital, and thus output, is highly sensitive to capital quality because it affects both current returns (first term) and holding gains on the undepreciated capital stock (second term). Total factor productivity, on the other hand, only affects the first term and therefore, has a much lower effect.

Another open question is the extent to which recent events reflect changes in beliefs or preferences. It is nearly impossible to disentangle the two with aggregate macroeconomic data. In this paper, we took what we believe to be the most fruitful approach – hold preferences fixed and discipline beliefs with data. But, new approaches to preference formation could shed more light on this question and provide a deeper understanding on the role of tail events.

Finally, there have been other surprising and extreme economic events. Exploring the extent to which they induced persistent responses and the ability of learning to explain those patterns is an important area for future work. Similarly, the Great Recession was not an extreme event along some dimensions. Pairing learning about tail events with limited attention might help uncover why agents focused on capital returns, instead of other, less extreme series. Perhaps this was because capital returns were both payoff-relevant and extreme, making them highly-informative events.

5 Conclusion

Economists typically assume that agents in their models know the distribution of shocks. In this paper, we showed that relaxing this assumption introduces persistent economic responses to tail events. The agents in our model behave like classical econometricians, re-estimating distributions as new data arrives. Under these conditions, observing a tail event like the 2008-09 Great Recession in the US, causes agents to assign larger weights to similar events in the future, depressing investment and output. Crucially, these effects last for a long time, even when the underlying shocks are transitory. The rarer the event that is observed, the larger and more persistent the revision in beliefs. The effects on economic activity are amplified when investments are financed with debt. This is because debt payoffs (and therefore, borrowing costs) are particularly sensitive to the probability of extreme negative outcomes.

When this mechanism is quantified using data for the US economy, the predictions of the model resemble observed macro and asset market outcomes in the wake of the Great Recession, suggesting that the persistent nature of the recent stagnation is due, at least partly, to the fact that the events of 2008-09 changed the way market participants think about tail risk.

References

Angeletos, G.-M. and J. La'O (2013). Sentiments. *Econometrica 81*(2), 739–779.

Anzoategui, D., D. Comin, M. Gertler, and J. Martinez (2015). Endogenous technology adoption and r & d as sources of business cycle persistence. New York University working paper.

Atkeson, A. G., A. d'Avernas, A. L. Eisfeldt, and P.-O. Weill (2018). Government guarantees and the valuation of american banks. *NBER Macroeconomics Annual 33*.

Backus, D. K., S. Foresi, and L. Wu (2008). Accounting for biases in black-scholes. *SSRN Working Paper Series*.

Bakshi, G., N. Kapadia, and D. Madan (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies 16*(1), 101–143.

Beaudry, P. and F. Portier (2004). An exploration into pigou's theory of cycles. *Journal of Monetary Economics 51*, 1183–1216.

Bennett, R. L. and H. Unal (2015). Understanding the components of bank failure resolution costs. *Financial Markets, Institutions & Instruments 24*(5), 349–389.

Bloom, N., M. Floetotto, N. Jaimovich, I. Sapora-Eksten, and S. Terry (2014). Really uncertain business cycles. NBER working paper 13385.

Brunnermeier, M. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *American Economic Review 104*(2).

Cho, I.-K., N. Williams, and T. J. Sargent (2002). Escaping nash inflation. *The Review of Economic Studies 69*(1), 1–40.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Cogley, T. and T. Sargent (2005). The conquest of us inflation: Learning and robustness to model uncertainty. *Review of Economic Dynamics 8*, 528–563.

Collin-Dufresne, P., M. Johannes, and L. Lochstoer (2016). Parameter learning in general equilibrium: The asset pricing implications. *American Economic Review 106*(3), 664–98.

Cooley, T. and E. Prescott (1995). *Frontiers of Business Cycle Research*, Chapter Economic Growth and Business Cycles. Princeton University Press.

Eaton, J. and M. Gersovitz (1981). Debt with potential repudiation: Theoretical and empirical analysis. *The Review of Economic Studies*, 289–309.

Eggertsson, G. and N. Mehrotra (2014). A model of secular stagnation. NBER Working Paper No. 20574.

Evans, G. W. and S. Honkapohja (2001). *Learning and expectations in macroeconomics*. Princeton University Press.

Fajgelbaum, P., E. Schaal, and M. Taschereau-Dumouchel (2014). Uncertainty traps. NBER Working Paper No. 19973.

Fernald, J. (2014). A quarterly, utilization-adjusted series on total factor productivity. *Federal reserve bank of San Francisco working paper 19.*

Gertler, M. and P. Karadi (2011). A model of unconventional montary policy. *Journal of Monetary Economics 58*, 17–34.

Gertler, M., N. Kiyotaki, et al. (2010). Financial intermediation and credit policy in business cycle analysis. *Handbook of monetary economics 3*(3), 547–599.

Gilchrist, S. and E. Zakrajek (2012, June). Credit spreads and business cycle fluctuations. *American Economic Review 102*(4), 1692–1720.

Gomme, P., B. Ravikumar, and P. Rupert (2011). The return to capital and the business cycle. *Review of Economic Dynamics 14*(2), 262–278.

Gordon, R. (2014). The demise of u. s. economic growth: Restatement, rebuttal, and reflections. NBER working paper 19895.

Gourio, F. (2012). Disaster risk and business cycles. *American Economic Review 102(6)*, 2734–66.

Gourio, F. (2013). Credit risk and disaster risk. *American Economic Journal: Macroeconomics 5(3)*, 1–34.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988). Investment, capacity utilization, and the real business cycle. *The American Economic Review*, 402–417.

Hall, R. (2015a). Macroeconomics of persistent slumps. *Handbook of Macroeconomics*.

Hall, R. E. (2015b, October). High discounts and high unemployment. Hoover Institution, Stanford University.

Hansen, B. E. (2015). *Econometrics*.

Hansen, L. (2007). Beliefs, doubts and learning: Valuing macroeconomic risk. *American Economic Review 97(2)*, 1–30.

Hansen, L. and T. Sargent (1999). *The Conquest of American Inflation*. Princeton University Press.

Huo, Z. and N. Takayama (2015). Higher order beliefs, confidence, and business cycles. University of Minnesota working paper.

Jaimovich, N. and S. Rebelo (2006). Can news about the future drive the business cycle? *American Economic Review 99(4)*, 1097–1118.

James, C. (1991). The losses realized in bank failures. *The Journal of Finance 46*(4), 1223–1242.

Johannes, M., L. Lochstoer, and P. Collin-Dufresne (2015). The asset pricing implications of priced structural parameter uncertainty.

Johannes, M., L. Lochstoer, and Y. Mou (2015). Learning about consumption dynamics. *Journal of Finance forthcoming*.

Khan, A., T. Senga, and J. K. Thomas (2014). Default risk and aggregate fluctuations in an economy with production heterogeneity. Technical report, Ohio State University working paper.

Kozlowski, J., L. Veldkamp, and V. Venkateswaran (2018). The tail that keeps the riskless rate low. *National Bureau of Economic Research, Macroeconomics Annual forthcoming*.

Lorenzoni, G. (2009). A theory of demand shocks. *American Economic Review 99*(5), 2050–84.

Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk taking?*. *The Quarterly Journal of Economics 126*(1), 373–416.

Merton, R. (1973). An intertemporal capital asset pricing model. *Econometrica 41*(5), 867–887.

Midrigan, V. and T. Philippon (2011). Household leverage and the recession.

Modigliani, F. and M. Miller (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review 48*(3), 261–297.

Moriera, A. and A. Savov (2015). The macroeconomics of shadow banking. New York University working paper.

Navarro, G. (2014). Financial crises and endogenous volatility. *manuscript, New York University*.

Nimark, K. (2014). Man-bites-dog business cycles. *American Economic Review 104(8)*, 2320–67.

Orlik, A. and L. Veldkamp (2014). Understanding uncertainty shocks and the role of the black swan. Working paper.

Pintus, P. and J. Suda (2015). Learning financial shocks and the great recession. Aix Marseille working paper.

Sargent, T. J. (2001). *The conquest of American inflation*. Princeton University Press.

Sarin, N. and L. H. Summers (2016). Understanding bank risk through market measures. *Brookings Papers on Economic Activity 2016*(2), 57–127.

Straub, L. and R. Ulbricht (2013). Credit crunches, information failures, and the persistence of pessimism. Toulouse School of Economics working paper.

Van Nieuwerburgh, S. and L. Veldkamp (2006). Learning asymmetries in real business cycles. *Journal of Monetary Economics 53(4)*, 753–772.

Veldkamp, L. and J. Wolfers (2007). Aggregate shocks or aggregate information? costly information and business cycle comovement. *Journal of Monetary Economics 54*, 37–55.

A Solution Method

The equilibrium is characterized by the following non-linear system:

$$1+\chi\mathcal{W}_t\frac{L_{t+1}}{\hat{K}_{t+1}}=\mathbb{E}[M_{t+1}R^k_{t+1}]+(\chi-1)lev_{t+1}q_t-(1-\theta)\mathbb{E}[M_{t+1}R^k_{t+1}h(\underline{v}_{t+1})]$$
$$-\mathbb{E}[M_{t+1}R^k_{t+1}\pi_{t+1}(\underline{v}_{t+1}(1-F(\underline{v}_{t+1}))+\theta h(\underline{v}_{t+1}))] \tag{16}$$

$$\chi\mathcal{W}_t=\mathbb{E}_t\left[M_{t+1}\left(1-\alpha\right)\phi^{\alpha}_{t+1}\left(\frac{\hat{K}_{t+1}}{L_{t+1}}\right)^{\alpha}\left(J^l(\underline{v}_{t+1})(1-\pi_{t+1})+\pi_{t+1}(1-h(\underline{v}_{t+1}))\right)\right] \tag{17}$$

$$\mathbb{E}_tM_{t+1}\left[(1-\theta)\underline{v}_{t+1}f\left(\underline{v}_{t+1}\right)\right]=\left(\frac{\chi-1}{\chi}\right)\mathbb{E}_t\left[M_{t+1}\left(1-F\left(\underline{v}_{t+1}\right)\right)\right]$$
$$-\mathbb{E}_t\left[M_{t+1}\pi_{t+1}(1-F(\underline{v}_{t+1})-(1-\theta)\underline{v}_{t+1}f(\underline{v}_{t+1}))\right] \tag{18}$$

$$M_{t+1}=\beta\left[\mathbb{E}\left(U^{1-\eta}_{t+1}\right)\right]^{\frac{\eta-\psi}{1-\eta}}U^{\psi-\eta}_{t+1}\left(\frac{u\left(C_{t+1},L_{t+1}\right)}{u\left(C_t,L_t\right)}\right)^{-\psi} \tag{19}$$

$$\mathcal{W}_t=L^{\gamma}_{t+1}\mathbb{E}M_{t+1} \tag{20}$$

where

$$C_t=\phi^{\alpha}_t\hat{K}^{\alpha}_tL^{1-\alpha}_t+\left(1-\delta\right)\phi_t\hat{K}_t-\hat{K}_{t+1} \tag{21}$$

$$U_t=\left[\left(1-\beta\right)\left(u\left(C_t,L_t\right)\right)^{1-\psi}+\beta\mathbb{E}\left(U^{1-\eta}_{t+1}\right)^{\frac{1-\psi}{1-\eta}}\right]^{\frac{1}{1-\psi}} \tag{22}$$

$$R^k_{t+1}=\frac{\phi^{\alpha}_{t+1}\hat{K}^{\alpha}_{t+1}L^{1-\alpha}_{t+1}+\left(1-\delta\right)\phi_{t+1}\hat{K}_{t+1}}{\hat{K}_{t+1}} \tag{23}$$

$$\underline{v}_{t+1}=\frac{lev_{t+1}}{\phi^{\alpha}_{t+1}\left(\frac{\hat{K}_{t+1}}{L_{t+1}}\right)^{\alpha-1}+\left(1-\delta\right)\phi_{t+1}} \tag{24}$$

$$J^l(\underline{v})=1+\underline{v}^2f\left(\underline{v}\right)\chi\left(1-\theta\right)-\left(1-\chi\theta\right)h\left(\underline{v}\right) \tag{25}$$

Solution Algorithm To solve the system described above at any given date t (i.e. after any observed history of (ϕ_t,π_t)), we recast it in recursive form with grids for the aggregate state (Π,L) and the shocks (ϕ,π). We then use the following iterative procedure:

- Estimate $\hat{g}$ on the available history using the kernel density estimator.
- Start with a guess (in polynomial form) for $U(\Pi,L),C(\Pi,L)$.
- Solve (16)-(18) for $\hat{K}'(\Pi,L),L'(\Pi,L),lev'(\Pi,L)$ using a non-linear solver.
- Update the guess for U,C using (21)-(22) and iterate until convergence.

Online Appendix

This material is for a separate, on-line appendix and not intended to be printed with the paper.

B Model solution and derivations 2
B.1 Firm Optimality . . . 2
B.2 Intermediation . . . 5

C Additional Results 7
C.1 Measurement of ϕ_t: Alternative price indices . . . 7
C.2 Numerical accuracy of solution method . . . 7
C.3 Steady State Analysis . . . 9
C.4 Behavior of Consumption . . . 10
C.5 Computing option-implied tail probabilities . . . 10
C.6 Role of Risk Aversion, Intertemporal Elasticity of Substitution . . . 11
C.7 Role of GHH preferences . . . 12
C.8 Debt and Shock Size . . . 13
C.9 Exogenous persistence in ϕ_t . . . 13
C.10 Learning with a Normal distribution . . . 15
C.11 Alternative kernels . . . 16
C.12 Tail Shocks vs Beliefs . . . 17
C.13 Asymmetry: Right vs left tail events . . . 18

D Other evidence 19
D.1 Internet search behavior . . . 19
D.2 Stock market . . . 19
D.3 Returns during the Great Recession . . . 20
D.4 De-trending . . . 21
D.5 TFP . . . 21

B Model solution and derivations

B.1 Firm Optimality

Define

$$R^k_{it+1} \equiv {\phi_{t+1}}^{\alpha} \left(\frac{l_{it+1}}{\hat{k}_{it+1}}\right)^{1-\alpha} + (1-\delta)\phi_{t+1}.$$

as the *ex-post* per-unit, pre-wage return on capital. Substituting for dividends and wages from (5) and (6), the firm's continuation value can be expressed as the solution to the following maximization problem:

$$\Gamma_{it} = \max_{\hat{k}_{it+1},\ lev_{it+1}, \frac{l_{it+1}}{\hat{k}_{it+1}}} \hat{k}_{it+1} \left(-1 - \chi \mathcal{W}_t \frac{l_{it+1}}{\hat{k}_{it+1}} + \chi q_{it} lev_{it+1} \ + \mathbb{E} M_{t+1} r_{it+1} \left(v_{it} R^k_{it+1} - lev_{it+1} + \frac{\Gamma_{it+1}}{\hat{k}_{it+1}}\right)\right)$$

where

$$q_{it} = \mathbb{E} M_{t+1}(1-\pi_{t+1}) \left[r_{it+1} + (1-r_{it+1})\, \theta \frac{v_{it+i} R^k_{it+1} + \frac{\Gamma_{it+1}}{\hat{k}_{it+1}}}{lev_{it+1}}\right].$$

Leverage lev_{it+1} includes debt and the wage promise made to workers. However, wage promises (or operating leverage) are different from debt, in that they does not earn a tax advantage. Since the above formulation credits the firm with tax advantage χ on all leverage, the wage obligation $\mathcal{W}_t$ needs to be multiplied by χ, i.e. the firm pays back the tax advantage from labor payments, so only external debt ends up accruing the subsidy.

We guess (and later verify) that $\Gamma_{it+1} = 0$.[45] Using the threshold characterization of the default decision,

$$\Gamma_{it} = \max\ \hat{k}_{it+1} \left(-1 - \chi \mathcal{W}_t \frac{l_{it+1}}{\hat{k}_{it+1}} + \chi q_{it} lev_{it+1} \ + \mathbb{E} M_{t+1} \int_{\underline{v}_{it+1}}^{\infty} \left(v R^k_{it+1} - lev_{it+1}\right) dF(v)\right)$$

$$= \max\ \hat{k}_{it+1} \left(-1 - \chi \mathcal{W}_t \frac{l_{it+1}}{\hat{k}_{it+1}} + \chi q_{it} lev_{it+1} \ + \mathbb{E} M_{t+1} \left[\left(1 - h\left(\underline{v}_{it+1}\right)\right) R^k_{it+1} - lev_{it+1}\left(1 - F\left(\underline{v}_{it+1}\right)\right)\right]\right)$$

where

$$\begin{aligned} q_{it} &= \mathbb{E} M_{t+1}(1-\pi_{t+1}) \left[1 - F\left(\underline{v}_{it+1}\right) + h\left(\underline{v}_{it+1}\right) \theta \frac{R^k_{it+1}}{lev_{it+1}}\right] \\ \underline{v}_{it+1} &= \frac{lev_{it+1}}{R^k_{it+1}}. \end{aligned}$$

[45] Intuitively, given constant returns to scale, the firm's problem turns out to be linear in capital. In equilibrium, therefore, in order for the firm's value to be bounded, we must have $\Gamma_{it} = 0$. See Navarro (2014).

Capital choice: Note that q_{it} is only a function of lev_{it+1} and R^k_{it+1} (which only depends on the labor-capital ratio $\frac{l_{it+1}}{\hat{k}_{it+1}}$). As a result, the objective function is linear in $\hat{k}_{it+1}$. At an interior optimum, we must have:

$$\begin{aligned}
1+\chi\mathcal{W}_t\frac{l_{it+1}}{\hat{k}_{it+1}} &= \chi q_{it}lev_{t+1}+\mathbb{E}M_{t+1}\left[\left(1-h\left(\underline{v}_{it+1}\right)\right)R^k_{it+1}-lev_{it+1}\left(1-F\left(\underline{v}_{it+1}\right)\right)\right] \\
&= \mathbb{E}M_{t+1}R^k_{it+1}+\chi q_{it}lev_{t+1}-\mathbb{E}M_{t+1}\left[h\left(\underline{v}_{it+1}\right)R^k_{it+1}+lev_{it+1}\left(1-F\left(\underline{v}_{it+1}\right)\right)\right] \\
&= \mathbb{E}M_{t+1}R^k_{it+1}+\chi q_{it}lev_{t+1}-lev_{it+1}\mathbb{E}M_{t+1}\left[h\left(\underline{v}_{it+1}\right)\frac{R^k_{it+1}}{lev_{it+1}}+\left(1-F\left(\underline{v}_{it+1}\right)\right)\right] .
\end{aligned}$$

Now,

$$\begin{aligned}
&\mathbb{E}M_{t+1}\left[h\left(\underline{v}_{it+1}\right)\frac{R^k_{it+1}}{lev_{it+1}}+\left(1-F\left(\underline{v}_{it+1}\right)\right)\right]= \\
&\mathbb{E}M_{t+1}\left\{\left(1-\pi_{t+1}\right)\left[\left(1-F\left(\underline{v}_{it+1}\right)\right)+h\left(\underline{v}_{it+1}\right)\frac{R^k_{it+1}}{lev_{it+1}}\right]+\pi_{t+1}\left[\left(1-F\left(\underline{v}_{it+1}\right)\right)+h\left(\underline{v}_{it+1}\right)\frac{R^k_{it+1}}{lev_{it+1}}\right]\right\} \\
&=q_{it}+\left(1-\theta\right)\mathbb{E}\left[M_{t+1}\left(1-\pi_{t+1}\right)h\left(\underline{v}_{it+1}\right)\frac{R^k_{it+1}}{lev_{it+1}}\right]+\mathbb{E}M_{t+1}\pi_{t+1}\left[\left(1-F\left(\underline{v}_{it+1}\right)\right)+h\left(\underline{v}_{it+1}\right)\frac{R^k_{it+1}}{lev_{it+1}}\right] .
\end{aligned}$$

The optimality condition becomes

$$\begin{aligned}
&1+\chi\mathcal{W}_t\frac{l_{it+1}}{\hat{k}_{it+1}}=\mathbb{E}M_{t+1}R^k_{it+1}+\chi q_{it}lev_{t+1} \\
&-lev_{it+1}\left(q_{it}+\left(1-\theta\right)\mathbb{E}M_{t+1}\left(1-\pi_{t+1}\right)\left[h\left(\underline{v}_{it+1}\right)\frac{R^k_{it+1}}{lev_{it+1}}\right]\right) \\
&-lev_{it+1}\left(\mathbb{E}M_{t+1}\pi_{t+1}\left[\left(1-F\left(\underline{v}_{it+1}\right)\right)+h\left(\underline{v}_{it+1}\right)\frac{R^k_{it+1}}{lev_{it+1}}\right]\right) \\
&=\mathbb{E}M_{t+1}R^k_{it+1}+\left(\chi-1\right)q_{it}lev_{t+1} \\
&-\left(1-\theta\right)\mathbb{E}M_{t+1}\left(1-\pi_{t+1}\right)\left[h\left(\underline{v}_{it+1}\right)R^k_{it+1}\right]-\mathbb{E}M_{t+1}\pi_{t+1}\left[lev_{it+1}\left(1-F\left(\underline{v}_{it+1}\right)\right)+h\left(\underline{v}_{it+1}\right)R^k_{it+1}\right] \\
&=\mathbb{E}M_{t+1}R^k_{it+1}+\left(\chi-1\right)q_{it}lev_{t+1}-\left(1-\theta\right)\mathbb{E}\left[M_{t+1}R^k_{it+1}h\left(\underline{v}_{it+1}\right)\right] \\
&-\mathbb{E}\left[M_{t+1}\pi_{t+1}R^k_{it+1}\left(\theta h\left(\underline{v}_{it+1}\right)\underline{v}_{it+1}\left(1-F\left(\underline{v}_{it+1}\right)\right)\right)\right] .
\end{aligned} \tag{26}$$

Note that this verifies our guess that $\Gamma_{it+1}=0$. Labor and leverage choice solve

$$\max_{lev_{it+1},\frac{l_{it+1}}{\hat{k}_{it+1}}} -1+\chi q_{it}lev_{t+1}-\chi\mathcal{W}_t\frac{l_{it+1}}{\hat{k}_{it+1}}+\mathbb{E}M_{t+1}\left[\left(1-h\left(\underline{v}_{it+1}\right)\right)R^k_{it+1}-lev_{it+1}\left(1-F\left(\underline{v}_{it+1}\right)\right)\right]$$

which, after substituting for q_{it}, becomes

$$\max_{lev_{it+1},\frac{l_{it+1}}{\hat{k}_{it+1}}} -1-\chi\mathcal{W}_t\frac{l_{it+1}}{\hat{k}_{it+1}}+\mathbb{E}\left[R^k_{it+1}M_{t+1}\left(J^k\left(\underline{v}_{it+1}\right)-\pi_{t+1}J^{k\pi}\left(\underline{v}_{it+1}\right)\right)\right]$$

where

$$\begin{aligned} J^k\left(\underline{v}\right) &= 1+\left(\chi-1\right)\underline{v}\left(1-F\left(\underline{v}\right)\right)+\left(\chi\theta-1\right)h\left(\underline{v}\right)\\ J^{k\pi}\left(\underline{v}\right) &= \chi\underline{v}\left(1-F\left(\underline{v}\right)\right)+\chi\theta h\left(\underline{v}\right)\ . \end{aligned}$$

Labor choice: The first order condition with respect to $\frac{l_{it+1}}{\hat{k}_{it+1}}$ is

$$\begin{aligned} \chi\mathcal{W}_t &= \mathbb{E}M_{t+1}R^k_{it+1}\frac{\partial J^k(\underline{v}_{it+1})}{\partial\frac{l_{it+1}}{\hat{k}_{it+1}}}+\mathbb{E}M_{t+1}\frac{\partial R^k}{\partial\frac{l_{it+1}}{\hat{k}_{it+1}}}J^k(\underline{v}_{it+1})\\ &\quad -\mathbb{E}\pi_{t+1}M_{t+1}R^k_{it+1}\frac{\partial J^{k\pi}(\underline{v}_{it+1})}{\partial\frac{l_{it+1}}{\hat{k}_{it+1}}}-\mathbb{E}\pi_{t+1}M_{t+1}\frac{\partial R^k}{\partial\frac{l_{it+1}}{\hat{k}_{it+1}}}J^{k\pi}(\underline{v}_{it+1}), \end{aligned}$$

Now,

$$\begin{aligned} R^k_{it+1}\frac{\partial J^k(\underline{v}_{it+1})}{\partial\frac{l_{t+1}}{\hat{k}_{it+1}}} &= R^k_{it+1}\frac{\partial\underline{v}_{it+1}}{\partial\frac{l_{t+1}}{\hat{k}_{it+1}}}\left(\left(\chi-1\right)\left(1-F\left(\underline{v}_{it+1}\right)\right)-\underline{v}_{it+1}\left(\chi-1\right)f\left(\underline{v}_{it+1}\right)+\left(\chi\theta-1\right)\frac{\partial h\left(\underline{v}_{it+1}\right)}{\partial\underline{v}_{it+1}}\right)\ ,\\ R^k_{it+1}\frac{\partial\underline{v}_{it+1}}{\partial\frac{l_{it+1}}{\hat{k}_{it+1}}} &= -\underline{v}_{it+1}\frac{\partial R^k_{it+1}}{\partial\frac{l_{it+1}}{\hat{k}_{it+1}}}\ ,\\ \frac{dh\left(\underline{v}_{it+1}\right)}{d\underline{v}_{it+1}} &= \underline{v}_{it+1}f\left(\underline{v}_{it+1}\right)\ ,\\ \frac{\partial R^k_{it+1}}{\partial\frac{l_{it+1}}{\hat{k}_{it+1}}} &= \left(1-\alpha\right){\phi_{t+1}}^{\alpha}\left(\frac{\hat{k}_{it+1}}{l_{it+1}}\right)^{\alpha}. \end{aligned}$$

Substituting and rearranging terms yields

$$\chi\mathcal{W}_t=\mathbb{E}\left[M_{t+1}\left(1-\alpha\right){\phi_{t+1}}^{\alpha}\left(\frac{\hat{k}_{it+1}}{l_{it+1}}\right)^{\alpha}\left(J^l\left(\underline{v}_{it+1}\right)\left(1-\pi_{t+1}\right)+\pi_{t+1}\left(1-h\left(\underline{v}_{it+1}\right)\right)\right)\right] \quad (27)$$

where

$$J^l\left(\underline{v}\right)=1+\chi\left(1-\theta\right)\underline{v}^2f\left(\underline{v}\right)+\left(\chi\theta-1\right)h\left(\underline{v}\right)\ .$$

Leverage choice: The first order condition with respect to lev_{it+1} is

$$\mathbb{E}M_{t+1}R_{it+1}^{k}\left(\frac{\partial J^{k}(\underline{v}_{it+1})}{\partial lev_{it+1}}-\pi_{t+1}\frac{\partial J^{k\pi}(\underline{v}_{it+1})}{\partial lev_{it+1}}\right)=0,$$

where

$$\begin{aligned}\frac{\partial J^{k}(\underline{v}_{it+1})}{\partial lev_{it+1}} &= \frac{\partial \underline{v}_{it+1}}{\partial lev_{it+1}}\left((\chi-1)\left(1-F\left(\underline{v}_{it+1}\right)\right)-(\chi-1)\,\underline{v}_{it+1}f\left(\underline{v}_{it+1}\right)+(\chi\theta-1)\,\underline{v}_{it+1}f\left(\underline{v}_{it+1}\right)\right)\\ &= \frac{1}{R_{it+1}^{k}}\left((\chi-1)\left(1-F\left(\underline{v}_{it+1}\right)\right)-\chi\left(1-\theta\right)\underline{v}_{it+1}f\left(\underline{v}_{it+1}\right)\right)\ .\end{aligned}$$

and

$$\begin{aligned}\frac{\partial J^{k\pi}(\underline{v}_{it+1})}{\partial lev_{it+1}} &= \frac{\partial \underline{v}_{it+1}}{\partial lev_{it+1}}\left(\chi\left(1-F\left(\underline{v}_{it+1}\right)\right)-\chi\underline{v}_{it+1}f\left(\underline{v}_{it+1}\right)+\chi\theta\frac{\partial h\left(\underline{v}_{it+1}\right)}{\partial\underline{v}_{it+1}}\right)\\ &= \frac{1}{R_{it+1}^{k}}\left(\chi\left(1-F\left(\underline{v}_{it+1}\right)\right)-\chi\left(1-\theta\right)\underline{v}_{it+1}f\left(\underline{v}_{it+1}\right)\right)\ .\end{aligned}$$

Substituting and re-arranging,

$$\begin{aligned}&(1-\theta)\,\mathbb{E}M_{t+1}\underline{v}_{it+1}f\left(\underline{v}_{it+1}\right)=\\ &\frac{\chi-1}{\chi}\mathbb{E}M_{t+1}\left(1-F\left(\underline{v}_{it+1}\right)\right)-\mathbb{E}M_{t+1}\pi_{t+1}\left(1-F\left(\underline{v}_{it+1}\right)-(1-\theta)\,\underline{v}_{it+1}f\left(\underline{v}_{it+1}\right)\right)\ .\end{aligned}\tag{28}$$

Finally, since all firms make symmetric choices, we can suppress the i subscript, so

$$\hat{k}_{it+1}=\hat{K}_{t+1}\qquad l_{it+1}=L_{t+1}\qquad lev_{it+1}=lev_{t+1}\qquad \underline{v}_{it+1}=\underline{v}_{t+1}\ .$$

Using this, equations (26) – (28) become (11) – (14) in the main text.

B.2 Intermediation

In this subsection, we describe financial intermediation in detail and derive the bond price schedule faced by firms.

There is a continuum of competitive intermediaries who live for 2 periods. We assume that they are members of the representative household (so any profits or rents they earn flow back to the household) but have no direct control over its resources. They compete to raise money and invest it in bonds issued by firms.

We assume that each intermediary maintains an exogenous (but potentially time-varying) leverage ratio. Let $\mu\equiv\frac{D}{Lq}$ be the target leverage, where D denotes debt issued by the intermediary, L is the face value of bonds purchased and q the corresponding price.

Thus, for each dollar invested in bonds of firm i in period t, the intermediary raises $q_{it}\mu_t$ from debtholders and the remaining from equity. The bond pays off $\left[r_{it+1} + (1 - r_{it+1}) \frac{\theta \tilde{V}_{it+1}}{B_{it+1}}\right]$ in period $t+1$. In each period, a fraction of intermediaries are subject to a 'default' shock, parameterized by π_{t+1}. As mentioned in the main text, this can be interpreted in several ways, but for concreteness, we adopt a particular microfoundation here: π_{t+1} is the probability that a given intermediary is able to divert the entire asset portfolio and makes zero payments to both equity- and debt-holders. This stark assumption is not critical for our results. It is possible to extend the model to allow for partial default, but given our focus, we chose the simpler formulation. Note also that since the intermediaries are also members of the household, the diverted funds flow back to the household, i.e. the losses here are private, not social.

If the intermediary is not hit by the default shock, debt-holders are repaid in full (along with a gross interest rate of R_t^d)[46] and equity-holders receive dividends. Competition among intermediaries implies that, in the absence of default, they receive zero payments, i.e. the proceeds from assets less payments to debt-holders are entirely paid out as dividends to equity-holders.[47] Thus, equity-holders' payoffs (per dollar invested in bonds of i) are given by

$$-(1-\mu)\, q_{it} + E_t\left[M_{t+1} Div_{it+1}\right]$$

$$\text{where} \quad Div_{it+1} = \begin{cases} \left[r_{it+1} + (1 - r_{it+1}) \frac{\theta \tilde{V}_{it+1}}{B_{it+1}}\right] - \mu_t q_{it} R_t^d & \text{w.p. } 1 - \pi_{t+1} \\ 0 & \text{otherwise.} \end{cases}$$

In equilibrium, we must have

$$(1-\mu)\, q_{it} = E_t\left[M_{t+1} Div_{it+1}\right] \tag{29}$$

Debt-holders receive R_t^d with probability $1 - \pi_{t+1}$. In equilibrium, therefore,

$$1 = E_t\left[M_{t+1}(1 - \pi_{t+1}) R_t^d\right] \qquad \Rightarrow R_t^d = \frac{1}{E_t\left[M_{t+1}(1 - \pi_{t+1})\right]}$$

[46] This is again a simplifying assumption and effectively makes intermediary leverage irrelevant for our purposes. This implicitly assumes that intermediary leverage is not too high relative to the shocks on the asset portfolio, which is not a bad approximation given our calibration of corporate default rates. While it is possible to relax this assumption and allow for default on intermediary debt even in the absence of the financial shock, Atkeson et al. (2018) and Sarin and Summers (2016) argue that the post-crisis changes in bank leverage have not really led to lower risk premia. Our specification is consistent with this finding.

[47] To see this, suppose an intermediary offers a contract where she receives a positive payment in the no-default state. Then, another intermediary can offer equity-holders a better deal by slightly reducing that payment.

Substituting in (29) and re-arranging yields the equilibrium price schedule,

$$q_{it} = E_t M_{t+1} \left(1 - \pi_{t+1}\right) \left[r_{it+1} + \left(1 - r_{it+1}\right) \frac{\theta \tilde{V}_{it+1}}{B_{it+1}} \right]$$

which is (9) in the main text.

C Additional Results

C.1 Measurement of ϕ_t: Alternative price indices

Figure 10 shows that the measurement of capital quality shocks is unaffected when we use the price index for GDP or Personal Consumption Expenditure to control for nominal price changes.

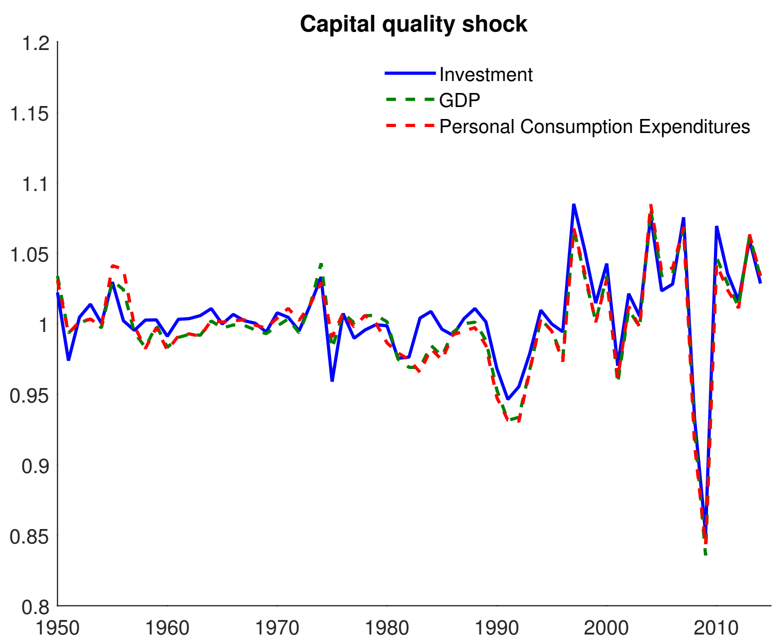


Figure 10: **Time series of shocks ϕ_t using different indices to control for nominal price changes.**

C.2 Numerical accuracy of solution method

To test the numerical accuracy of our solution method, we perform the following exercise. Starting from the steady state of $\hat{g}_{2007}$, we simulate time paths for two different economies. In Model I, as new data arrives, we update beliefs and policy functions at each date and history. In

Model II, beliefs and policy functions are fixed at $\hat{g}_{2007}$. In our solution, we essentially assume that agents use Model II as an approximation for Model I, while evaluating continuation values. Table 5 shows the sample mean and coefficient of variation for output at different horizons for these two versions.[48] It is easy to see that aggregates (or at least, the first two moments thereof) are very well-approximated by replacing the sequence of future distributions with their conditional mean. Recall that this numerical procedure works reasonable well thanks to the martingale property of beliefs.

	Horizon			
	$s=1$	$s=5$	$s=10$	$s=15$
$\mathbb{E}_t\left[y_{t+s}\right]$				
Model I:	6.378	6.385	6.393	6.397
Model II:	6.378	6.385	6.394	6.398
$\mathbb{CV}_t\left[y_{t+s}\right]$				
Model I:	0.010	0.032	0.042	0.046
Model II:	0.010	0.031	0.040	0.044

Table 5: **Numerical accuracy**

The rows labeled Model I show the actual moments under the assumption that beliefs $\hat{g}_{2007+s}$ are re-estimated at each date. Model II corresponds to the assumption underlying our solution method, where future beliefs are replaced by $\hat{g}_{2007}$.

[48] These are averages over 4000 paths. Other aggregate variables, e.g. capital and labor, show similar patterns.

C.3 Steady State Analysis

To dig a little deeper into why long-run outcomes are so sensitive to ϕ, we turn to a special case - a deterministic version of our economy without debt. The level of steady state capital is given by the following equation[49]

$$\ln k_{ss} = \text{Const.} + \left(\frac{1+\gamma}{\gamma}\frac{\alpha}{1-\alpha}\right)\ln\phi_{ss} - \left(\frac{1}{1-\alpha}\frac{\alpha+\gamma}{\gamma}\right)\ln\left(\frac{1}{\beta} - (1-\delta)\,\phi_{ss}\right). \tag{30}$$

Hence, the effect of the mean shock on steady sate capital is given by

$$\frac{d\ln k_{ss}}{d\ln\phi_{ss}} = \left(\frac{1+\gamma}{\gamma}\frac{\alpha}{1-\alpha}\right) + \left(\frac{1}{1-\alpha}\frac{\alpha+\gamma}{\gamma}\right)\frac{(1-\delta)}{1/\beta-(1-\delta)\phi_{ss}}.$$

Under our parameterization,

$$\frac{1+\gamma}{\gamma}\frac{\alpha}{1-\alpha} = 2, \qquad \frac{1}{1-\alpha}\frac{\alpha+\gamma}{\gamma} = 3, \qquad \left(\frac{(1-\delta)}{1/\beta-(1-\delta)\phi_{ss}}\right)_{\phi_{ss}=1} = 7.5$$

which implies $\frac{d\ln k_{ss}}{d\ln\phi_{ss}} = 2 + 3(7.5) = 24.5$. This simple calculation shows the source of the high sensitivity - the fact that capital quality shock affects not just the current return component but also the portion that comes from the undepreciated stock.

[49] In steady state, $M_t = 1$ and the intertemporal Euler equation and labor optimality conditions reduce to

$$\begin{aligned} 1 &= \beta\left(\alpha\phi_{ss}^{\alpha}k_{ss}^{\alpha-1}l_{ss}^{1-\alpha} + \phi_{ss}(1-\delta)\right) \\ l_{ss}^{\gamma} &= W_{ss} = (1-\alpha)\,\phi_{ss}^{\alpha}k_{ss}^{\alpha}l_{ss}^{-\alpha}. \end{aligned}$$

Substituting for l_{ss} from the second into the first and re-arranging yields the expression (30).

C.4 Behavior of Consumption

Figure 11 shows that the behavior of consumption, as predicted by the model and the corresponding pattern in the data. The model over-predicts the drop in consumption in the years immediately following impact – the flip side of its inability to match the full extent of the drop in investment during that time – but over a longer horizon, the predicted drop lines up quite well with the data.

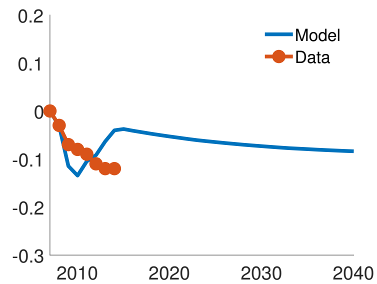


Figure 11: **Response of consumption.**

C.5 Computing option-implied tail probabilities

To compute tail probabilities, we follow Backus et al. (2008) and use a Gram-Charlier expansion of the distribution function.[50] This yields an approximate density function for the standardized random variable, $\omega = \frac{x-\mu}{\sigma}$:

$$f(\omega) = \varphi(\omega)\left[1 - \gamma\frac{(3\omega - \omega^3)}{6}\right] \qquad \text{where} \qquad \gamma = E\left[\frac{x-\mu}{\sigma}\right]^3$$

where $\varphi(\omega)$ is the standard normal density and γ is the skewness.[51] The VIX and the SKEW indices provide the standard deviation and the skewness of the implied risk-neutral distribution of the returns on the S&P 500. The numbers reported for tail probabilities in Table 3 are computed using this distribution.

[50]The CBOE follows this method to compute implied probabilities in their white paper on the SKEW Index.

[51]The Gram-Charlier expansion also includes a term for the excess kurtosis, but is omitted from the expansion because, as shown by Bakshi et al. (2003), it is empirically not significant.

C.6 Role of Risk Aversion, Intertemporal Elasticity of Substitution

	Data	Epstein-Zin			CRRA	
Risk Aversion (η)		10	10	5	0.5	2
IES($1/\psi$)		2	1	2	2	0.5
Real economy						
Output	-12.0%	-10.6%	-8.4%	-6.7%	-4.2%	-3.7%
Labor	-7.0%	-7.5%	-6.0%	-4.7%	-2.9%	-2.7%
Investment	-17.0%	-12.0%	-8.6%	-6.4%	-2.7%	-1.9%
Credit spreads						
Credit spreads	0.34%	0.30%	0.29%	0.19%	0.13%	0.15%
Intermediaries spreads	0.58%	0.33%	0.32%	0.21%	0.14%	0.16%
Tail risk						
Third moment ($\times 10^2$)	-0.24	-0.24	-0.24	-0.16	-0.11	-0.12
Third risk ($\times 10^2$)	1.60	1.63	1.59	0.49	0.12	0.19

Table 6: Role of risk aversion and intertemporal elasticity of substitution. The second panel reports the difference in the value of the variable in the new (post-crisis) and old (pre-crisis) stochastic steady states.

Risk aversion, IES and debt all play a role in determining the magnitude of the effects of increased tail risk. In order to show how much, here we compare our baseline results to a number of alternative parameterizations. The results for the role of recursive preferences and assumptions are collected here in Table 6. The first column reproduces our benchmark results, which sets risk aversion = 10 and $IES = 2$. The next two columns vary, respectively, risk aversion holding IES constant and IES holding risk aversion constant. The last 2 columns show results under CRRA utility, with a risk aversion coefficient of 2 and 0.5 respectively.

Our estimate for the IES is drawn from the macro and asset-pricing literature – see, e.g., Bansal and Yaron (2004), Barro (2009), Baron et. al. (2014). In order to assess the robustness of our results to this parameter, we ran the model with an IES of 1 – the results are presented in Column 2. Under this parameterization, the model predicts a slightly lower, but importantly just as persistent, drop in GDP (8.4% vs 10.6% in the benchmark). This is due to a precautionary channel – agents dislike intertemporal fluctuations in consumption, so faced with the increased likelihood of a tail event, they have an incentive to hold *more* capital to mitigate the potential consumption drop. This channel is stronger, the lower is IES. In fact, as the IES approaches 0, this channel becomes so powerful that it can overwhelm the disincentives to invest and can lead agents to *increase* investment in response to higher tail risk. However, in the region that the macro/asset-pricing literature typically focuses on, the effects of varying IES are relatively modest.

Analogously, column 3 reveals that the size of the drop in economic activity from increased tail risk is lower when agents are less risk averse. This is intuitive – the extent to which agents dislike the increased riskiness of investment depends on their aversion to risk. However, as with IES, the magnitude of our effects is not particularly sensitive to this parameter.

The previous two exercises show that the magnitude of effects of increased tail risk on the macro economy are increasing, albeit modestly, in both risk aversion and IES. Under CRRA utility, of course, the two are tightly (and negatively) linked – a high risk aversion necessarily implies a low IES and vice-versa. For example, in Column 4 of Table 6, we show results for a CRRA specification with the same IES as the benchmark parameterization. However, this now comes with a much lower risk aversion (0.5 vs 10), which attenuates the long-run drop in GDP. Finally, Column 5 shows results for a CRRA specification with an IES of 0.5 (or equivalently, risk aversion of 2). This also implies a small drop in economic activity.

C.7 Role of GHH preferences

The GHH specification of utility has criticized as being inconsistent with the facts on long run growth, specifically the observation that labor input is more or less constant (or maybe, slightly declining) in most advanced economies. One resolution is the following specification proposed by Jaimovich and Rebelo (2006):

$$u(C_t, L_t) = C_t - X_t \frac{L_t^{1+\gamma}}{1+\gamma} \qquad\qquad X_t = X_{t-1}^{1-\varrho} C_t^{\varrho}$$

Now, on the balanced growth path, the state variable X_t grows at the same rate as wages, ensuring labor stays constant. The parameter ϱ governs the strength of wealth effects on labor supply away from the long run. The lower value of ϱ, the closer the behavior of the economy is to the GHH specification in the short-to-medium run. In their baseline calibration, Jaimovich and Rebelo use $\varrho = 0.001$ at a quarterly frequency.

Solving this version of our model with learning involves an additional state variable and considerable computational complexity. However, a simple back-of-the-envelope calculation suggests that the drop in GDP and consumption over a 30 year horizon would only be slightly lower than our baseline (GHH) specification (about 9% instead of 11%). To see why, a 10% drop in consumption, along with $\varrho = 0.001$, implies a change in X_t over 30 years of approximately $0.1(1 - 0.999^{120}) = 0.011$. Assuming that wages change by about the same as in the baseline, the optimality conditions for labor and capital imply that the drops in L_t and K_t are about 2% lower than under GHH (5% instead of 7% and 13% instead of 15%, respectively), consistent with the conjectured drops in GDP and consumption. Over shorter horizons, e.g. in the decade immediately after the recession, the two specifications would be virtually indistinguishable. [52]

[52] As an additional robustness exercise, we repeated the steady-state exercise in Appendix C.3 with Cobb-

C.8 Debt and Shock Size

Debt also helps explain why some shocks generate more persistent responses than others, a central question of the paper. The attractiveness of debt (and therefore, the incentives to borrow) is affected disproportionately by perceived tail risk. As we discussed in Section 4.2, belief changes from larger shocks are not only larger but also occur further out in the tail. As a result, they are amplified by debt, further increasing the persistent component.

In Figure 12, we plot the long-run effects of adverse capital quality shocks ranging in size from 1 to 5 standard deviations in an economy without intermediation. The initial position corresponds to the steady state under the beliefs induced by data through 2007. The responsiveness to small shocks is almost the same with and without debt, but larger shocks see significant amplification from the non-linearity induced by debt. Thus, debt makes the severity and persistence of unusual events differ from more common downturns.

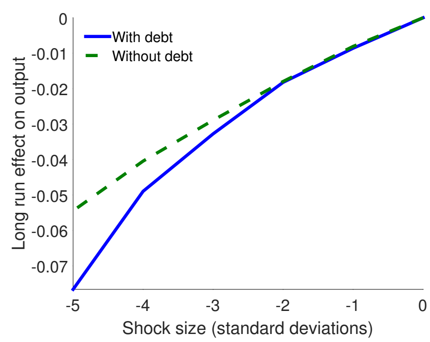


Figure 12: **Debt amplifies belief revisions from large shocks.**
Change in long-run GDP both with (solid line) and without debt (dashed line) in response to negative shocks of various sizes. The initial condition is the $\hat{g}_{2007}$ *steady state.*

C.9 Exogenous persistence in ϕ_t

In this section, we show that the observed degree of persistence in the data is just not enough to explain the prolonged stagnation since 2008-'09: in other words, learning is key to generating persistence. To do this, we solved a version of our model without learning where the ϕ_t shocks are no longer iid. Formally, we fit an AR(1) process to the observed ϕ_t series (this produces an autocorrelation estimate of 0.15) and then non-parametrically estimate the joint distribution of innovations to this process and the financial shock π_t (which is still assumed to be iid). As in the standard rational expectations setting, agents are assumed to know this process from the

Douglas preferences: $u(C_t, L_t) = C_t^{\varkappa}(1 - L_t)^{1-\varkappa}$. The responsiveness of capital and output to a change in the steady-state level of ϕ is about 70% of the elasticity in the baseline case. In other words, even with wealth effects on labor supply, the effects of increased tail risk in the long run are quite significant.

very beginning. In Figure 13, we plot the resulting impulse responses (the green dashed line). As the graph shows, the implications are quite similar to the iid, no-learning case – investment surges and the economy slowly but steadily recovers back to the pre-crisis level. Even if we used a shock process that was twice as persistent ($\rho = 0.30$) as the data, the results do not change significantly, as we see in Figure 14. These results suggest that persistence of the shock itself is an unlikely candidate to explain the prolonged stagnation.

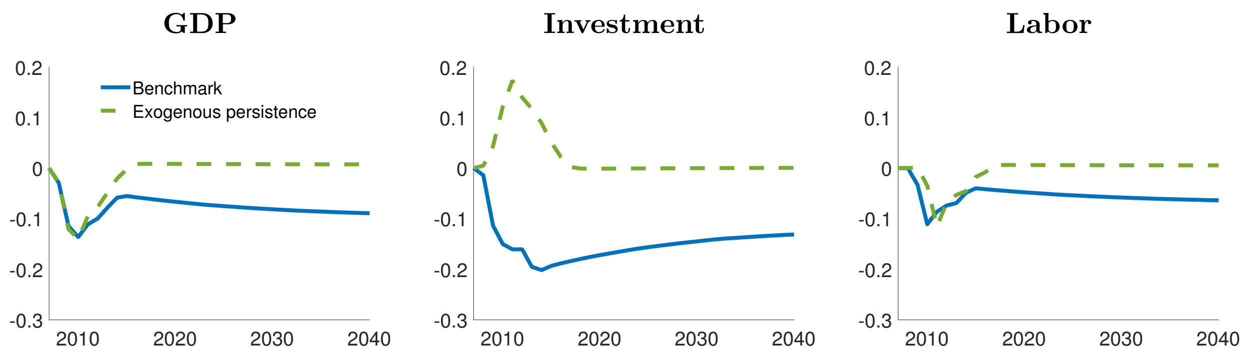


Figure 13: No learning model, with persistent shocks (dashed line, $\rho = 0.15$) vs. learning model with iid shocks.

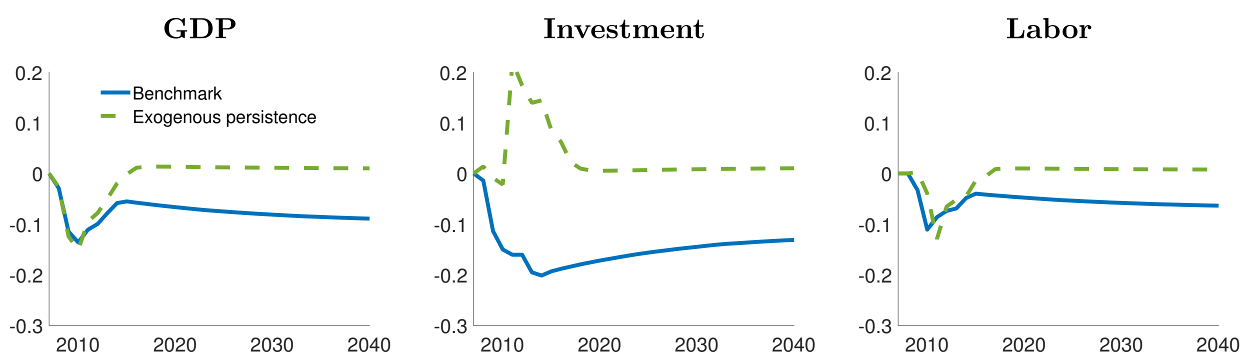


Figure 14: No learning model, with 2× estimated persistence (dashed line, $\rho = 0.30$) vs. learning model with iid shocks (solid line).

C.10 Learning with a Normal distribution

Here, we repeat our analysis under the assumption that agents fit a normal distribution. Formally, they assume that $(\phi_t, \ln \pi_t)$ is an iid draw from a joint normal distribution and estimate its parameters with the available data. The resulting beliefs for the marginal pdf of the capital quality shock are shown in the second panel of figure 15 (the first panel reproduces the baseline kernel density estimates). The large, negative tail realizations observed in 2008-'09 lowers the mean and increases the variance of the estimated normal distribution. Qualitatively, these belief revisions are also long-lived, for the same reason as those under the kernel density estimation. The economic implications are also sizable and similar to our baseline, especially in the short run. This is partly the result of the direct impact of the shock itself and partly from the fact that changes in the first two moments have an substantial effect in this highly non-linear setting.

However, the two procedures imply different time paths for beliefs and economic activity. This is seen most clearly in the exercise where we simulate the economy by drawing time paths from the pre-crisis distribution. The third panel compares the average path for GDP when agents estimate a lognormal distribution to the baseline (kernel density) case. The graph shows faster recovery for macro variables under the former. This is because realizations anywhere in the support contain information about the mean and variance of the normal distribution. The kernel estimate of the distribution at a particular point in the support, on the other hand, places relatively more weight on the observed history close to it, making learning more 'local'. The non-parametric procedure captures the idea that tail events are harder to learn about, because they are, by definition, rare. Imposing a parametric form on the distribution essentially allows the agent to learn about the probability of disasters from more normal times, and therefore, ties learning about tail risk much more closely to learning about the rest of the distribution. Obviously, if the parametric form of the distribution was known, this is the efficient thing to do, but this exercise illustrates how the assumption can have a significant effect.

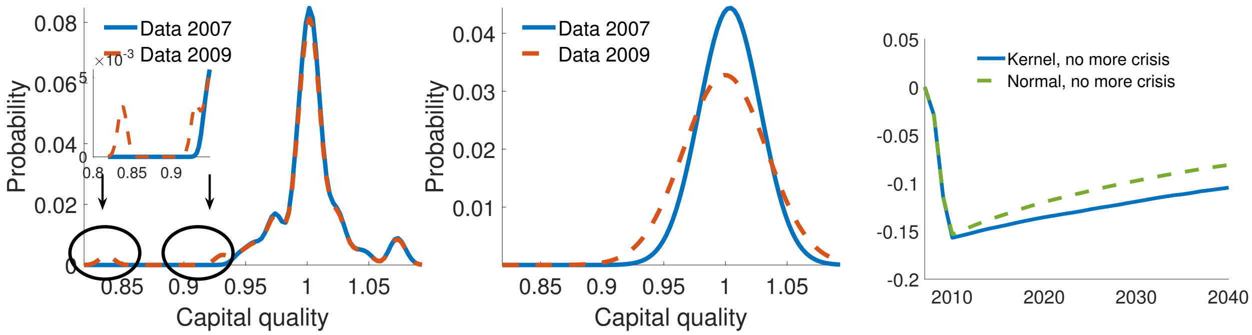


Figure 15: **Learning with a Normal distribution.**
Beliefs under our baseline non-parametric proecudure (first panel) and assuming a normal distribution (second panel). The third panel shows the exercise where we simulate the economy by drawing time paths from the pre-crisis distribution.

C.11 Alternative kernels

We estimated our belief process using alternative kernel densities. In the benchmark solution $\Omega(\cdot)$ is the normal density. This appendix consider another two common kernel densities: (i) the Epanechnikov, $\Omega(x) = \frac{3}{4}(1-x^2)$, and (ii) the box or uniform density, $\Omega(x) = \frac{1}{2}$.[53] Figure 16 shows that these approaches yielded similar changes in tail probabilities and therefore, similar predictions for economic outcomes. A Bayesian approach is conceptually similar – posterior beliefs exhibit the martingale property, the key source of persistence. However, the departure from normality needed to capture tail risk, requires particle filtering techniques, making it difficult to integrate it into any but the simplest economic environments. For a detailed discussion of non-parametric estimation, see Hansen (2015).

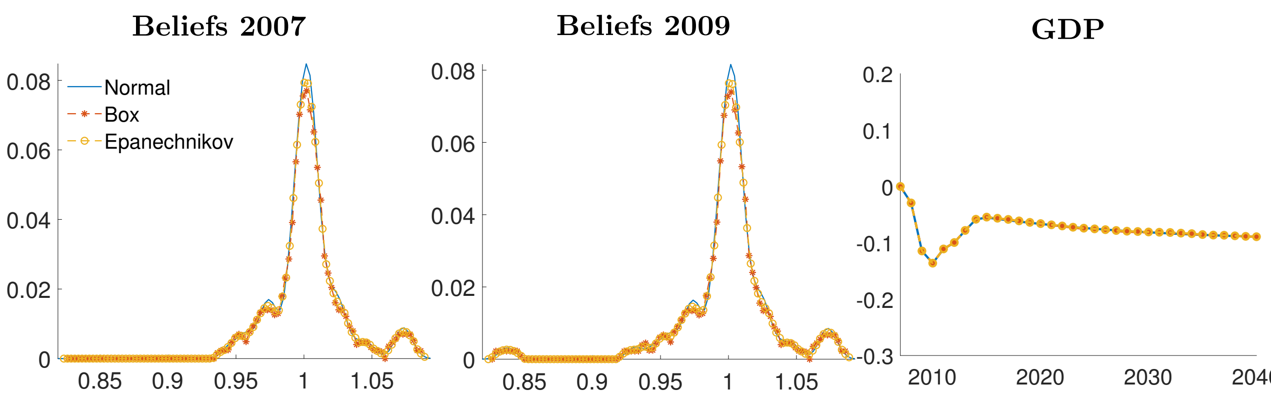


Figure 16: **Alternative Kernel densities**
The solution of the model is very similar under normal, box or epanechnikov kernel densities.

[53] We find similar results under further additional distributions: (i) the Champernowne transformation (which is designed to better capture tail risk), (ii) semi-parametric estimators, e.g. with Pareto tails and (iii) the g-and-h family of distributions which allows for a flexible specification of tail risk using various transformations of the normal distribution.

C.12 Tail Shocks vs Beliefs

The time path of aggregate variables in our baseline model – e.g. in Figure 6 – reflect the combined effects of the tail realizations of the capital quality shock in 2008-'09 (which directly reduce productive capacity) as well as the belief changes they induce. Here, we perform an exercise intended to isolate the role of the latter and show that they account for a significant fraction of the persistence. In Figure 17, we plot the time path of GDP under the assumption that beliefs (and therefore, policy functions) jump in 2009 to $\hat{g}_{2014}$ from $\hat{g}_{2007}$ but the realizations in 2008 and 2009 are not tail events (instead, we simulate shocks for those two years by drawing from $\hat{g}_{2014}$ and average over time paths). Now, output does not drop immediately but the economy steadily reduces its capital (by investing less) and converges to the new steady state. Importantly, even without the large negative shocks, we obtain a sizable and persistent drop in economic activity, underscoring the key role played by belief changes in driving our results.

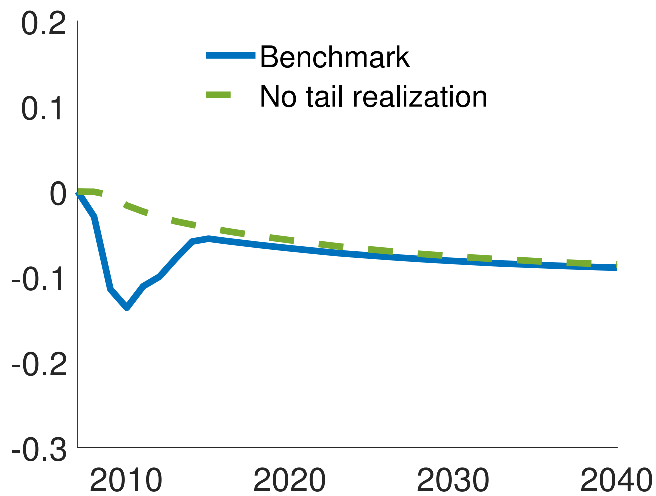


Figure 17: **Decomposition: Beliefs vs Shocks**
Response of GDP under the counterfactual in which beliefs change in 2009 but there is no actual realization of tail events ϕ_t that hit the capital stock.

C.13 Asymmetry: Right vs left tail events

Figure 18 compares (the absolute value of) responses of the economy to left and right tail events starting from the stochastic steady state associated with $\hat{g}_{2007}$. The former uses the actual realizations during 2008 and 2009 ($\phi_{2008} = 0.93$ and $\phi_{2009} = 0.84$) while the latter considers positive shocks of a similar magnitude ($\phi_{2008} = 1.07$ and $\phi_{2009} = 1.16$). The graph shows that long-run changes are smaller when the economy is hit by a positive tail event than under negative ones. This asymmetry is the result of non-linearities in the model stemming from curvature in utility and the presence of debt.

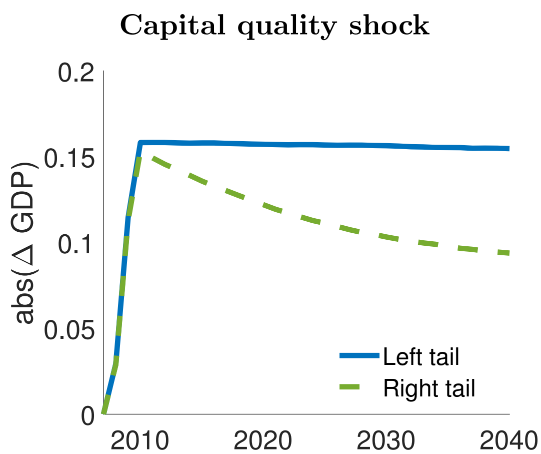


Figure 18: **Response of GDP to left and right tail shocks.**
Solid (dashed) line shows the absolute change in GDP after a left (right) tail event.

D Other evidence

D.1 Internet search behavior

Data on internet search behavior lends support to the idea that assessments of tail risk are persistently higher after the financial crisis. Figure 19 shows that the frequency of searches for the terms “financial crisis,” “economic crisis,” and “systemic risk” spiked during the crisis and then came back down. But this search frequency did not return to its pre-crisis level. In each case, there was some sustained interest in crises at a higher level than pre-2007. We find similar results for searches on the terms “economic collapse,” “financial collapse,” and “tail risk” yielded similar results.

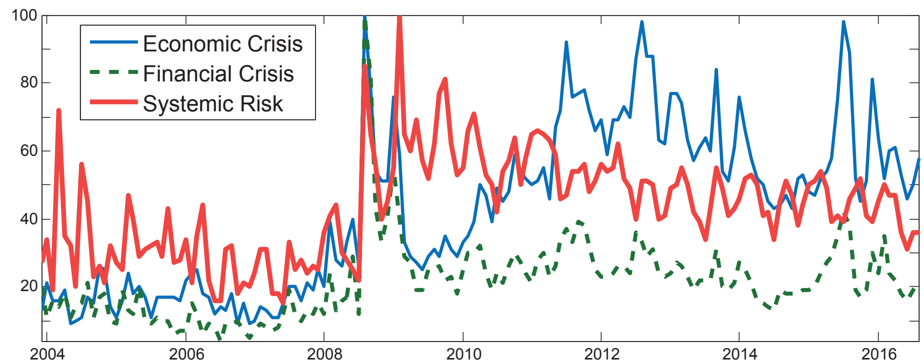


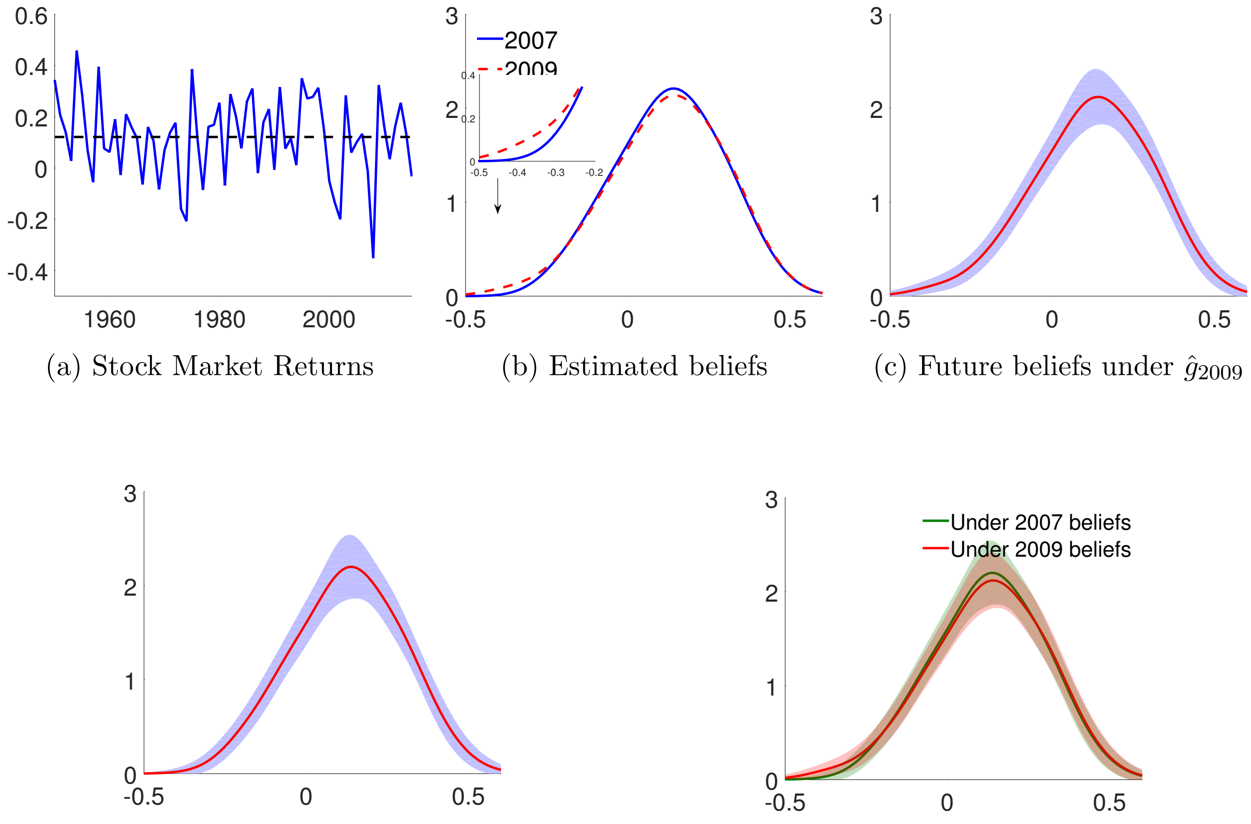
Figure 19: **Tail risk-related Google searches rose permanently after 2008.** *Search frequency for the terms 'financial crisis,' 'economic crisis,' and 'systemic risk' world-wide, from December 2003 - September 2016. Each series is normalized so that the highest intensity month is set to 100. Source: Google trends.*

D.2 Stock market

One question that often arises is whether other unusual events, such as the large stock market drop in 2008, might trigger a persistent economic response. Here, we illustrate what belief revisions would look like for agents learning about the distribution of stock returns. Of course, we acknowledge that this is not the driving force in our model. It is only intended to further illustrate possible future applications of our persistence mechanism.

Figure 20 shows the belief revision after observing 2008-09 equity returns, and the distribution of future beliefs under two different assumptions about the true distribution of shocks. Annual returns 1950-2009 come from Robert Shiller's website.

The figure shows that the negative equity returns during 2008-09, while large, were not all that unusual. The stock market has plunged many times. Seeing one more drop was not



(a) Stock Market Returns

(b) Estimated beliefs

(c) Future beliefs under $\hat{g}_{2009}$

(d) Future beliefs under $\hat{g}_{2007}$

(e) Future beliefs under $\hat{g}_{2007}$ and $\hat{g}_{2009}$

Figure 20: **Estimated beliefs about stock market returns.**
The first panel shows the realized returns on the stock market. The second panel shows the estimated kernel density for 2007 and 2009. The third panel shows the mean belief (along with a 2 standard deviation band) in 2039 (computed by simulating data for the period 2010-2039 using the estimated distribution in 2009).

unusual enough to change beliefs by much. We conclude that while stock returns can also generate some persistence through belief updating, this force is not a likely candidate for the recent stagnation, relative to the capital quality shock, because the downturn in stock prices was less unusual.

D.3 Returns during the Great Recession

Not all authors agree that the Great Recession was an unusual event. For example, Gomme et al. (2011) present a series for returns on capital that show adverse realizations for 2008-09 that are not as extreme as our measures. The difference stems from their measurement strategy. To compute capital gains, they use data from the NIPA, which values non-residential capital (structures, equipment and software) at *replacement* cost. During 2008-09, we saw massive declines in the *market value* (particularly, for commercial real estate), even though the replacement cost of structures fell only modestly. While appropriate for their purposes, these NIPA measures miss one of the unusual aspects of the Great Recession – large declines in the market value of business capital, notably commercial real estate.

D.4 De-trending

Our learning mechanism generates persistent movements in aggregate variables after extreme events. Therefore, in order to make a meaningful comparison with the data, the choice of the right de-trending procedure for the data is very important. We use a log-linear trend, which removes only the lowest-frequency (permanent) part of the series. A common approach in business cycle analysis is to non-linear filters (like the Hodrick-Prescott filter), which take out more of the persistent movements in the series. By design, what is left will not have much persistence left. In figure 21, we illustrate this using aggregate non-residential investment (other aggregate series show very similar patterns). As the graph reveals, the trend component of the HP filter (smoothing parameter 100) picks up some of the deviation from the linear trend. Given that our focus is on low-frequency or persistent components, a linear detrending procedure seems most appropriate.

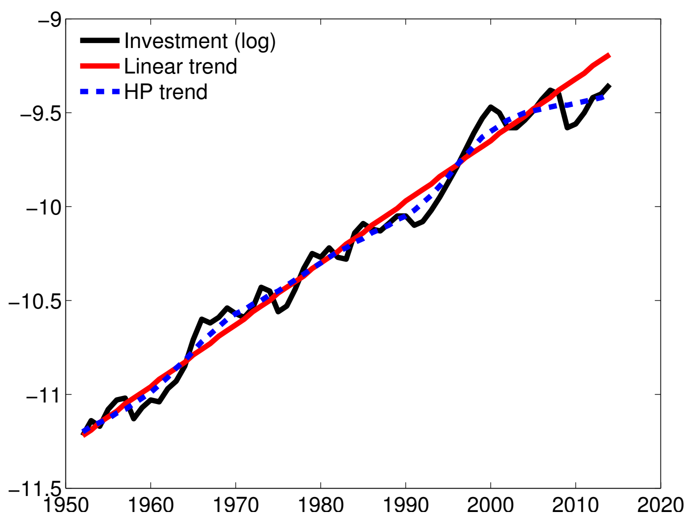


Figure 21: Non-residential Investment, with log-linear and HP trends.

D.5 TFP

While a productivity slowdown may have contributed to low output, it does not explain the timing or the rise in tail risk indicators. Figure 22 shows the time series of raw total factor productivity, constructed as $dTFP_t = dY_t - \alpha_t dK_t - (1-\alpha_t)(dHours_t + dLQ_t)$ from Fernald

(2014). When we examine instead utilization-adjusted TFP, we find a slight decline during the recession, but a decline that is will within two-standard deviation bands of the distribution of TFP changes. Productivity did not have a precipitous decline that could be considered a tail event.

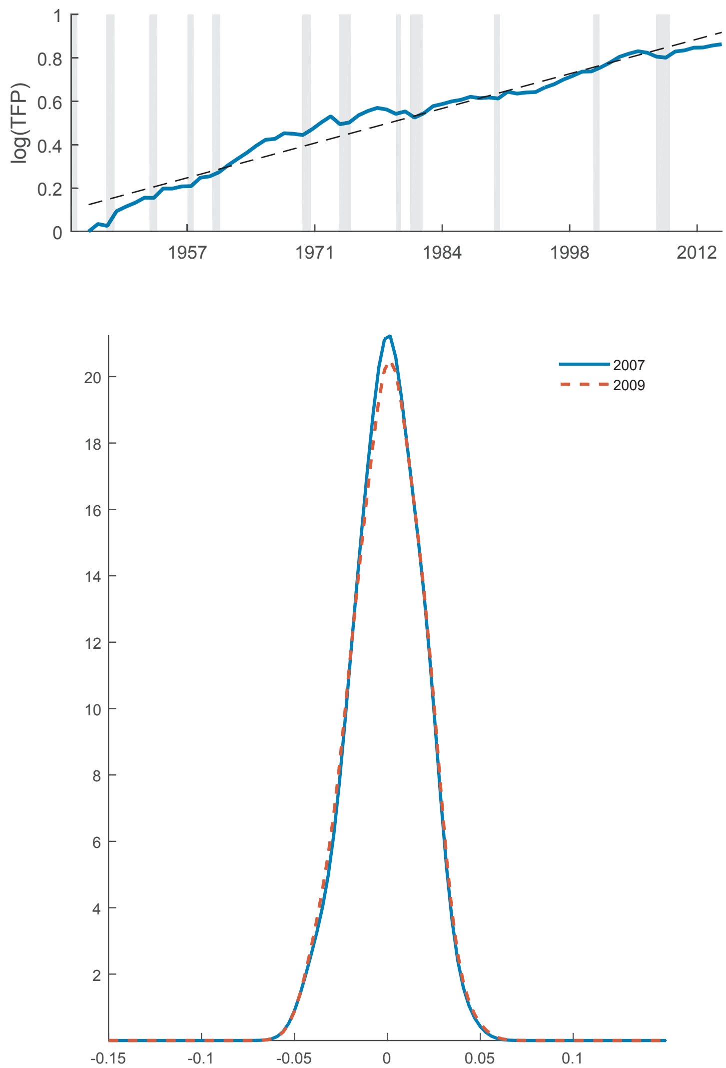


Figure 22: **Productivity.**