Understanding Uncertainty Shocks and the Role of Black Swans

Anna Orlik and Laura Veldkamp [1]

July 15, 2014

[1]Please send comments to anna.a.orlik@frb.gov and lveldkam@stern.nyu.edu. We are grateful for comments from NBER EF&G meetings, Wharton, Yale SOM, Banque de France, HEC, Bank of England, Toulouse School of Economics, U. Quebec at Montreal, Conference on Macroeconomic Uncertainty at Dallas Federal Reserve, the NBER Summer Institute forecasting group, NYU Alumni Conference, North American and European meetings of the Econometric Society, AEA meetings, Becker-Friedman Institute Policy Uncertainty Workshop, Federal Reserve Bank of Boston, the NYU macro lunch, and also to Nick Bloom, Tom Stark, our EF&G discussant Jennifer La'O and our AEA discussant, Lars Hansen. Many thanks to Isaac Baley, David Johnson, Callum Jones and Nic Kozeniauskas for outstanding research assistance. We acknowledge grant assistance from the Stern Center for Global Economy and Business. The views expressed herein are those of the authors and do not necessarily reflect the position of the Board of Governors of the Federal Reserve or the Federal Reserve System.

Abstract

Many recent theories postulate that economic and financial fluctuations result from changes in uncertainty, confidence, sentiment or the perceived probability of a disaster. One reason that belief-based theories have not eclipsed the standard paradigm is that, without direct evidence about what agents believe, many question whether beliefs fluctuate cyclically and why. This paper asks: What if agents form beliefs like a Bayesian econometrician? Each quarter, they use real-time RGDP data to re-estimate their forecasting model, forecast next quarter's growth, and estimate their forecast's standard deviation (uncertainty). When the forecasting model admits only normally-distributed outcomes, we find small, acyclical changes in uncertainty. But when agents can estimate a parameter that regulates skewness, uncertainty fluctuations become large and counter-cyclical. The reason is that small changes in estimated skewness whip around probabilities of unobserved tail events (black swans). The resulting forecasts resemble those of professional forecasters. Our real-time measure of GDP forecast uncertainty reveals that changes in the risk of a black swan explain most of the shocks to uncertainty.

A fruitful emerging literature has taught us that shocks to uncertainty can explain asset price fluctuations, business cycles and financial crises. But these theories would have stronger predictive power if we understood where these shocks come from. Instead, researchers typically assume that agents become more uncertain because they suddenly know, for certain, that future economic shocks will be more volatile. But real people do not know what the true distribution of economics is, when it changes, or by how much. They observe economic information and estimate the probabilities of outcomes. Much of their uncertainty comes from not knowing if their estimates are correct. This paper argues that equating uncertainty with stochastic volatility misses most uncertainty shocks. We use real-time data and develop a new technique for non-linear Bayesian forecasting to measure the uncertainty (conditional standard deviation) that arises from not knowing the true macro model. A combination of statistical and probability theory results explain why uncertainty is more variable and more counter-cyclical than stochastic volatility alone. The results uncover the following recipe for large uncertainty shocks: A long sequence of mild positive realizations, followed by one or more negative observations causes the forecaster's skewness estimate to become more negative. When skewness dips, the probability of left-tail events (black swans) surges. Because black swan events are so far from the mean forecast, changes in their probability have an outsized effect on the standard deviation, which is uncertainty. In short, it is not the variance of the realized data, but rather the time-varying risk of the unobserved black swans that explains most fluctuations in uncertainty.

We illustrate and quantify these uncertainty effects with a forecasting model that has two key features: 1) Outcomes are not conditionally normally distributed, and 2) agents use new data to re-estimate parameters that govern the distribution's higher moments, such as skewness. For each quarter, we use all U.S. GDP growth data that was available at that date to estimate the forecast model, update the forecast, and compute uncertainty. We define macroeconomic uncertainty as the standard deviation of next-period GDP growth y_{t+1}, conditional on all information observed through time t: $Std[y_{t+1}|\mathcal{I}_t]$. We use this definition because in most models, this is the theoretically-relevant moment. When there is an option value of waiting, forecasts with a higher conditional variance (larger expected forecast error) raise the value of waiting to observe additional information. We use the

resulting estimates to study how uncertainty changes and why.

This conceptually simple econometric exercise makes three contributions. (1) It provides a unified framework to explore the origins of and connections between uncertainty shocks, news shocks (changes in the forecasts of future outcomes) and disaster risk. These strands of the literature have evolved separately and have all suffered from the criticism that the right beliefs can rationalize almost any economic outcome. Allowing all three shocks to arise from observed macro outcomes offers the prospect of a unified information-based macro theory and a way to discipline the shocks to beliefs. (2) The results teach us about a natural amplification mechanism. They show how small economic events can trigger large changes in beliefs. It is not obvious that parameter learning would amplify shocks. Because most macro data is announced only quarterly and is highly persistent, parameter learning is a slow, gradual process. Thus, one might think that learning would make uncertainty shocks smoother than changes in volatility. Instead, we find that the opposite is true. This finding complements models that rely on large, counter-cyclical shocks to uncertainty to generate interesting economic and financial effects. (3) The results replicate the puzzling -1/2% bias observed in professional forecasts of GDP growth. Our theoretical results use a change-of-measure argument to prove that the combination of parameter uncertainty and skewness produces such a bias. When the estimated model matches the degree of skewness observed in the GDP growth data, it also matches the size of the forecast bias. The finding resolves a puzzle in the forecasting literature. It also produces beliefs that look similar to what an ambiguity-averse agent might report. But, just as importantly, this evidence suggests that the model accurately describes how people form beliefs.

To build up intuition for how parameter learning creates uncertainty shocks, we start (section 1) with one of the simplest forecasting models: a linear model of GDP growth with a continuous hidden state and homoskedastic innovations. We hold the volatility of the innovations fixed so that we can isolate the effect of parameter uncertainty shocks. Each period t, our forecaster observes time-t GDP growth and uses Bayesian updating and the complete history of GDP data to estimate her model and forecast GDP growth in $t+1$. In such a setting, an agent with known model parameters faces no uncertainty shocks because the estimated model is a homoskedastic one. But the forecaster with parameter uncertainty experiences changes in the conditional variance of her forecasts. These are uncertainty

shocks. The source of uncertainty shocks is surprises: events that are improbable under the previous period's estimated model. If an event occurs that a model says is improbable, standard errors on the parameter estimates often rises, estimated shock variances go up and uncertainty increases. While this linear model illustrates the possibility that parameter learning can create uncertainty shocks, those shocks are small and not counter-cyclical.

Next, we relax the assumption that the forecasting model generates normal outcomes. Instead, we use a linear model with an exponential twist (section 2). The parameters of the exponential twist regulate the skewness of outcomes. Because skewness is hard to learn in small samples, learning does not converge quickly. The resulting uncertainty series remains both volatile and counter-cyclical throughout the sample. The main source of uncertainty fluctuations is something we call "black swan risk," which is the conditional probability of an extreme growth realization that has not been observed. When the forecasting model implies a normal distribution of outcomes, the probability of an n-standard-deviation event is constant. But when we allow our forecaster to estimate a non-normal model, the probability of negative outliers can fluctuate. A new piece of data can lead the forecaster to estimate more negative skewness, which makes extreme negative outcomes more likely and raises uncertainty. When we apply this model to GDP data, we find that 75% of the variation in uncertainty can be explained by changes in the estimated probability of black swans.

The skewed forecasting model appears to be a plausible model of belief formation because it matches an important feature of professional economic forecasts: The average forecast is nearly half a percent lower than the average GDP growth realization. This bias has been a puzzle in the forecasting literature because an unbiased forecaster with a linear model and more than sixty years of data should not make such large systematic errors. We offers a new explanation for this forecasting puzzle: Forecast bias arises from rational Bayesian updating when forecasters believe outcomes have negative skewness and are uncertain about model parameters. While this bias might prompt one to use another estimation procedure, keep in mind that the objective in this paper is to describe a belief-formation process. The fact that our model has forecasts that are just as biased as professional forecasts suggests that Bayesian estimation might offer a good approximation to human behavior.

Section 2.6 investigates how volatility changes and parameter learning interact. To do that, we estimate the same model as before, but with two volatility states and Markov transitions between the states. Adding stochastic volatility makes uncertainty shocks 1/3rd larger on average. It also helps the model's performance in two key respects. First, it prevents a downwards trend in uncertainty. When all parameters are constant, uncertainty trends down partly because parameters are being better estimated over time, but mostly because the 70's and early 80's were much more volatile times for real GDP than the 90's and 2000's. So, the forecaster revises down the variance parameters over time and uncertainty trends down. When there are two volatility regimes, the forecaster infers that the 70's and 80's were a high volatility regime and that the regime switches in the mid 80's. As a result, the variance parameter estimates are more stable. Second, this model produces a larger surge in uncertainty during the financial crisis. With constant volatility, uncertainty rises slightly. But upon seeing a few pieces of highly-volatile data, the stochastic volatility forecaster quickly shifts probability weight to the high-volatility regime, causing uncertainty to spike.

Section 3 compares our model-based uncertainty series to commonly-used uncertainty proxies and find that it is less variable, but more persistent than the proxy variables. The most highly correlated proxies are the price of a volatility option (VIX) and the Baker, Bloom and Davis (2012) uncertainty index.

Our message is that understanding and measuring economic uncertainty requires relaxing the full-information assumptions of rational expectations econometrics. In such a full-information world, agents have no uncertainty about what the distribution of economic outcomes is. Their only uncertainty is about what realization will be drawn from a known distribution. To measure the uncertainty of such a forecaster, it makes sense to estimate a model on as much data as possible, take the parameters as given, and estimate the conditional standard deviation of model innovations. This is what stochastic volatility estimates typically are. But in reality, the macroeconomy is not governed by a simple, known model and we surely don't know its parameters. Instead, our forecast data (from the Survey of Professional Forecasters or SPF) suggests that forecasters estimate simple models to approximate complex processes and constantly use new data to update beliefs. Forecasters are not irrational. They simply do not know the economy's true data-generating process.

In such a setting, uncertainty and volatility can behave quite differently. Our findings teach us that one important source of financial and business cycle fluctuations may be changes in beliefs that come from learning about the distribution of economic outcomes.

Related literature A new and growing literature uses uncertainty shocks as a driving process to explain business cycles (e.g., Bloom, Floetotto, Jaimovich, Sapora-Eksten, and Terry (2012), Basu and Bundick (2012) Christiano, Motto, and Rostagno (2012), Bianchi, Ilut, and Schneider (2012), or Bachmann and Bayer (2012b) which argues against the role of uncertainty), to explain investment dynamics Bachmann and Bayer (2012a), to explain asset prices (e.g., Bansal and Shaliastovich (2010), Pastor and Veronesi (2012)), and to explain banking panics (Bruno and Shin, 2012). These papers are complementary to ours. We explain where uncertainty shocks come from, while these papers trace out the economic and financial consequences of the shocks.

A small set of related theories explain why uncertainty fluctuates using nonlinearities in a production economy (Van Nieuwerburgh and Veldkamp (2006), Fajgelbaum, Schaal, and Taschereau-Dumouchel (2013), Jovanovic (2006)) or active experimentation (Bachmann and Moscarini (2012)). In Nimark (2012), the key assumption is that only extreme events are reported. Thus, the publication of a signal reveals that the true event is extreme, which raises uncertainty. Our model differs because it does not depend on an economic environment, only on a forecasting procedure. Our uncertainty shocks arise from agents applying Bayes law when parameters are unknown. Our paper also contributes a framework that connects uncertainty with disaster risk and news shocks, unifying the literature on the role of beliefs in macroeconomics.

Our exercise also connects with a set of papers that measure uncertainty shocks in various ways. Bloom (2009), Baker, Bloom, and Davis (2012), Stock and Watson (2012), Jurado, Ludvigson, and Ng (2013) Fernandez-Villaverde, Kuester, Guerron, and Rubio-Ramirez (2012), Fernandez-Villaverde, Guerron-Quintana, Rubio-Ramirez, and Uribe (2011), Nakamura, Sergeyev, and Steinsson (2012), Justiniano and Primiceri (2008), Born, Peter, and Pfeifer (2011) and Rossi and Sekhposyan (2014) document the properties of uncertainty shocks in the U.S. and in emerging economies, while Bachmann, Elstner, and Sims (2012) use forecaster data to measure ex-ante and ex-post uncertainty in Germany. While our paper also engages in a measurement exercise, we primarily contribute a quantitative

model of why such shocks arise.

The idea that the probability of rare events is time-varying also appears in models of discrete, rare events (e.g., Rietz (1998), Barro (2006), Wachter (2013) for asset pricing and Gourio (2012) for business cycles). However, these papers assume that there is a time-varying probability of a rare event and derive the consequences of that for asset prices. Our paper is a complement to this literature; it describes a learning process that can generate the necessary volatile, counter-cyclical beliefs about the probability of extreme, negative outcomes.

Our approach is inspired by two preceeding papers that both estimate Bayesian forecasting models to describe agents' beliefs. Cogley and Sargent (2005) use such a model to understand the behavior of monetary policy, while Johannes, Lochstoer, and Mou (2011) estimate a model of consumption growth to capture properties of asset prices. While the concept is similar, our use of a non-normal model, our measurement of uncertainty, and our study of the effects of learning about skewness distinguish our work from its predecessors. This approach is also motivated by Chen, Dou, and Kogan (2013), which critiques models that give agents knowledge of parameters that econometricians cannot identify and Hansen's Ely Lecture (Hansen, 2007) which advocates modeling agents who struggle with real-time model selection.

1 A Linear-Normal Forecasting Model

The purpose of the model is to explain why relaxing rational expectations and assuming that agents do not know the true distribution of outcomes with certainty opens up an additional source of uncertainty shocks. To isolate this new source of uncertainty shocks, we consider first a simple, homoskedastic model. If an agent knew the true model (they had rational expectations), their uncertainty would be only about the random realizations of future shocks. Since the variance of these shocks is assumed to be constant over time, there would be no uncertainty shocks.

We consider a forecaster who observes real-time GDP growth data, in every quarter and forecasts the next period's growth. The agent knows that GDP growth comes from a linear model with a hidden state, but does not know the parameters of this model. Each period, he starts with prior beliefs about these parameters and the current state, observes

the new GDP data, and updates his beliefs using Bayes' law.

1.1 Definitions

A model, denoted $\mathcal{M}$, has a vector a parameters θ. Together, $\mathcal{M}$ and θ determine a probability distribution over a sequence of outcomes y_t. Let $y^t \equiv \{y_\tau\}_{\tau=0}^t$ denote a series of data available to the forecaster at time t. In every model, agent i's information set $\mathcal{I}_{it}$ will include the model $\mathcal{M}$ and the history y^t of y observations up to and including time t. The state S_t, innovations, and the parameters θ are never observed.

The agent, who we call a forecaster and index by i, is not faced with any economic choices. He simply uses Bayes' law to forecast future y outcomes. Specifically, at each date t, the agent conditions on his information set $\mathcal{I}_{it}$ and forms beliefs about y_{t+1}. We call the expected value $E\left(y_{t+1}|\mathcal{I}_{it}\right)$ an agent i's *forecast* and the square root of the conditional variance $Var\left(y_{t+1}|\mathcal{I}_{it}\right)$ is what we call *uncertainty*. Forecasters' forecasts will differ from the realized growth rate. This difference is what we call a forecast error.

Definition 1. *An agent i's forecast error is the distance, in absolute value, between the forecast and the realized growth rate: $FE_{i,t+1} = |y_{t+1} - E[y_{t+1}|\mathcal{I}_{it}]|$.*

We date the forecast error $t+1$ because it depends on a variable y_{t+1} that is not observed at time t. Similarly, if there are N_t forecasters at date t, an average forecast error is

$$\bar{FE}_{t+1} = \frac{1}{N_t}\sum_{i=1}^{N_t} FE_{i,t+1}. \tag{1}$$

We define forecast errors and uncertainty over 1-period-ahead forecasts because that is the horizon we focus on in this paper. But future work could use these same tools to measure uncertainty at any horizon.

Definition 2. *Uncertainty is the standard deviation of the time-$(t+1)$ GDP growth, conditional on an agent's time-t information: $U_{it} = \sqrt{E\left[\left(y_{t+1} - E[y_{t+1}|\mathcal{I}_{it}]\right)^2 \middle| \mathcal{I}_{it}\right]}$.*

Volatility is the same standard deviation as before, but now conditional on the history y^t, the model $\mathcal{M}$ and the parameters θ:

Definition 3. *Volatility is the standard deviation of the unexpected innovations in y_{t+1}, taking the model and its parameters as given:*
$V_t = \sqrt{E\left[\left(y_{t+1} - E[y_{t+1}|y^t, \theta, \mathcal{M}]\right)^2 \middle| y^t, \mathcal{M}, \theta\right]}$.

If an agent knew the parameters (i.e., if $\mathcal{I}_{it} = \{y^t, \mathcal{M}, \theta\}$), then uncertainty and volatility would be identical. Thus, the only source of uncertainty shocks would be volatility shocks.

Many papers equate volatility, uncertainty and squared forecast errors. These definitions allow us to understand the conditions under which these are equivalent. Volatility and uncertainty are both ex-ante measures because they are time-t expectations of $t+1$ outcomes, which are time-t measurable. However, forecast errors are an ex-post measure because it is not measurable at the time when the forecast is made. Substituting definition 1 into definition 3 reveals that $U_{it} = \sqrt{E[FE^2_{i,t+1}|\mathcal{I}_{it}]}$. So, uncertainty squared is the same as the expected squared forecast error. Of course, what people measure with forecast errors is typically not the *expected* squared forecast error. It is an average of *realized* squared forecast errors: $\sqrt{1/N_t \sum_i FE^2_{i,t+1}}$.

1.2 The linear forecasting model

We begin by examining the following continuous hidden-state process for y_t[1]

$$y_t = \alpha + s_t + \sigma\varepsilon_{y,t} \quad (2)$$

$$s_t = \rho s_{t-1} + \sigma_s \varepsilon_{s,t} \quad (3)$$

where $\varepsilon_{y,t}$ and $\varepsilon_{s,t}$ are standard normal random variables independent of each other. In this model, the parameters are $\theta \equiv [\alpha, \rho, \sigma, \sigma_s]'$.

Information assumptions Each forecaster has an identical information set: $\mathcal{I}_{it} = \{y^t, \mathcal{M}\}$, $\forall i$. The model $\mathcal{M}$ is described by (2) and (3). The state s_t and the parameters θ are never observed.

To compute forecasts and the process for uncertainty, we use Bayesian updating. A forecast is a conditional expectation of next-period growth, where the expectation is taken

[1] We have also explored a version of the model with a discrete hidden state. The results are very similar and are reported in Appendix B.

over unknown parameters, states, and growth realizations. Using the law of iterated expectations, we can write this forecast as:

$$E\left(y_{t+1}|y^t\right) = \int\int\int y_{t+1} p\left(y_{t+1}|\theta, s_{t+1}, y^t\right) p\left(s_{t+1}|\theta, y^t\right) p\left(\theta|y^t\right) d\theta ds_{t+1} dy_{t+1} \quad (4)$$

The first probability density function, $p\left(y_{t+1}|\theta, s_{t+1}, y^t\right)$, is the probability of $t+1$ GDP growth, given the state and the parameters. From 2, we know that, conditional on the state, GDP growth is normally distributed with mean $\alpha + s_{t+1}$ and variance σ^2.

The second probability density function, $p\left(s_{t+1}|\theta, y^t\right)$, is the probability of a hidden state in a Kalman filtering system. It is also conditionally normally distributed. When the parameters are known, (2) and (3) form the observation and state equations of a Kalman filtering system. The following recursive equations describe the conditional mean and variance of the first two probability terms, jointly:

$$E[y_{t+1}|y^t, \theta, \mathcal{M}] = \rho E[y_t|y^{t-1}, \theta, \mathcal{M}] + \rho K_t y_t \quad (5)$$

where the term $K_t = K_t = (I + \sigma^2 Var[y_t|y^{t-1}, \theta, \mathcal{M}])^{-1}$ is the Kalman gain and the conditional variance of the estimate is

$$Var[y_{t+1}|y^t, \theta, \mathcal{M}] = \left((1-\rho^2)\sigma^2 + \sigma_S^2\right) I \quad (6)$$

Volatility is $\sqrt{Var[y_{t+1}|y^t, \theta, \mathcal{M}]}$, which is a constant. Constant volatility may or may not be a realistic feature of the data. But it is a helpful starting point because it will allow us to isolate the fluctuations in uncertainty that come from parameter uncertainty.

Finally, the third probability density function is the probability of the parameter vector θ, conditional on the t-history of observed GDP data. To compute posterior beliefs about parameters, we employ a Markov Chain Monte Carlo (MCMC) technique.[2] At each date t, the MCMC algorithm produces a sample of possible parameter vectors, $\left\{\theta^d\right\}_{d=1}^{D}$, such that the probability of any parameter vector θ^d being in the sample is equal to the posterior probability of those parameters, $p\left(\theta^d|y^t\right)$. Therefore, we can compute an approximation to any integral by averaging over sample draws: $\int f(\theta)p(\theta|y^t)d\theta \approx 1/D \sum_d f(\theta^d)$.

[2]More details are presented in the Appendix.

To estimate uncertainty, we compute these probability density terms and integrate numerically to get a forecast. In similar fashion, we also calculate $E\left(y_{t+1}^2|y^t\right)$. Applying the variance formula $Var\left(y_{t+1}|y^t\right) = E\left(y_{t+1}^2|y^t\right) - E\left(y_{t+1}|y^t\right)^2$, and taking the square root yields uncertainty: $U_t = \sqrt{Var\left(y_{t+1}|y^t\right)}$.

1.3 Data Description

There are two pieces of data that we use to evaluate and estimate our forecasting model. The first is real-time GDP data from the Philadelphia Federal Reserve. The variable we denote y_t is the growth rate of GDP. Specifically, it is the log-difference of the real GDP series, times 400, so that it can be interpreted as an annualized percentage change. We use real-time data because we want to accurately assess what agents know at each date. Allowing them to observe final GDP estimates, that are not known until much later, is not consistent with the goal. Therefore, y_t represents the estimate of GDP growth between the end of quarter $t-1$ and quarter t, based on the GDP estimates available at time t. Similarly, y^t is the history of GDP growth up to and including period t, based on the data available at time t.

We use the second set of data, professional GDP forecasts, to evaluate our forecasting models. We describe below the four key moments that we use to make that assessment. The data come from the Survey of Professional Forecasters, released by the Philadelphia Federal Reserve. The data are a panel of individual forecaster predictions of real US output for both the current quarter and for one quarter ahead from quarterly surveys from 1968 Q4 to 2013 Q4. In each quarter, the number of forecasters varies from quarter-to-quarter, with an average of 40.5 forecasts per quarter.

Formally, $t \in \{1, 2, \ldots, T\}$ is the quarter in which the survey of professional forecasters is given. Let $i \in \{1, 2, \ldots, I\}$ index a forecaster and $I_t \subset \{1, 2, \ldots, I\}$ be the subset of forecasters who participate in a given quarter. Thus, the number of forecasts made at time t is $N_t = \sum_{i=1}^{I} \mathbb{I}(i \in I_t)$. Finally, let y_{t+1} denote the GDP growth rate over the course of period t. Thus, if GDP_t is the GDP at the end of period t, observed at the start of quarter $t+1$, then $y_{t+1} \equiv ln(GDP_t) - ln(GDP_{t-1})$. This timing convention may appear odd. But we date the growth $t+1$ because it is not known until the start of date $t+1$.

1.4 Estimation and results: linear-normal model

Our forecaster needs prior distributions over all the parameters to start the updating process. We start with a flat prior, estimate each parameter on GDP growth data from 1947:2-1968:3, and use the mean and variance of this estimate as the mean and variance of prior beliefs. (See appendix for more details and prior estimation results.) Starting in quarter 4 of 1968, each period, the agent observes y_t and revisions of previous quarters' data and updates his beliefs about future GDP growth using (4). We start the estimation of the model in 1968:4 because this is the first quarter for which we have forecasts from the Survey of Professional Forecasters.

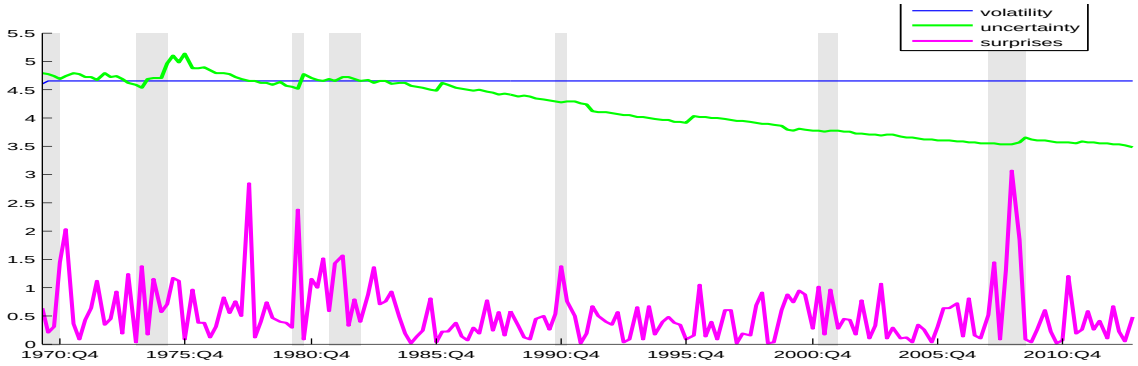


Figure 1: Uncertainty shocks in the linear model.
Volatility, uncertainty and surprises are calculated using (6), definition 2 and (7).

To understand what properties of our forecasts come from parameter uncertainty, Figure 1 compares our uncertainty to that of an agent who takes the model parameters as given (labeled θ known). To keep the results comparable, the parameters that are used to compute volatility are the same as the mean prior beliefs in our model.

The main takeaway is that parameter uncertainty generates uncertainty shocks, albeit small ones. With known parameters, stdev$(U_t) = 0$. When parameters are updated every period, stdev$(U_t) = 0.48$. What triggers these uncertainty shocks is primarily surprises. We define a surprise to be an event that was unlikely, given the previous period's estimated model:

$$Surprise_t = \frac{|y_t - E(y_t|y^{t-1})|}{U_{t-1}}. \tag{7}$$

Figure 1 plots this surprise measure against the uncertainty measure. What it reveals is that periods with large surprises are periods when uncertainty increases. But also that the uncertainty generated by the surprise persists, long after the surprise has passed.

Our results also expose three aspects of our forecasts that do not look realistic. 1) Our forecasters' uncertainty is not counter-cyclical (Correl(U_t,GDP) = 13%). Every common proxy for uncertainty is counter-cyclical and most theories use uncertainty to explain the onset of a recession. So, a forecasting model that fails to deliver this feature is suspect. 2) The model does not explain the low average forecasts of GDP observed in the professional forecaster data. The true average of GDP growth over 1968:Q4-2012:Q4 is 2.68%. The average professional forecast of GDP growth is 2.24%, almost half a percentage point lower.[3] This model fails to explain that gap. 3) There is obviously no dispersion in forecasts. There is only one forecaster. Yet, forecast dispersion is a prominent and interesting feature of the data. Later, we will introduce heterogeneous signals to examine its role.

In the next section, we examine a nonlinear forecasting model that remedies the first two problems: It generates counter-cyclical uncertainty, forecasts that are low on average, and larger uncertainty shocks.

2 A Skewed Forecasting Model

The linear-normal model of the previous section was a starting point. It is a commonly-used framework where we shut down all sources of uncertainty shocks besides our estimation uncertainty and could see that mechanism at work. But that model misses important features of the forecast data and its uncertainty shocks are modest. One reason that uncertainty varies so little is that all the random variables are normally distributed. The normal distribution has the unusual properties that the conditional variance is the same, no matter what the conditional mean is and that an n-standard deviation event is always equally unlikely. Since uncertainty is a conditional variance, the normal distribution shuts down much scope for changes in uncertainty. Therefore, we explore a model that has non-normal shocks. This skewed model is still too simplistic to be a state-of-the art forecasting

[3]This gap only arises in final GDP estimates. The average initial GDP announcement has 2.3% growth on average, in line with the forecasts. But if these initial announcements are themselves BEA forecasts of what the final GDP estimate will be, there is still a puzzle about why early estimates are systematically lower than final estimates.

model. For one, the forecaster only uses one source of information, past GDP growth, to forecast. But the simplicity of the model allows us to understand its mechanics and to better understand how parameter revisions affect the probabiliity of extreme events, why forecasts are downward biased, and what triggers large uncertainty shocks.

Moving away from normality, there is an infinite set of possibilities. But one natural choice is to focus on a distribution with skewness. Negative skewness is a salient feature of GDP data. In the real GDP (1968:Q4-2013:Q4) data we use for our forecasting model simulations, the skewness of GDP growth is -0.30. Skewness is also a feature of many models. Models where workers lose jobs quickly and find jobs gradually, or models where borrowing constraints amplify downturns are just a couple of examples of models that generate booms that are more gradual than crashes. Finally, we find that a model with skewness both does a better job of matching features of forecast data and generates much larger uncertainty shocks.

Updating non-normal variables is typically cumbersome. Combining this with parameter uncertainty typically requires particle filtering, which is possible, but slow. We make this problem tractable by doing a change of measure. The Radon-Nikodym theorem tells us that, for any measure g that is absolutely continuous with respect to a measure induced by a normal distribution, we can find a change-of-measure function f such that $g(x) = \int f(x)d\Phi(x)$, where Φ is a normal cdf. If we estimate such an f function, then we use f^{-1} to take data from a skewed distribution and transform it into normal data that we can then use the tools from the last section to learn about.

Of course, allowing a forecaster to explore the whole function space of possible f's is not viable. We focus the problem by considering a family of functions and allowing the forecaster to consider parameter estimates that govern the properties of the distribution. The change of measure function should have three desirable properties: 1) Its range is the real line; 2) it is monotone; and 3) it can be either globally concave or globally convex, depending on the estimated parameters. A class of transformations that satisfy this criteria is

$$f(\widetilde{X}_t) = c + b\ exp(-\widetilde{X}_t) \tag{8}$$

If we have estimates for b and c, we can do a change of variable: Use $f^{-1}(y_t)$ to transform GDP growth into a variable $\widetilde{X} = S_t + \sigma\epsilon_t$, which is a normally-distributed continuous

variable with a hidden persistent state. The we can write our skewed forecasting model as

$$\begin{aligned} y_t &= c + b\, exp\left(S_t + \sigma\epsilon_t\right) \\ S_t &= \rho S_{t-1} + \sigma_S \varepsilon_t \end{aligned}$$

where $\epsilon_t \sim N(0,1)$ and $\varepsilon_t \sim N(0,1)$. This change-of-variable procedure allows our forecaster to consider a family of non-normal distributions of GDP growth and convert each one into a linear-normal filtering problem with unknown parameters that can be solved using the same tools as in the previous section. The only additional complication is that the parameters b and c also need to be estimated.

The b parameter governs the mean of the $\widetilde{X}$ process. To see this, note that for $b < 0$, we can rewrite $b\, exp(-\widetilde{X}_t) = -exp(-\widetilde{X}_t + ln(|b|))$. To streamline our code, we simply remove the time-t sample mean of the $\widetilde{X}_t$ and set $b = -1$. After estimating the parameters of the mean-zero process, we add back in the sample mean. This approach is supported by the fact that when we have estimated b in less complex settings, we come up with consistently negative values and quantitatively similar estimates.

We start with priors (see Appendix) and use MCMC techniques to form beliefs about the ρ, σ, σ_s and c parameters. For each parameter draw θ_i from the MCMC algorithm, we compute $E[y_t|\mathcal{I}_t, \theta_i]$ and $E[y_t^2|\mathcal{I}_t, \theta_i]$. We average these expectations over all parameter draws and compute uncertainty as $U_t = E[y_t^2|\mathcal{I}_t] - E[y_t|\mathcal{I}_t]^2$.

The key feature of this model is that it produces a skewed distribution of outcomes and that the forecaster has to estimate parameters that govern the skewness. Skewness in this model is most sensitive to the b and c parameters because they govern the curvature of the transformation f of the normal variable. Any function with similar curvature, such as a polynomial or sine function, would have a similar mechanism. Our work in progress is forecasting with a concave, piecewise polynomial function. Results are available upon request.

2.1 Results: Uncertainty Shocks

Figure 2 compares the time series of uncertainty in both models. Column (3) of Table 1 shows that updating beliefs about the skewness of the GDP growth distribution has a large

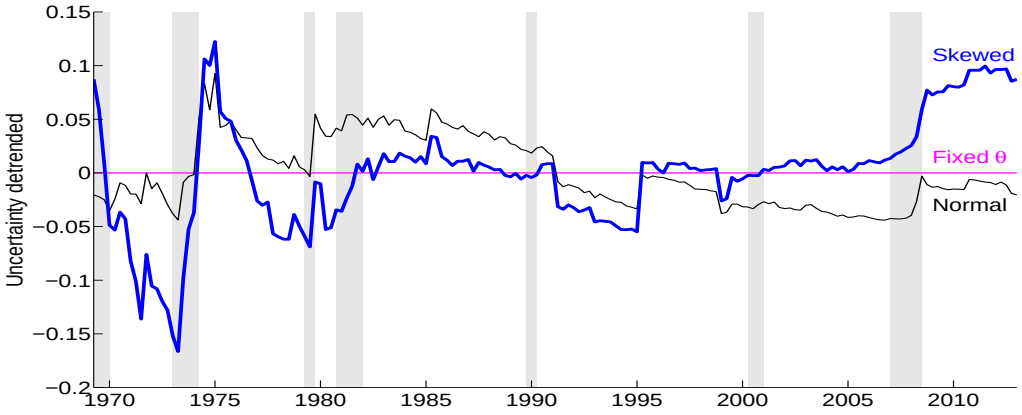


Figure 2: Uncertainty (U_t) implied by linear and skewed models in log deviations from trend.

effect on uncertainty. Such learning increases the average level of uncertainty by only 8%. But it amplifies uncertainty shocks. The standard deviation of the uncertainty series was 0.48% with normally-distributed outcomes and rises to 1.50% when our forecaster updates beliefs about skewness. One can interpret the magnitude of this standard deviation relative to the mean. A 1-standard deviation shock to uncertainty raises uncertainty 33% above its mean. That is quite a volatile process and offers a stark contrast to the relatively modest changes in volatility typically measured.

Since using growth rates of GDP is a form of trend-removal, it makes sense to correlate a stationary series with another stationary series. Therefore, we detrend volatility and uncertainty in order to discern the nature of their cyclical components (Table 1, middle panel). We remove the trend in uncertainty using log deviations from an exponential trend:

$$\tilde{U}_t \equiv ln(U_t) - ln(U_t^{trend}) \tag{9}$$

The resulting series, plotted in figure 2, reveals large, highly counter-cyclical uncertainty shocks. Not only is the level higher, uncertainty rose noticeably during each of the recessions since 1970.

Keep in mind that there is still no stochastic volatility in this model. To the extent that we believe that there are volatility shocks to GDP, this would create additional shocks to uncertainty, above and beyond those we have already measured. These series are not yet

model:	unc/vol	normal	skewed
Mean	U_t	4.20%	4.53%
	V_t	4.65%	4.01%
Std deviation	U_t	0.48%	1.50%
	V_t	0%	0.05%
Autocorrelation	U_t	0.99	0.97
	V_t	0	0.93
	Detrended uncertainty/volatility		
Corr($\tilde{U}_t$, $E_t[y_{t+1}]$)		0.04	-0.78
Corr($\tilde{V}_t$, $E_t[y_{t+1}]$)		0	-0.74
	Forecast properties		
	data	normal	skewed
Mean forecast	2.29%	2.82%	2.27%
Mean $\lvert F\,Err\rvert$	1.87%	2.25%	2.51%
Std forecast	2.25%	1.17%	0.64%
Std $\lvert F\,Err\rvert$	1.46%	2.17%	2.39%

Table 1: **Properties of model uncertainty series.** Forecasts are computed using equation (4). Forecast error is (forecast - final GDP growth). Uncertainty, denoted U_t is computed as in Definition 2. Volatilities, denoted V_t, are computed as in Definition 3 assuming that the parameters θ are known and equal to the the mean posterior beliefs at the end of the sample for the parameter learning models.

a complete picture of macroeconomic uncertainty. Instead, they are a look at what part of uncertainty is missed when we just measure volatility.

Clearly, the model is not forecasting GDP as accurately as the forecasters in the survey of professional forecasters do (Table 1, bottom panel). However, this is a problem that we can remedy, without changing our main message. The forecasts in the model are based only on prior GDP releases. In reality, forecasters have access to other sources of data that improve the accuracy of their forecasts. The fact that the model produces a forecast error that is too large and too volatile reflects this problem. Kozeniauskas, Orlik, and Veldkamp (2014) develop a non-normal forecasting model with additional forecasting information that remedies this forecast accuracy problem and still generates large, counter-cyclical uncertainty shocks.

Finally, uncertainty is very persistent. It is highest at the start of the sample, when data is most scarce, and then slowly decays over the rest of the sample. As noted by Collin-Dufresne, Johannes, and Lochstoer (2013), the persistent uncertainty process comes

from the nature of learning: A single large shock to GDP growth results in a quick reevaluation of the parameter and model probabilities. Because beliefs about parameters are martingales, these parameter revisions act as permanent, non-stationary shocks even when the underlying shock is transitory.

2.2 Skewness Shocks and Time-Varying Black Swan Risk

To understand why uncertainty varies so much, it is helpful to look at the probability of tail events. Since our estimated probability distribution is negatively skewed, negative outliers are more likely than positive ones. For a concrete example, let's consider the probability of a particular negative growth event. The mean of GDP growth is 2.68%, while its standard deviation is 3.32%. If GDP growth were normally distributed, then $y_{t+1} \leq -6.8\%$ would be a 1-in-100-year event (Pr= 0.0025 quarterly). Let's call this rare event a black swan.

$$Black\,Swan\,Risk_t = Prob[y_{t+1} \leq -6.8\%|\mathcal{I}_t]. \tag{10}$$

The correlation between black swan risk and uncertainty is 97% (75% for the detrended series). This illustrates that uncertainty shocks arise in times when the estimated probabilities of extreme events change. Our model suggests that uncertainty builds up gradually over time as more and more unusual observations are realized.

When the distribution of GDP growth is non-normal and states and parameter estimates change over time, the probability of this black swan event fluctuates. Figure 3 plots the estimated black swan probability each period. The black swan probability varies con-

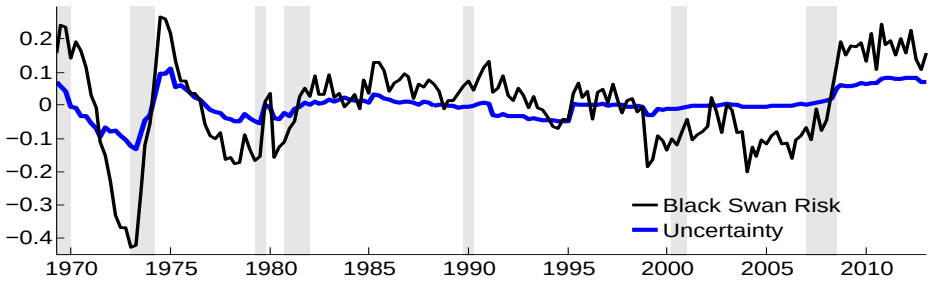


Figure 3: When the probability of a black swan event is high, uncertainty is high. *Black Swan Risk* is defined in (10). It is evaluated in the model where the skewness parameter c is updated every period. Both black swan risk and the uncertainty series are exponentially detrended.

siderably. It reaches a high of 18% in early 1970. Leading up to the 2008 financial crisis, the black swan probability rose from 3.5% in 2007 Q1 to over 4.6% in Q3 of 2009.

These results teach us that when we include parameter uncertainty in our notion of economic uncertainty, and we consider a model with skewed outcomes, then most changes in uncertainty coincide with changes in the estimated probability of rare events. Most of these uncertainty shocks were not present when we did not allow the forecaster to update his skewness beliefs. When we allow for learning about skewness, new pieces of data cause changes in the skewness estimates. Tail event probabilities are very sensitive to this skewness parameter. When the probability of extreme events is high, uncertainty is high as well.

This explanation raises the question: What types of data realizations make estimated skewness more negative, increase black swan risk, and thereby generate uncertainty shocks? We find two types of episodes that set up large uncertainty shocks. The first is simply a large negative GDP growth realization. When a negative outlier is observed, the forecaster revises skewness to be more negative and increases the estimated variance of shocks, both of which cause the probability of a black swan event and uncertainty to rise. This is what happens in 2008 and in the early 1980's. But there is a second, more subtle, cause of uncertainty shocks that comes from a sequence of mild positive GDP growth realizations in a row. These observations cause the forecaster to increase the estimated mean of the distribution. When the mean increases, the existing negative outlier data points become further from the mean. Because the previously-observed negative realizations are more extreme, the estimate of skewness rises and the probability of rare negative events can rise as well. This is what happens in the early 1970's. A sequence of positive growth realizations causes a rise and then a fall in uncertainty. But the persistence of the high estimated skewness set the stage for the large rise in uncertainty in the second half of the 1970's. This mechanism provides one explanation for why uncertainty seems to rise particularly at the end of long spells of consistently positive growth.

2.3 Negative Skewness as a Force for Counter-Cyclical Uncertainty

One way of understanding the cyclical effect skewness has on uncertainty is by thinking about the skewed distribution as a non-linear transformation of a normal distribution. The

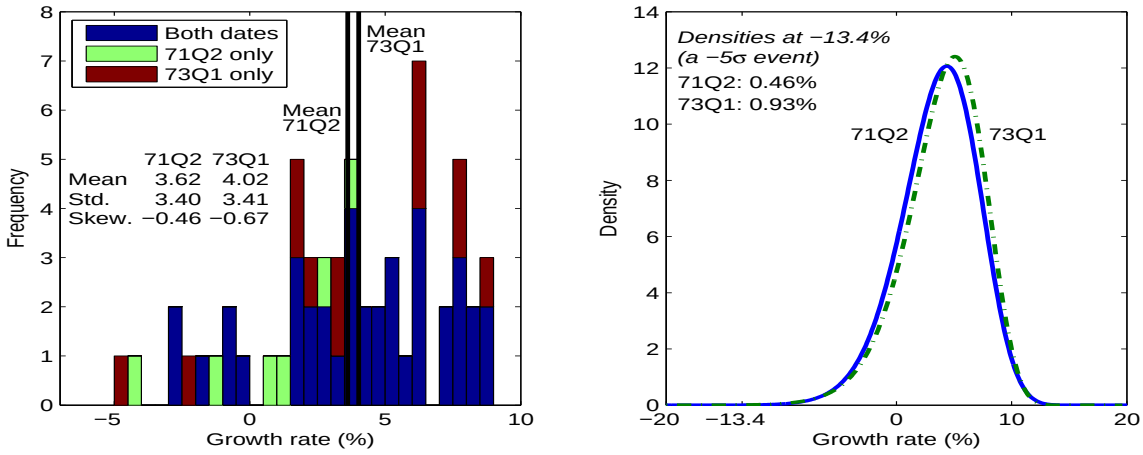


Figure 4: An example of a positive growth episode that increased the estimated mean, skewness and black swan probability.

transformation has no economic interpretation. It does not represent a utility function, production function or anything other than an estimated change-of-measure function that regulates the skewness of outcomes.[4] But since many problems in economics use normal shocks and compute means and variances of concave functions of these shocks, we can leverage that intuition here to understand the role of skewness. (See Albagli, Hellwig, and Tsyvinski (2013) for a similar approach.) The following result shows that a concave transformation of a variable with a normal probability density function results in a variable whose distribution has negative skewness.

Lemma 1. *Suppose that y is a random variable with a probability density function $\phi(g^{-1}(y))$ where ϕ is a standard normal density and g is an increasing, concave function. Then, $E[(y - E[y])^3] < 0$.*

The unconditional distribution of GDP growth rates is negatively skewed. Therefore, when we estimate the change of measure function that maps a normal variable x into

[4]Although this paper does not try to explain the negative skewness of outcomes, many other theories do. Negative skewness can arise when the economy is functioning very well (high $\widetilde{X}_t$), then improving its efficiency results in a small increase in GDP. But if there is a high degree of dysfunction or inefficiency (low $\widetilde{X}$), then the economy can fall into depression. Many models generate exactly this type of effect through borrowing or collateral constraints, other financial accelerator mechanisms, matching frictions, or information frictions. Even a simple diminishing returns story could explain such skewness.

GDP growth (8), we consistently find that the coefficient b is negative, meaning that the transformation is increasing and concave. A concave transformation of a normal variable puts more weight on very low realizations and makes very high realizations extremely unlikely. In other words, the concave transformation creates a negatively-skewed variable.

Breaking the probability density into a normal and a concave function is helpful because it allows us to understand where counter-cyclical uncertainty comes from. We can use the Radon-Nikodym theorem to characterize the conditional variance of a skewed variable as the conditional variance of a normal variable, times a Radon-Nikodym derivative.

$$Var[y_{t+1}|y^t] = \int (y_{t+1} - E[y_{t+1}])^2 f(y_{t+1}|y^t) dy_{t+1}$$

If $f(y_{t+1}|y^t) = f(g(x_{t+1})|y^t) = \phi(x_{t+1}|x^t)$, then by the Radon-Nikodym theorem,

$$Var[y_{t+1}|y^t] = \int (x_{t+1} - E[x_{t+1}])^2 \frac{dg}{dx} \phi(x_{t+1}|x^t) dx_{t+1}$$

$$Var[y_{t+1}|y^t] = E\left[\frac{dg(x_{t+1})}{dx} \middle| x^t \right] Var[x_{t+1}|x^t] + cov(\frac{dg}{dx}, (x_{t+1} - E[x_{t+1}])^2)$$

The variance of the normal variable x_{t+1} obviously depends on its history x^t, but it is not affected by what the expected value of x_{t+1} is. Normal variables have the property that their conditional variance is the same throughout the state-space. Conditional variance is not mean-dependent. That is not true of the skewed variable y. Because g is an increasing, concave function, dg/dx is largest when x is low and falls as x rises. This tells us that $Var[y_{t+1}|y^t]$ is largest when $E[y_{t+1}|y^t]$ is low and falls as the expected GDP growth rate y_{t+1} rises. This is the origin of counter-cyclical uncertainty. It arises naturally if a variable has a negatively-skewed distribution that can be characterized as a concave transformation of a normal variable.

Figure 5 illustrates why uncertainty is counter-cyclical. The concave line is a mapping from x into GDP growth, y. The slope of this curve is a Radon-Nikodym derivative. A given amount of uncertainty is like a band of possible x's. (If x was uniform, the band would represent the positive-probability set and the width of the band would measure uncertainty about x.) If that band is projected on to the y-space, the implied amount of uncertainty about y depends on the state x. When x is high, the mapping is flat, and the resulting

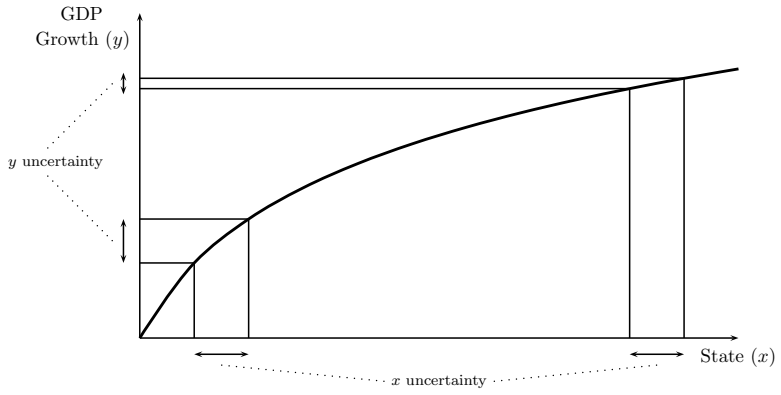


Figure 5: Nonlinear change of measure and counter-cyclical uncertainty. A given amount of uncertainty about x creates more uncertainty about y when x is low than it does when x is high.

width of the band projected on the y-axis (y uncertainty) is small. When x is low, the band projected on the y axis is larger and uncertainty is high. This mechanism for generating coutner-cyclical uncertainty is related to Straub and Ulbricht (2013), except that in his model, the concave function arises from assumptions about an economic environment. In this paper, the concave function is estimated and captures only the fact that GDP growth data is negatively skewed.

Learning about skewness causes this concave curve to shift over time. When a negative outlier is observed, the estimated state falls and estimated skewness becomes more negative. More skewness translates into more curvature in the change of measure function. Combined with a low estimated state, this generates even more uncertainty. Thus, bad events trigger larger increases in uncertainty. This is reflected is the more negative correlation between forecasts and uncertainty in the learn skew model in Table 1.

2.4 Why Skewness and Model Uncertainty Lower Forecasts

Aside from generating larger uncertainty shocks, the model with skewness also explains the low GDP growth forecasts in the professional forecaster data. The average forecast is 2.27% in the model and 2.29% in the forecaster (SPF) data.[5] These forecasts are puzzling because the average GDP growth rate is 2.68%. It cannot be that over 70 years of post-war

[5]Elliott and Timmermann (2008) confirm our finding of forecast bias. They argue that stock analysts over-estimate earnings growth and the Federal Reserve under-estimates GDP growth.

history, forecasters have not figured out that the sample mean is 0.4% higher than their forecasts on average. Our next result shows that these low forecasts are entirely rational for a Bayesian who believes that outcomes are negatively skewed and faces parameter uncertainty.

Lemma 2. *Suppose that y is a random variable with a probability density function f that can be expressed as $f(y|\mu,\sigma) = \phi((g^{-1}(y)-\mu)/\sigma)$ where ϕ is a standard normal density and g is a concave function. Let the mean of y be $\bar{y} \equiv \int y f(y|\mu,\sigma)dy$. A forecaster does not know the true parameters μ and σ, but estimates probability densities $h(\mu)$ and $k(\sigma)$, with means μ and σ. The forecaster uses these parameter densities to construct a forecast: $\hat{y} \equiv \int\int\int y\, f(y|\mu',\sigma')\, h(\mu')\, k(\sigma')\, dy\, d\mu'\, d\sigma'$. Then $\hat{y} < \bar{y}$.*

The logic of the result is the following: If GDP growth is a concave transformation of a normal underlying variable, Jensen's inequality tells us that expected values will be systematically lower than the median realization. But by itself, Jensen's inequality does not explain the forecast bias because the expected GDP growth and the mean GDP growth should both be lowered by the concave transformation (see figure 6, left panel). It must

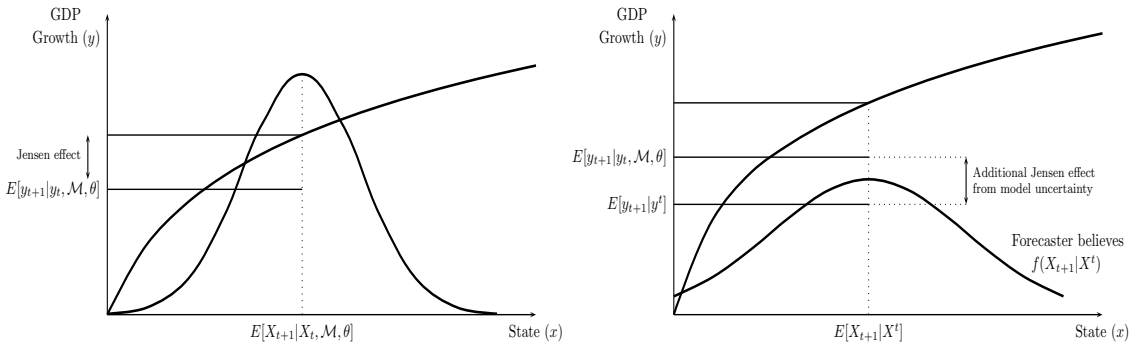


Figure 6: Explaining why average forecasts are lower than mean GDP growth. The result has two key ingredients: The forecaster faces more uncertainty than he would if he knew the true distribution of outcomes, and a Jensen inequality effect from the concave change of measure.

be that there is some additional uncertainty in expectations, making the Jensen inequality effect larger for forecasts than it is for the unconditional mean of the true distribution (see figure 6, right panel). This would explain why our results tell us that most of the time the sample mean is greater than the average forecast. If the agent knew the true parameters, he would have less uncertainty about y_{t+1}. Less uncertainty would make the Jensen effect

smaller and raise his estimate of y_{t+1}, on average. Thus, it is the combination of parameter uncertainty and a skewed distribution that can explain the forecast bias.[6]

2.5 Convergence and the Downward Trend in Uncertainty

Since the true parameters in this model are constant, eventually agents will learn them, if the model is correctly specified. Even in our 44-year sample, there is evidence of convergence. There is a downward trend in uncertainty, some of which comes from the decline in the uncertainty about the parameter values. Between 1970 and 2013, uncertainty falls from 6.2% to 3.5%. Does this decline imply that all parameter uncertainty should be resolved in the near future and these effects will disappear? There are three reasons why parameter uncertainty would persist.

First, our forecasting model is clearly not a complete or correct description of the macroeconomy. Our simple specification represents the idea that people use simple models to understand complex economic processes. Bayesian learning converges when the model is correctly specified. But when the estimated model and the true data-generating process differ, there is no guarantee that parameter beliefs will converge to the truth. Even as the data sample becomes large, parameter beliefs can continue to fluctuate, generating uncertainty shocks.

Second, much of the trend decline in uncertainty comes from lower estimated volatility. The mean estimate of the transitory shock variance (σ^2) falls by 46% between 1970 Q1 and 2013 Q4. The mean estimate of variances decline because GDP growth just becomes less volatile in the second half of the sample and agents react to that by revising down their estimates of the variance parameters. Lower innovation variance also reduces uncertainty.

Finally, simply adding time-varying parameters can avoid convergence. If we assume that some or all of the parameters drift over time, then beliefs about these parameters will continue to change over time, just like beliefs about the hidden state s_t fail to converge over time. One example of a model with time-varying parameters is a stochastic volatility model. We turn to these results next.

[6]When we compare SPF forecasts to first-release GDP data, the difference in means disappears. But that is consistent with an interpretation of this first release as itself a forecast. The first release may be lower, on average because it is more uncertain and the added uncertainty, combined with the non-linear model, lowers the average initial announcement.

2.6 Adding Stochastic Volatility

So far, we have explored homoskedastic models, in order to isolate the uncertainty shocks that come from parameter learning. But both changes in volatility and in parameter estimates can contribute to uncertainty shocks. To quantify the contribution of each, we estimate a model with stochastic volatility and parameter learning. The result is an uncertainty series that is a bit more volatile than before, but without the downward trend in uncertainty and with a larger spike in uncertainty around the time of the financial crisis.

We make a small change to the forecasting model (9): We add a subscript to the standard deviation σ_S. The variance of the persistent hidden state innovations now takes on one of two values $\sigma_{St} \in \{\sigma_{SL}, \sigma_{SH}\}$. State changes are governed by a Markov transition matrix whose entries are also estimated by our forecaster. The agent starts with prior beliefs that low volatility is 0.15%, high volatility is 2.0%, the transition probability from low to low volatility is 95% and from high to high volatility is 50%. Just like for the other parameters, these priors come from estimating their value on the data from 1947-68. After updating each quarter, the end-of-sample low volatility estimate is 0.30% and the high volatility estimate is 2.42%. The low-to-low transition probability rises slightly to 98.5% and the high-to-high transition probability falls to 5.5%. In order to make this estimation with additional parameters tractable, we simplify the problem slightly by fixing the constant c in (9) at 35, which is the posterior belief of c at the end of the sample in the previous model.

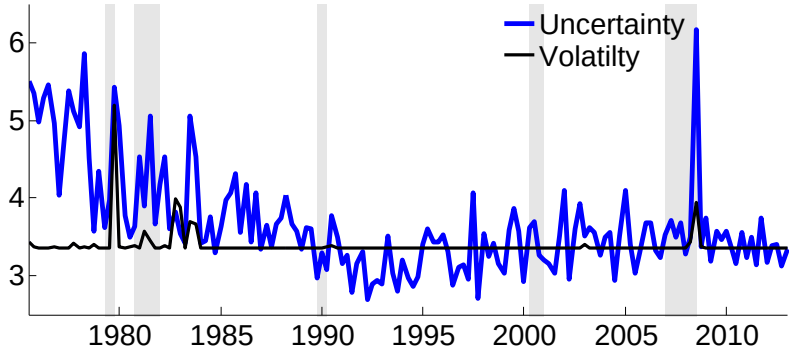


Figure 7: Uncertainty U_t and volatility V_t in the skewed model with stochastic volatility.

Figure 7 plots the uncertainty that results with parameter learning and stochastic

volatility in the skewed model, from 1976-2013. We start in 1976 because this is when the uncertainty estimates stabilize. This plot is not detrended, and yet, we see no downward trend in uncertainty after 1990. The average level of uncertainty is 4.29%, which is lower primarily because the forecaster viewed the highly-volatile 1970's data as a transitory state, not a permanent feature of the data. The forecaster with the homoskedastic model needs to accumulate lots of low-volatility observations to revise down her estimate of σ_S over time. The forecaster with the stochastic volatility model realizes that the economy is no longer in the high-volatility state, increases the probability of being in the low-volatility state and in doing so, lowers her uncertainty within a few quarters. Allowing volatility to be stochastic does make uncertainty fluctuate more. The standard deviation of U_t rises from 1.5% in the homoskedastic model to 2.0% with stochastic volatility. But adding stochastic volatility has only a small effect on the correlation of uncertainty with GDP growth (-0.72).

	Uncertainty	Volatility
Mean	2.06%	3.43%
Std deviation	2.00%	0.34%
Autocorrelation	0.83	0.22
Mean forecast	2.05%	2.35%
Mean $\|F\,Err\|$	2.61%	2.42%
Std forecast	0.54%	0.09%
Std $\|F\,Err\|$	2.37%	2.30%

Table 2: **Properties of stochastic volatility model.** Forecasts are computed using equation (4). Uncertainty, denoted U_t is computed as in Definition 2. Volatilities, denoted V_t, are computed as in Definition 3 assuming that the parameters θ are known and equal to the the mean posterior beliefs at the end of the sample for the parameter learning model.

The main lessons from combining the stochastic volatility view with the parameter learning view of uncertainty shocks are that 1) Both channels contribute to our understanding of uncertainty shocks. 2) Stochastic volatility allows the model to explain high uncertainty during the financial crisis; and 3) incorporating stochastic volatility helps to avoid the downward trend in uncertainty that arises with a homoskedastic model. It prevents uncertainty from converging to a constant level. The more realistic version of this effect is that all parameters of the model can change or drift over time. Such a model would keep learning active in all model parameters and might be a better description of reality.

But we can't estimate such a rich model. The hope is this this simple first step in that direction might give us some insight about how time-varying parameters and parameter learning might interact more generally.

3 Data Used to Proxy for Uncertainty

Our model generates a measure of economic uncertainty. Next, we compare our measure to others. Commonly used measures for uncertainty include the VIX, forecast dispersion, squared forecast errors, and stochastic volatility estimates. The introduction and section 1 describe at length the difference between our uncertainty measure and stochastic volatility. To measure stochastic volatility, we use Jurado, Ludvigson, and Ng (2013), the state-of-the-art in macro volatility measurement. We refer the reader to their paper for details on how that measure is constructed. For the remaining measures, we describe each one and examine its theoretical difference with ours. Then, we compare the statical properties.

3.1 Forecast dispersion

Some authors use forecast dispersion (D_t in equation 12) as a measure of uncertainty.[7] One advantage of this measure is that it is typically regarded as "model-free." It turns out that dispersion is only equivalent to uncertainty in models with uncorrelated signal noise and no parameter uncertainty.

Any unbiased forecast can be written as the difference between the true variable being forecast and some forecast noise that is orthogonal to the forecast. In other words,

$$y_{t+1} = E[y_{t+1}|\mathcal{I}_t] + \eta_t + e_{it} \tag{11}$$

where the forecast error ($\eta_t + e_{it}$) is mean-zero and orthogonal to the forecast. We can further decompose any forecast error into a component that is common to all forecasters η_t and a component that is the idiosyncratic error e_{it} of forecaster i.

Dispersion D_t is the average squared difference of each forecast from the average forecast. We can write each forecast as $y_{t+1} - \eta_t - e_{it}$. Then, with a large number of forecasters,

[7] See e.g. Baker, Bloom, and Davis (2012) or Diether, Malloy, and Scherbina (2002), or Johnson (2004).

we can apply the law of large numbers, set the average e_{it} to 0 and write the average forecast as $\bar{E}[y_{t+1}|\mathcal{I}_t] = y_{t+1} - \eta_t$. Thus,

$$D_t \equiv \frac{1}{N}\sum_i (E[y_{t+1}|\mathcal{I}_t] - \bar{E}[y_{t+1}|\mathcal{I}_t])^2 = \frac{1}{N}\sum_i e_{it}^2 \tag{12}$$

Note that dispersion reflects only private noise e_{it}, not public noise η_t. Uncertainty is the conditional standard deviation of the forecast error, which is $\sqrt{E[(\eta_t + e_{it})^2|\mathcal{I}]}$ and depends on both sources of noise. Thus, whether dispersion accurately reflects uncertainty depends on private or public nature of information.

3.2 Mean-squared forecast errors

A measure related to forecast dispersion that captures both private and common forecast errors is the forecast mean-squared error.

We define a forecast mean-squared error (MSE_{t+1}) of a forecast of y_{t+1} made in quarter t as the square root of the average squared distance between the forecast and the realized value

$$MSE_{t+1} = \sqrt{\frac{\sum_{i\in I_t}(E[y_{t+1}|\mathcal{I}_{it}] - y_{t+1})^2}{N_t}}. \tag{13}$$

If forecast errors were completely idiosyncratic, with no common component, then dispersion in forecasts and mean-squared forecasting errors would be equal. To see this, note that $FE_{jt}^2 = (E[y_{t+1}|\mathcal{I}_{jt}] - y_{t+1})^2$. We can split up FE_{jt}^2 into the sum $((E[y_{t+1}|\mathcal{I}_{jt}] - \bar{E}_t[y_{t+1}]) + (\bar{E}[y_{t+1}] - y_{t+1}))^2$, where $\bar{E}_t[y_{t+1}] = \int_j E[y_{t+1}|\mathcal{I}_{jt}]$ is the average forecast. If the first term in parentheses is orthogonal to the second, $1/N\sum_j FE_{jt}^2 = MSE_t^2$ is simply the sum of forecast dispersion and the squared error in the average forecast: $E[y_{t+1}|\mathcal{I}_{jt}] - \bar{E}_t[y_{t+1}])^2 + (\bar{E}_t[y_{t+1}] - y_{t+1})^2$.

We can then use this insight along with our forecast data to evaluate the extent to which variation in mean-squared errors (MSE) comes from changes in the accuracy of average forecasts and how much comes from changes in dispersion. We estimate a regression of the GDP forecast mean-squared error, MSE^2, defined in (13) on $(\bar{E}_t[y_{t+1}] - y_{t+1})^2$. We find that the R^2 of this regression is 80%. The remaining variation is due to changes in forecast dispersion. Since most of the fluctuation in MSE comes from changes in average

forecast errors, it tells us that using forecast dispersion as a proxy for uncertainty will miss an important source of variation.

3.3 VIX and confidence measures

We discuss the volatility and forecast-based measures in most detail because there are measures that we can relate explicitly to our theory. Other proxy variables for uncertainty are interesting but have a less clear connection to our model. The market volatility index (VIX) is a traded blend of options that measures expected percentage changes of the S&P500 in the next 30 days. It captures expected volatility of equity prices. But it would take a rich and complicated model to link macroeconomic uncertainty to precise movements in the VIX. Nevertheless, we can compare its statistical properties to those of the uncertainty measure in our model. Figure 8 does just this.

Another commonly cited measure of uncertainty is business or consumer confidence. The consumer confidence survey asks respondents whether their outlook on future business or employment conditions is "positive, negative or neutral." Likewise, the index of consumer sentiment asks respondents whether future business conditions and personal finances will be "better, worse or about the same." While these indices are indeed negatively correlated with the GARCH-implied volatility of GDP, they are not explicitly questions about uncertainty. Furthermore, we would like to use a measure that we can compare to the forecasts in our model. Since it is not clear what macro variable "business conditions" or "personal finances" corresponds to, it is not obvious what macro variable respondents are predicting.

3.4 Comparing uncertainty proxies to model-generated uncertainty

Figure 8 plots each of the uncertainty proxies. There is considerable comovement, but also substantial variation in the dynamics of each process. These are clearly not measures of the same stochastic process, each with independent observation noise. Furthermore, they have properties that are quite different from our model-implied uncertainty metric. Table 3 shows that our uncertainty metric is negatively correlated with traditional measures of volatility, but is highly correlated with the volatility index (VIX) and the Jurado, Ludvigson, and Ng (2013) stochastic volatility measure.

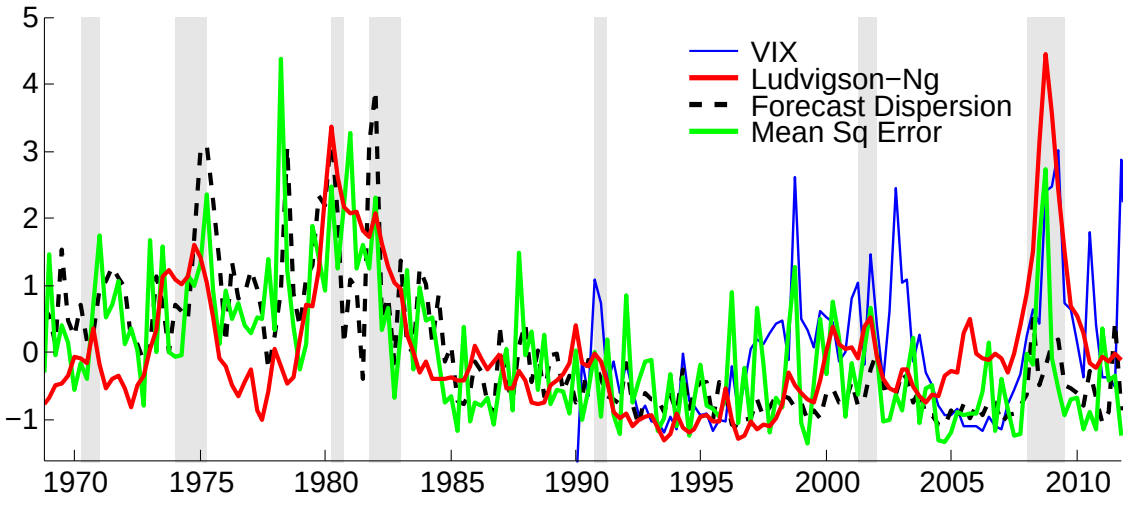


Figure 8: Comparing variables used to measure uncertainty in the literature. See Table 3 for definitions and sources.

3.5 Inferring uncertainty from probability forecasts

One way to infer the uncertainty of an economic forecaster is to ask them about the probabilities of various events. The survey of professional forecasters does just that. They ask about the probability that GDP growth exceeds 6%, is between 5-5.9%, between 4-4.9%, ..., between -1 and -2%, and below -2%. However, the survey only reports a single probability weight that is averaged across all forecasters.

Since this data does not completely describe the distribution of y_{t+1} beliefs, computing a variance requires some approximation. The most obvious approximation is to assume a discrete distribution. For example, when agents assign a probability to $1-2\%$ GDP growth, we treat this as if that is the probability placed on the outcome of 1.5% GDP growth. When the agent says that there is probability $p_{6.5}$ of growth above 6%, we treat this as probability $p_{6.5}$ placed on the outcome $y_{t+1} = 6.5\%$. And if the agent reports probability $p_{-2.5}$ of growth below -2%, we place probability of $p_{-2.5}$ on $y_{t+1} = -2.5\%$. Then the expected rate of GDP growth is $\bar{y} = \sum_{m \epsilon M} p_m m$ for $M = \{-2.5, -1.5, \ldots, 6.5\}$. Finally, the conditional variance of beliefs about GDP growth are $var[y|\mathcal{I}] = \sum_{m \epsilon M} p_m (m - \bar{y})^2$.

The resulting conditional variance series is not very informative. It hardly varies (range is $[0.0072, 0.0099]$). It does not rise substantially during the financial crisis. In fact, it

	Mean	Standard deviation	autocorr	correlation with y_{t+1}	correlation with $\tilde{U}_t$
Ludvigson-Ng	69.78%	9.54%	0.32	-0.51	30.6%
forecast MSE	2.64%	1.53	0.48	0.04	-15.4%
forecast dispersion	1.54%	0.95	0.74	-0.19	-15.2%
VIX	20.55	7.81	0.58	-0.41	40.2%

Table 3: **Properties of uncertainty measures used in the literature.** Ludvigson-Ng is the uncertainty measure from Jurado, Ludvigson, and Ng (2013). Forecast MSE and dispersion are defined in (13) and (12) and use data from 1968q4-2011q4. Growth forecast is constructed as $ln(E_t(GDP_t)) - ln(E_t(GDP_{t-1}))$. VIX_t is the Chicago Board Options Exchange Volatility Index closing price on the last day of quarter t, from 1990q1-2011q4. y_{t+1} is GDP growth, defined in Table 1. $\tilde{U}_t$ is uncertainty from our skewed model, measured as the log deviation from trend (eq. 9).

suggests that uncertainty in 2008 was roughly the same as it was in 2003. The reason this measure does not detect high uncertainty in exteme events is that the growth rates are top- and bottom-coded. All extremely bad GDP events are grouped in the bin "growth less than 2%." All the uncertainty was about how bad this recession might be. Instead, what the probability bins reveal is a high probability weight on growth below 2%. Since most of the probability is concentrated in one bin, it makes the uncertainty look low. In sum, while using surveys to ask about ex-ante probabilities of GDP events is a promising approach to measuring uncertainty, the available data does not seem useful for our purposes.

4 Conclusions

Most approaches to measuring economic uncertainty ignore parameter estimation uncertainty. Sometimes referred to as "rational expectations econometrics," the traditional approach entails estimating a model on the full sample of data and then treating the estimated parameters as truth to infer what the volatility of innovations was in each period in the past. In equating the volatility with uncertainty, the econometrician is assuming that the uncertain agent knows the true distribution of outcomes at every moment in time and is only uncertain about which outcome will be chosen from this distribution. Assuming such precise knowledge of the economic model rules out most uncertainty and ignores many sources of uncertainty shocks.

We explore the uncertainty shocks that arise when an agent is not endowed with knowledge of the true economic model and needs to estimate it, just like an econometrician. The conditional variance of this agent's forecast, his uncertainty, is much higher and varies more than volatility does. When the agent considers skewed distributions of outcomes, new data or real-time revisions to existing data can change his beliefs about the skewness of the distribution, and thus the probability of extreme events. Small changes in the estimated skewness can increase or decrease the probability of these tail events many-fold. Because tail events are so far from the mean outcome, changes in their probability have a large effect on conditional variance, which translates into large shocks to uncertainty. Thus, our message is that beliefs about black swans, extreme events that are never observed, but whose probability is inferred from a forecasting model, are responsible for much of the shocks to macroeconomic uncertainty.

This paper stripped out utilities, endowments and production technologies, in order to focus purely on the belief formation process. But Orlik, Veldkamp, and Kozeniauskas (2013) show how this mechanism can be embedded in a production economy with heterogeneous information, forecast dispersion and heterogeneous firm outputs. Our mechanism could also help explain important economic phenomena such as debt crises and equity premia. In a model where firms have debt, changes in "black swan" probabilities would become changes in default probabilities. The model would then tell us what kinds of events trigger high default premia and debt crises. Similarly, in a consumption-based asset pricing model, we know that the right stochastic process for time-varying rare event probabilities can match many features of equity returns. Our mechanism could be embedded in these models to explore what properties rare event probabilities really have and whether they might be able to resolve the equity return puzzles.

References

Albagli, E., C. Hellwig, and A. Tsyvinski (2013): "A Theory of Asset Pricing Based on Heterogeneous Information," Toulouse School of Economics working paper.

Bachmann, R., and C. Bayer (2012a): "Investment Dispersion and the Business Cycle," University of Aachen working paper.

——— (2012b): "Wait-and-See Business Cycles?," University of Aachen working paper.

Bachmann, R., S. Elstner, and E. Sims (2012): "Uncertainty and Economic Activity: Evidence from Business Survey Data," University of Aachen working paper.

Bachmann, R., and G. Moscarini (2012): "Business Cycles and Endogenous Uncertainty," Yale University working paper.

Baker, S., N. Bloom, and S. Davis (2012): "Measuring Economic Policy Uncertainty," Stanford University working paper.

Bansal, R., and I. Shaliastovich (2010): "Confidence Risk and Asset Prices," *American Economic Review*, 100(2), 537–41.

Barro, R. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Econometrica*, 77(3), 623–685.

Basu, S., and B. Bundick (2012): "Uncertainty Shocks in a Model of Effective Demand," Boston College working paper.

Bianchi, F., C. Ilut, and M. Schneider (2012): "Ambiguous business cycles and uncertainty shocks," Stanford University working paper.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77(3), 623–685.

Bloom, N., M. Floetotto, N. Jaimovich, I. Sapora-Eksten, and S. Terry (2012): "Really Uncertain Business Cycles," NBER working paper 13385.

Born, B., A. Peter, and J. Pfeifer (2011): "Fiscal News and Macroeconomic Volatility," Bonn econ discussion papers, University of Bonn, Germany.

Bruno, V., and H. Shin (2012): "Capital Flows, Cross-Border Banking and Global Liquidity," Princeton University working paper.

Chen, H., W. Dou, and L. Kogan (2013): "Measuring the Dark Matter in Asset Pricing Models," MIT working paper.

Christiano, L., R. Motto, and M. Rostagno (2012): "Risk Shocks," Northwestern University working paper.

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 (2013): "Parameter Learning in General Equilibrium: The Asset Pricing Implications," Columbia University working paper.

Diether, K., C. Malloy, and A. Scherbina (2002): "Differences of Opinion and the Cross-Section of Stock Returns," *Journal of Finance*, 57, 2113–2141.

Elliott, G., and A. Timmermann (2008): "Economic Forecasting," *Journal of Economic Literature*, 46(1), 3–56.

Fajgelbaum, P., E. Schaal, and M. Taschereau-Dumouchel (2013): "Uncertainty Traps," NYU working paper.

Fernandez-Villaverde, J., P. Guerron-Quintana, J. F. Rubio-Ramirez, and M. Uribe (2011): "Risk Matters: The Real Effects of Volatility Shocks," *American Economic Review*, 101(6), 2530–61.

Fernandez-Villaverde, J., K. Kuester, P. Guerron, and J. Rubio-Ramirez (2012): "Fiscal Volatility Shocks and Economic Activity," University of Pennsylvania working paper.

Gourio, F. (2012): "Disaster Risk and Business Cycles," *American Economic Review*, 102(6), 2734–66.

Hansen, L. (2007): "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97(2), 1–30.

Johannes, M., L. Lochstoer, and Y. Mou (2011): "Learning About Consumption Dynamics," Columbia University working paper.

Johnson, T. (2004): "Forecast Dispersion and the Cross Section of Expected Returns," *Journal of Finance*, 59, 1957–1978.

Jovanovic, B. (2006): "Asymmetric Cycles," *Review of Economic Studies*, 73, 145–162.

Jurado, K., S. Ludvigson, and S. Ng (2013): "Measuring Uncertainty," New York University working paper.

Justiniano, A., and G. Primiceri (2008): "Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98(3), 604–641.

Kozeniauskas, N., A. Orlik, and L. Veldkamp (2014): "Black Swans and the Many Shades of Uncertainty," NYU working paper.

Nakamura, E., D. Sergeyev, and J. Steinsson (2012): "Growth-Rate and Uncertainty Shocks in Consumption: Cross-Country Evidence," Columbia University working paper.

Nimark, K. (2012): "Man Bites Dog Business Cycles," Working paper.

Orlik, A., L. Veldkamp, and N. Kozeniauskas (2013): "Micro and Macro Uncertainty Shocks," Working paper.

Pastor, L., and P. Veronesi (2012): "Uncertainty about Government Policy and Stock Prices," *Journal of Finance*, forthcoming.

Roberts, G., A. Gelman, and W. Gilks (1997): "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms," *Annals of Applied Probability*, 7(1), 110–120.

Rossi, B., and T. Sekhposyan (2014): "Macroeconomic Uncertainty Indices," UPF working paper.

Stock, J., and M. Watson (2012): "Disentangling the Channels of the 2007-2009 Recession," NBER working paper 18094.

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.

Wachter, J. (2013): "Can Time-Varying Risk or Rare Disasters Explain Aggregate Stock Market Volatility?," *Journal of Finance*, 68(3), 987–1035.

A Proofs

Lemma 1: Skewness and the concave change of measure We can write the skewness of y (times the variance, which is always positive) as

$$E[(y - E[y])^3] = \int (y - E[y])^3 \phi(g^{-1}(y)) dy \tag{14}$$

where $\phi(g^{-1}(y))$ is the probability density of y, by assumption. Using the change of variable rule, we can replace y with $g(x)$.

$$E[(g(x) - E[g(x)])^3] = \int (g(x) - E[g(x)])^3 \frac{\partial g}{\partial x} \phi(x) dx \tag{15}$$

Note that we replaced $\phi(g^{-1}(g(x))) = \phi(x)$, meaning that x is a standard normal variable.

Because g is increasing and concave, $\partial g/\partial x$ is positive and decreasing in x.

If $\partial g/\partial x$ were a constant, then 15 would be the skewness of a normal variable, which is zero. Thus,

$$-\int_{-\infty}^{0}(g(x)-E[g(x)])^3\phi(x)dx = \int_{0}^{\infty}(g(x)-E[g(x)])^3\phi(x)dx$$

Since $\partial g/\partial x$ is positive and decreasing, it is higher for any $y<0$ than it is for any $y>0$ and since both sides of the inequality are positive

$$-\int_{-\infty}^{0}(g(x)-E[g(x)])^3\frac{\partial g}{\partial x}\phi(x)dx > \int_{0}^{\infty}(g(x)-E[g(x)])^3\frac{\partial g}{\partial x}\phi(x)dx$$

Adding the negative of the left side to both sides of the inequality reveals that

$$E[(g(x)-E[g(x)])^3] = \int(g(x)-E[g(x)])^3\frac{\partial g}{\partial x}\phi(x)dx < 0.$$

Lemma 2: Forecast bias. In the forecast $\hat{y} \equiv \int\int\int yf(y|\mu',\sigma')g(\mu')h(\sigma')dyd\mu'd\sigma'$, we can substitute $g(x)$ for y and substitute $x = g^{-1}(y)$ into $\phi((g^{-1}(y)-\mu)/\sigma) = f(y)$ to get

$$\hat{y} = \int\int\int g(x)\phi((x-\mu')/\sigma')g(\mu')h(\sigma')dg(x)d\mu'd\sigma'$$

Then, we can define $\tilde{x} = (x-\mu)/\sigma$ and substitute it in for x:

$$\hat{y} = \int\int\int g(\mu'+\sigma'\tilde{x})\phi(\tilde{x})g(\mu')h(\sigma')dg(x)d\mu'd\sigma'$$

Note that the inside integral evaluated at $\mu' = \mu$ and $\sigma' = \sigma$ is the true mean of y: $\bar{y} \equiv \int yf(y|\mu,\sigma)dy = \int g(\mu+\sigma\tilde{x})\phi(\tilde{x})dg(x)$. Let us use the notation $\tilde{y}(\mu',\sigma') = \int g(\mu'+\sigma'\tilde{x})\phi(\tilde{x})dg(x)$ to denote the mean of y, given any mean and variance parameters μ' and σ'. Notice that since g is assumed to be a concave function, $\tilde{y}$ is concave in the parameters μ' and σ'. Then, by Jensen's inequality, we know that for any concave function $\tilde{y}$, $E[\tilde{y}(\mu,\sigma)] < \tilde{y}(\mu,\sigma)$. Note by inspection that $E[\tilde{y}(\mu,\sigma)] = \hat{y}$ and $\tilde{y}(\mu,\sigma) = \bar{y}$ and the result follows.

B Estimating the model

In what follows we show how to use Metropolis-Hastings algorithm to generate samples from $p\left(\theta|y^t\right)$ for each $t = 0,1,2,..,T$. [8]

[8] We drop here the dependence on $\mathcal{M}$ hoping that no confusion arises; the algorithm is applied independently to generate the respective samples under each model.

The general idea of MCMC methods is to design a Markov chain whose stationary distribution, π (with $\pi T = \pi$ where T is a transitional kernel), is the distribution p we are seeking to characterize. In particular, the Metropolis-Hastings sampling algorithm constructs an ergodic Markov chain that satisfies a detailed balance property with respect to p and, therefore, produces the respective approximate samples. The transition kernel of that chain, T, is constructed based on sampling from a proposal conditional distribution $q\left(\theta|\theta^{(d)}\right)$ where d denotes the number of the sampling step. Specifically, given the d-step in the random walk $\theta^{(d)}$ the next-step $\theta^{(d+1)}$ is generated as follows

$$\theta^{(d+1)} = \begin{cases} \theta' & \text{with probability } \alpha\left(\theta^{(d)}, \theta'\right) = \min\left(1, \frac{p\left(\theta'|y^t\right)}{p\left(\theta^{(d)}|y^t\right)} \frac{q\left(\theta^{(d)}|\theta'\right)}{q\left(\theta'|\theta^{(d)}\right)}\right) \\ \theta^{(d)} & \text{with probability } 1 - \alpha\left(\theta^{(d)}, \theta'\right) \end{cases}$$

where $\theta' \sim q\left(\theta|\theta^{(d)}\right)$.

In our application, the simulation of the parameters is done through simple random walk proposals. In particular, for the means the proposed move is

$$\mu'_S = \mu_S + \tau_u \varepsilon_S \tag{16}$$

where $S \in \{1, ..., N\}$, $\varepsilon_S \sim \mathcal{N}(0, 1)$.

For the variance σ^2, the proposed move is a multiplicative random walk

$$\log \sigma' = \log \sigma + \tau_\sigma \xi_\sigma \tag{17}$$

where $\xi_\sigma \sim \mathcal{N}(0, 1)$.

In the case of the transition probability matrix, the move is slightly more involved due to the constraint on the sum of rows. We reparameterize each row $(q_{i1}, ..., q_{iN})$ as

$$q_{ij} = \frac{\omega_{ij}}{\sum_j \omega_{ij}}, \quad \omega_{ij} > 0, \quad j \in \{1, ..., N\}$$

so that the summation constraint does not hinder the random walk. The proposed move on ω_{ij} is then given by

$$\log \omega'_{ij} = \log \omega_{ij} + \tau_\omega \xi_\omega \tag{18}$$

where $\xi_\omega \sim \mathcal{N}(0, 1)$. Note that this reparametrization requires that we select a prior distribution on ω_{ij} rather than on q_{ij}.

The parameters τ_u, τ_σ, and τ_ω can be adjusted to optimize the performance of the sampler. Choosing a proposal with small variance would result in relatively high acceptance rates but with strongly correlated consecutive samples. See Roberts, Gelman, and Gilks (1997) for the results on optimal scaling of the random walk Metropolis algorithm.

Since the proposal is symmetric in all the cases (16), (17), and (18) and since ε_S, ξ_σ, and ξ_ω are drawn independently from one another, we have $q(\theta|\theta') = q(\theta'|\theta)$, and the

acceptance probability simplifies to $\min\left(1, \frac{p\left(\theta'|y^t\right)}{p\left(\theta^{(d)}|y^t\right)}\right)$. To compute that acceptance ratio, note that the posterior distribution $p\left(\theta|y^t\right)$ is given by

$$p\left(\theta|y^t\right) = \frac{p\left(y^t|\theta\right)p\left(\theta\right)}{p\left(y^t\right)}$$

where $p\left(y^t\right) = \int p\left(y^t|\theta\right)p\left(\theta\right)d\theta$ is the marginal likelihood (or data density).

Continuous state filtering details In turn, the predictive distribution of the data, $p\left(y_{t+1}|y^t,\theta\right)$ can be obtained as an integral against the filtering distribution

$$p\left(y_{t+1}|y^t,\theta\right) = \int\int p\left(y_{t+1}|s_{t+1},\theta\right)p\left(s_{t+1}|s_t,\theta\right)p\left(s_t|y^t,\theta\right)ds_t ds_{t+1}$$

We apply Kalman filtering techniques to obtain the distribution of the filtered values of the hidden state, $p\left(s_t|y^t,\theta\right)$, which will be characterized by the following moments $\widehat{s}_t \equiv E\left[s_t|y^{t-1},\theta\right]$ and $Var_{s,t} \equiv E\left[\left(s_t-\widehat{s}_t\right)\left(s_t-\widehat{s}_t\right)'\right]$. The predictive distribution of the data is then given by the following moments $E\left[y_t|y^{t-1},\theta\right] = \widehat{s}_t$ and $Var\left[y_t|y^{t-1},\theta\right] = Var_{s,t} + \sigma^2$.

Estimating Prior Beliefs To discipline the priors, we use historical data, i.e. the vintage of the data as of 1968q3 (1947q2-1968q2). We use uniform priors on all the parameters, and estimate respective models using Bayesian techniques described in the Appendix. The mean and standard deviations of the posterior parameter distributions as of 1968q3 become the moments of the prior distributions for respective parameters that will be used in the real-time estimation from 1968q4 onwards. The results for the respective models are reported in the tables below.

To compute volatility in these models, we fix parameters at the estimated means of these prior distributions. Figures 8 and 9 plot the priors and the evolution of parameter beliefs over the sample for the linear-normal and skewed models.

Parameter	Mean	Stdev
α	2.35	0.68
ρ	0.47	0.12
σ^2	4.89	3.45
σ_s^2	15.92	4.47

Table 4: Moments of the prior distributions in the linear-normal model

Parameter	Mean	Stdev
c	41.27	6.97
ρ	0.05	0.07
σ^2	0.02	0.01
σ_s^2	0.005	0.007

Table 5: Moments of the prior distributions in the model with skewness.

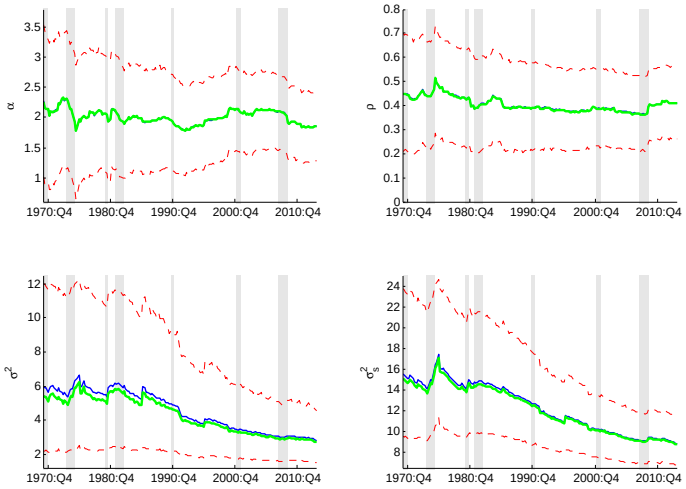


Figure 9: Linear-Normal Model: Posterior Parameter Means, Medians, and 95%- Credible Sets.

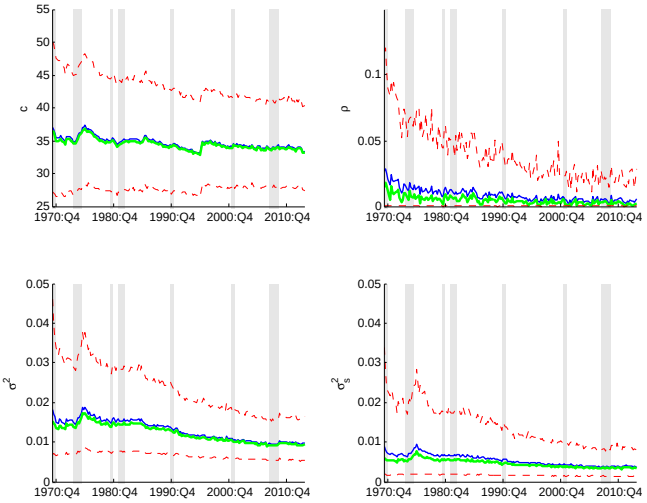


Figure 10: Skewed Model: Posterior Parameter Means, Medians, and 95%- Credible Sets.