Understanding Uncertainty Shocks and the Role of Black Swans

Anna Orlik and Laura Veldkamp [1]

October 22, 2015

[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, SED meetings, NBER Universities Research Conference on The Macroeconomic Consequences of Risk and Uncertainty, Barcelona GSE Summer Forum, CEF meetings, UCL Uncertainty and Economic Forecasting Workshop, Federal Reserve Bank of Boston, the NYU macro lunch, and also to Nick Bloom, Michele Lenza, Ignacio Presno, Thomas Sargent, Matthew Smith, Tom Stark, our EF&G discussant Jennifer La'O, our NBER Universities Research Conference discussant, Rudi Bachmann, and our AEA discussant, Lars Hansen. Many thanks to Edward Atkinson, Isaac Baley, David Johnson, Callum Jones, Nic Kozeniauskas and Pau Roldan 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

Economic uncertainty is a powerful force in the modern economy. Recent work shows that surges in uncertainty can trigger business cycles, bank runs and asset price fluctuations. But where do sudden surges in uncertainty come from? This paper provides a data-disciplined theory of belief formation that explains large fluctuations in uncertainty. It argues that people do not know the true distribution of macroeconomic outcomes. Like Bayesian econometricians, they estimate a distribution. Our main contribution is to explain why real-time estimation of distributions with non-normal tails are prone to large uncertainty fluctuations. We use theory and data to show how small changes in estimated skewness whip around probabilities of unobserved tail events (black swans). Our estimates, based on real-time GDP data, reveal that revisions in the estimates of black swan risk explain most of the fluctuations in uncertainty.

Economic uncertainty is a powerful force in the modern economy. Recent work shows that surges in uncertainty can trigger business cycles, bank runs and asset price fluctuations.[1] But the way uncertainty shocks are typically modelled is that one day, every agent suddenly knows that future outcomes will be less predictable than in the past. This mysterious belief shock is isomorphic to a preference shock. Such belief shocks are not disciplined by data, making the theories hard to test. Furthermore, if certainty is the precision of beliefs that arises from accumulating a body of information, a sudden rise in uncertainty seems to imply a sudden loss of information. Just like the loss of productivity associated with real business cycle recessions is puzzling, so is the loss of information associated with uncertainty-driven downturns.

This paper provides a data-disciplined theory of belief formation that explains large fluctuations in uncertainty. It starts from the premise that real people do not know what the true distribution of economic outcomes is, when it changes, or by how much. They observe economic information and, conditional on that information, estimate the probabilities of alternative outcomes. Much of their uncertainty comes from not knowing if their estimates are correct. Because everyday occurrences are observed frequently, their probabilities are easy to learn. After a short period, new data does not significantly alter those estimates. In contrast, the tails of a distribution are rarely observed; so their size and shape is difficult to assess. When people use observed data to infer the probabilities of unobserved tail events, new data can "wag the tail" of the distribution: It causes large revisions in tail probabilities. Since variance is expected squared distance from the mean, changes in the probabilities of events far from the mean have outsized effects on conditional variance and thus uncertainty. Thus, everyday fluctuations in a data series can produce large fluctuations in conditional variance for an agent who is constantly re-estimating the tails of the distribution.

We use real-time data to measure the uncertainty (conditional standard deviation) that arises from not knowing the true model. Then, we use a combination of data and probability theory results to explain why uncertainty varies so much. These results reveal that it is the combination of parameter uncertainty and tail risk that makes uncertainty more variable and more counter-cyclical than stochastic volatility alone. We learn why the greatest contribution to uncertainty fluctuations comes not from changes in the variance

[1]See e.g., Bloom, Floetotto, Jaimovich, Sapora-Eksten, and Terry (2012), Fajgelbaum, Schaal, and Taschereau-Dumouchel (2014), or Bacchetta, Tille, and van Wincoop (2012).

of the data, but rather from the time-varying risk of the unobserved tail events – the black swans.

To explore uncertainty, we use a forecasting model with two key features: First, outcomes are not conditionally normally distributed, and second, agents use real-time data to re-estimate parameters that govern the distribution's higher moments, such as skewness. For each quarter, we use the vintage of U.S. (real) GDP growth data that was available at that date to estimate the forecasting 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. In order to study how uncertainty changes and why, we feed GDP data into our forecasting model and compute this standard deviation.

This conceptually simple measurement 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 that when agents do not know the distribution of shocks, re-estimating beliefs can amplify changes. 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 are consistent with the observed forecast data, in particular with the puzzling forecast 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.

While using data to infer beliefs is a key strength of our approach, this is not about measuring uncertainty in the most sophisticated possible way. Rather, we use a simple framework to describe a theoretical mechanism, supported by data, to explain *why* uncertainty, beliefs and tail risk vary. The key assumption of the mechanism is that agents use everyday events to revise their beliefs about probabilities over the entire state space. This is what allows small changes in data to trigger large changes in black swan probabilities and sizeable fluctuations in uncertainty. The idea that data in normal times would change how we assess tail risk might strike one as implausible. But there is an abundance of evidence that perceptions of tail risks vary on a daily basis.[2] If we think that tail risks fluctuate in times when no extreme events occur, then either beliefs are random and irrational, or there is some information in the everyday data that agents use to update their beliefs.[3]

In section 2, we build our forecasting model. Using a change-of measure technique, we amend a standard class of models where GDP growth is assumed to be conditionally normally distributed (whether with homoscedastic or heteroskedastic innovations) by adding an exponential twist, with parameters that regulate the conditional skewness of outcomes. Each period t, our forecaster uses the complete history of GDP data as seen at time t and Bayes law to estimate her model and forecast GDP growth in $t+1$. Initially, we hold the volatility of the innovations fixed so that we can isolate the changes in uncertainty that come from parameter learning.

Even when the forecaster is certain that the variance of innovations is constant, we find large changes in conditional variance of forecasts – big uncertainty shocks. Then we ask how much of these fluctuations comes from skewness, how much comes from parameter updating and how much from their interaction. To tease this out, we turn off parameter

[2]See data based on firm-level asset prices Kelly and Jiang (2014) or on index options Gao and Song (2015).

[3]Of course, it is possible that the everyday data that is informative about tail outcomes is not GDP data. But the same principles apply to other series. One could apply the same framework and estimate tail risk from some other series to amplify its effect on uncertainty.

learning and skewness, one-by-one. We find that skewness alone generates a tiny fraction of changes in conditional variance. Parameter learning alone accounts for about one-third of our result. Most of the changes in conditional variance come from the interaction of skewness and parameter updating.

Our results reveal that the main source of uncertainty fluctuations is something we call "black swan risk," which is the conditional probability of a rare event, in this case an extremely low growth realization. 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 percentage point 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 offer a new explanation for this forecasting puzzle: Forecast bias arises from rational Bayesian belief 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 4 investigates how volatility changes and parameter learning interact. To do that, we estimate the full model, with two hidden Markov volatility states. Adding stochastic volatility makes uncertainty shocks one third larger on average. It also helps the model's performance in two key respects. First, it prevents a downward trend in uncertainty. When all parameters are believed to be constant, uncertainty trends down partly because parameters are being more precisely estimated over time, but mostly because the 70s and early

80s were much more volatile times for real GDP than the 90s and 2000s. So, the forecaster revises down the variance parameters over time and uncertainty trends down. When there are two (unobserved) volatility regimes, the forecaster infers that the 70s and early 80s were likely a high volatility regime and that the regime switches in the mid 80s. 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 5 compares our model-based uncertainty series to commonly-used uncertainty proxies and finds that it is less variable, but more persistent than the proxy variables. The most highly correlated proxies are Baker, Bloom, and Davis (2015) policy uncertainty index, the price of a volatility option (VIX), and Jurado, Ludvigson, and Ng (2015) macro uncertainty index.

Our message is that understanding the sources of economic uncertainty requires relaxing the full-information assumptions of rational expectations hypothesis. In such a full-information world, agents are assumed to know what the true 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 (Born and Pfeifer, 2012). But in reality, the macroeconomy is not governed by a simple, known model and we surely do not 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 learning about the distribution of economic outcomes may itself generate fluctuations.

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 (2014), Ilut and

Schneider (2014), Bidder and Smith (2012)), investment dynamics (Bachmann and Bayer, 2014), price-setting (Baley and Blanco, 2015), asset prices (e.g., Bansal and Shaliastovich (2010), Pastor and Veronesi (2012)), or to explain banking panics (Bruno and Shin, 2015). A related literature uses tail risk to explain asset pricing puzzles (e.g., Rietz (1988), Barro (2006), and Wachter (2013)) and business cycle fluctuations Gourio (2012). These theories are complementary to ours. We explain where uncertainty shocks come from, while these papers trace out the many economic and financial consequences of these shocks.

A growing literature in macroeconomics and finance explores how agents use information to form beliefs, with tools such as rational inattention (e.g., Maćkowiak and Wiederholt (2009), Matejka and McKay (2015), Kacperczyk, Nosal, and Stevens (2015)), inattentiveness (Reis, 2006), sentiments (Angeletos and La'O, 2013) or information diffusion (Amador and Weill, 2010). Our mechanism is not inconsistent with any of these frictions, all of which have Bayesian updating as a foundation. Instead, our paper shows how enriching the set of variables updated, to include parameters that govern tail risk, can link these dynamics to fluctuations in uncertainty as well. An advantage of our approach is that the belief formation process that we postulate is strictly disciplined by and consistent with the data.

A small subset of these theories explains why uncertainty fluctuates using nonlinearities in a production economy (Van Nieuwerburgh and Veldkamp (2006), Fajgelbaum, Schaal, and Taschereau-Dumouchel (2014), Jovanovic (2006)), active experimentation (Bachmann and Moscarini (2012)) or multiple equilibria (Bacchetta, Tille, and van Wincoop (2012)). Bachmann and Bayer (2013) support this endogenous uncertainty approach by arguing that uncertainty Granger-causes recessions, but not the other way around. In Nimark (2014), 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. In addition, our paper 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 (2015), Giglio, Kelly, and Pruitt (2015), Stock and Watson (2012), Jurado, Ludvigson, and Ng (2015), Justiniano and Primiceri (2008), Born and Pfeifer (2014) document the properties of uncertainty shocks in the

U.S. and in emerging economies, while Bachmann, Elstner, and Sims (2013) 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.

Our methodological approach is motivated by Hansen (2007) and Chen, Dou, and Kogan (2013), which critique models that give agents knowledge of parameters that econometricians cannot identify. We were also inspired by two preceding papers that 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 (forthcoming) estimate a model of consumption growth to capture properties of asset prices. While the concept is similar, our use of a model with skewness is what allows non-extreme data to whip tail risk estimates around. When the model is normal or discrete-state (as in Collin-Dufresne, Johannes, and Lochstoer (2013)), only potential disasters affect beliefs about tail probabilities. Furthermore, disaster states cannot be too extreme. Otherwise, agents will never believe they might be in the disaster. This severely limits the size of uncertainty fluctuations that result. In our model, the probability of every tail event, no matter how extreme, fluctuates when new data is observed.

Our work further draws on tools and ideas in finance models with learning and non-normal distributions, such as Breon-Drish (2015), Straub and Ulbricht (2013) and Chabakauri, Zachariadis, and Yuan (2015). In our model, agents learn about parameters instead of states. We also draw on ideas in the economic forecasting literature about model comparisons, e.g., Giacomini and Rossi (2013) and in the Bayesian estimation literature in macroeconomics (e.g., Del Negro and Schorfheide (2011)). Finally, the black swan metaphor and its relation to tail risk is of course borrowed from Taleb (2010).

1 Definitions and Data Description

A model, denoted $\mathcal{M}$, has a vector of parameters θ. Together, $\mathcal{M}$ and θ determine a probability distribution over a sequence of outcomes y_t. Let $y^t \equiv \{y_\tau\}_{\tau=1}^t$ denote a series of data (in our exercises, the GDP growth rates) 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 observations up to and including time t. The state S_t, innovations, and the parameters θ

are never observed.

The agent, whom we call a forecaster and index by i, is not faced with any economic choices. He simply uses Bayes' law to forecast future outcomes. Specifically, at each date t, the agent conditions on his information set $\mathcal{I}_{it}$ and forms beliefs about the distribution of 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

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

We define forecast errors and uncertainty over one-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. 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 (time-t measurable). However, forecast error is an ex-post measure because it is not measurable at the time when the forecast is made. Combining definition 1 and definition 2 reveals that $U_{it} = \sqrt{E[FE_{i,t+1}^2|\mathcal{I}_{it}]}$. So, uncertainty squared is the same as the expected squared forecast error.[4]

There are two pieces of data that we use to estimate and to evaluate our forecasting models. 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, which are not known until much later, is not consistent with the goal.[5] 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 (SPF), released by the Philadelphia Federal Reserve. The data are a set 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

[4] 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: $\sqrt{1/N_t \sum_i FE_{i,t+1}^2}$.

[5] Naturally, forecasters may use other information in conjunction with past GDP growth realizations to compute their forecasts. We explore a model with additional signals in Kozeniauskas, Orlik, and Veldkamp (2014). Another approach would be to take many series, extract a principle component or predictive quantile factor as in Jurado, Ludvigson, and Ng (2015) or Giglio, Kelly, and Pruitt (2015) and apply this Bayesian methodology to that factor. While this would likely produce a higher-precision forecast, the complexity would obscure the main message, which is about how the uncertainty shocks arise.

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 $t+1$.

2 A Skewed Forecasting Model with Parameter Uncertainty

The purpose of the paper is to explain why relaxing rational expectations and assuming that agents do not know the true distribution of outcomes opens up an additional source of uncertainty shocks. The key ingredients for our mechanism to operate are parameter uncertainty and skewness in the distribution. To isolate this new mechanism, we consider as simple a model as possible that has these two ingredients. We set the model up with stochastic volatility so that we can eventually explore the interaction and relative magnitudes of volatility and uncertainty fluctuations. But our results begin by shutting down the stochastic volatility so that we can see what comes from parameter updating alone. Later, we turn stochastic volatility back on to get a more complete picture of the sources of 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 contemplates a simple hidden state model as a true data generating process for GDP growth, but does not know the parameters of this model.[6] Each period, he starts with prior beliefs about these parameters and the current state, observes the new GDP data and the new revisions of past GDP data, and updates his beliefs using Bayes' law.

A key question is which forecasting model the agent should use. Once we move away from a linear-normal model, there is an infinite set of possibilities. We narrow this set by focusing on a simple distribution with skewness. In the real GDP (1968:Q4-2013:Q4) data we use for our forecasting model estimation, the skewness of GDP growth is strong: -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

[6] A related question is what happens if the agent does not know the form of the model. However, families of models can be indexed by parameters. A parameter could even be an indicator for one of two non-nested models. The point is that model uncertainty can be represented as parameter uncertainty. We can always roll the question back to one deeper level and question further assumptions. This paper is a first step in that direction.

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.

Estimating the parameter uncertainty in skewed distributions typically requires particle filtering, which is possible, but typically burdensome. We make this problem tractable by using a change of measure to introduce skewness. 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, we can use f^{-1} to take skewed data and transform it into normal data, so that we can then use standard tools from Kalman filtering and Bayesian econometrics to estimate the model parameters. Concave functions of normal variables will produce negatively skewed variables and convex functions of normals will produce positively skewed variables.

Thus, we consider the following general forecasting model that is a standard linear hidden state model with a functional operator f that can be non-linear to capture skewness.

$$\begin{aligned} y_t &= c + b\, f\left(X_t\right) \qquad (1) \\ X_t &= x(S_t) + \sigma(S_t)\epsilon_t \end{aligned}$$

where $\epsilon_t \sim N(0,1)$ is an i.i.d. random variable. We explore linear and non-linear transformations f that induce either conditionally normal or skewed distributions for y_t.

Of course, allowing a forecaster to explore the whole function space of non-linear f's is not viable. Instead, we use an approximating function. We focus the problem by considering a function f whose log is a linear approximation to many functions that would fit the data. If this approximate function generates large uncertainty shocks, it tells us that the set of functions f approximates likely do as well.

Following textbook Bayesian statistics practices (e.g., Headrick (2010), Hoaglin, Mosteller, and Tukey (1985)), we use an exponential f function to approximate the class of skewed distributions.[7] Exponential models are used because they have three desirable properties:

[7] What we are doing is estimating a probability density from a set of discrete data. A typical approach is to use a Kernel density estimator. But we want to account for parameter uncertainty. Standard Kernel densities have too many parameters to feasibly estimate their joint distribution. Therefore, Bayesian statisticians use the g-and-h family to estimate distributions with skewness, using a small number of parameters.

(1) The domain is the real line (so it can take a normal variable as an argument); (2) it is monotone; and (3) it can be either globally concave or globally convex, depending on the estimated parameters. For our purposes, the simplicity allows us to better understand why the combination of skewness and parameter uncertainty generates large, countercyclical uncertainty shocks, even though the underlying process that we estimate is homoscedastic. Thus, our baseline skewed forecasting model is (1) with the following specific assumptions.

Model 1 assumptions ($\mathcal{M}_1$): *Skewed*

$$
\begin{aligned}
f(X_t) &= exp\,(-X_t) \\
\sigma(S_t) &= \bar{\sigma} \;\; \forall t \\
x(S_t) &= S_t \\
S_t &= \rho S_{t-1} + \sigma_S \varepsilon_t
\end{aligned}
\tag{2}
$$

where $\varepsilon_t \sim N(0,1)$ is an i.i.d. random variables, also independent of ϵ_t.

This is a simplified representation, a model, of how an agent forms beliefs. Specifically, note that shocks have no time-varying volatility (constant $\bar{\sigma}$). We want to understand the fluctuations in conditional variance that come from skewness and parameter estimation alone. This assumption allows us to see how uncertainty and the skewness of y_t (2) depend sensitively on the parameter values in (1).

To isolate the role of parameter uncertainty relative to skewness in Model 1, we make two comparisons. First, we compare the results from the estimation of this model with results when parameter uncertainty is ignored and the forecaster fixes model parameters. This exercise allows us to contrast uncertainty (Definition 2) with volatility (Definition 3). Next, to isolate the nonlinear transformation (skewness) effect, we compare these results to those from a model where f is linear.

Our transformation is a simple, limiting case of this g-and-h transformation where $h = 0$.

Model 2 assumptions ($\mathcal{M}_2$): *Linear-Normal*

$$\begin{aligned} f(X_t) &= X_t \\ \sigma_t &= \bar{\sigma} \;\; \forall t \\ x(S_t) &= S_t \\ S_t &= \rho S_{t-1} + \sigma_s \varepsilon_t \end{aligned} \tag{3}$$

Contrasting $\mathcal{M}_1$ and $\mathcal{M}_2$ results shows that it is the combination of parameter uncertainty and skewness that whips around tail risk (section 3.1), causes uncertainty to fluctuate countercyclically (section 3.2), and makes forecasts downward biased (section 3.3).

So far, we held all innovation variances fixed. This was useful to illustrate and isolate the effects of our mechanism. But estimating a model that has stochastic volatility, skewness and parameter uncertainty teaches us about how these ingredients interact.

Model 3 assumptions ($\mathcal{M}_3$): *Stochastic volatility*

$$\begin{aligned} f(X_t) &= exp\,(-X_t) \\ \sigma(S_t) &\in \{\sigma(H), \sigma(L)\} \\ x(S_t) &= \bar{x} \;\; \forall t \end{aligned}$$
$$P(S_t = H | S_{t-1} = H) = \pi_{HH}, \qquad P(S_t = L | S_{t-1} = L) = \pi_{LL}$$

In this model, our forecaster estimates the Markov transition probabilities π_{HH} and π_{LL} that govern changes in variance, instead of the ρ and σ_s parameters that governed the hidden AR(1) process in the previous models.

Information sets and updating in skewed model ($\mathcal{M}_1$): Each forecaster has an identical information set, $\mathcal{I}_{it} = \{y^t, \mathcal{M}_1\}$, $\forall i$. The state S_t and the parameters $\theta = [c, b, \rho, \bar{\sigma}, \sigma_S]'$ are never observed. The model structure (f, $x(s)$, $\sigma(s)$) is known.

Our forecaster needs prior distributions over all the parameters to start the updating process. We start with a flat prior, estimate each parameter[8] on GDP growth data from 1947:Q2-1968:Q3, 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:Q4 because this is the first quarter for which we have forecasts from the Survey of Professional Forecasters. Recall that we do not use SPF data in the estimation but only to evaluate our forecasting model.

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

$$E\left(y_{t+1}|y^t\right) = \int\int\int\int y_{t+1} p\left(y_{t+1}|S_{t+1}, S_t, \theta, y^t\right) p\left(S_{t+1}|S_t, \theta, y^t\right) p\left(S_t|\theta, y^t\right) p\left(\theta|y^t\right) d\theta dS_t dS_{t+1} dy_{t+1} \tag{4}$$

The first probability density function, $p\left(y_{t+1}|S_{t+1}, S_t, \theta, y^t\right)$, is the probability of $t+1$ GDP growth, given the state and the parameters. This function is a composition of f^{-1} and a standard normal density, denoted ϕ. Conditional on estimates for b and c, we can do a change of variable: Construct $f^{-1}((y_t - c)/b)$ to transform GDP growth y_t into a variable $X_t = x(S_t) + \sigma(S_t)\epsilon_t$, which we have constructed as a normally-distributed continuous variable with a persistent hidden state. 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 (Kalman) filtering problem with unknown parameters that can be estimated jointly using the standard Bayesian estimation techniques.

The second probability density function, $p\left(S_{t+1}|S_t, \theta, y^t\right)$, is the probability of a hidden state. In models 1 and 2, the hidden state has a linear law of motion and normally-distributed shocks. Thus, the Kalman filter delivers the mean and variance of the (condi-

[8] In the results we present, we introduced one modification. Notice that the b parameter governs the mean of the X_t process. To see this, note that for $b < 0$, we can rewrite $b\,exp(-X_t) = -exp(-X_t + ln(|b|))$. To streamline our code, we simply remove the time-t sample mean of the 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 more complex settings, we come up with consistently negative values and quantitatively similar estimates.

tional) normal density. In model 3, the hidden discrete Markov filter delivers a closed-form solution for the probability of each state (H or L).

Finally, the last probability density function is the probability of the parameter vector θ, conditional on the t-history of observed GDP data. To estimate the posterior parameters distribution, we employ Markov Chain Monte Carlo (MCMC) techniques.[9] At each date t, the MCMC algorithm produces a sample of 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: $\int f(\theta)p(\theta|y^t)d\theta \approx 1/D\sum_d f(\theta^d)$.

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: $U_t = \sqrt{Var\left(y_{t+1}|y^t\right)}$.

Beliefs in skewed model ($\mathcal{M}_1$), conditional on parameters. The exponential form of f in model 1 allows us to describe the conditional mean and variance jointly

$$E\left[y_{t+1}|y^t,\theta,\mathcal{M}\right] = c + b\exp\left(-E\left[S_{t+1}|y^t,\theta,\mathcal{M}\right] + \frac{1}{2}Var\left[S_{t+1}|y^t,\theta,\mathcal{M}\right] + \frac{1}{2}\sigma^2\right)$$

where the following recursion characterizes the updating of state belief $E\left[S_t|y^t,\theta,\mathcal{M}\right] = (1-K_t)\,E\left[S_t|y^{t-1},\theta,\mathcal{M}\right] + K_t\, ln((y_t - c)/b)$, and where the term $K_t = Var\left[ln((y_t-c)/b)|y^{t-1},\theta,\mathcal{M}\right]$ $\left(Var\left[ln((y_t-c)/b)|y^{t-1},\theta,\mathcal{M}\right] + \sigma_s^2\right)^{-1}$ is the Kalman gain. The conditional variance is

$$Var\left[ln((y_t-c)/b)|y^{t-1},\theta,\mathcal{M}\right] = \rho^2\left[\frac{1}{Var\left[ln((y_{t-1}-c)/b)|y^{t-2},\theta,\mathcal{M}\right]} + \frac{1}{\sigma_s^2}\right]^{-1} + \sigma^2 \quad (5)$$

and volatility is $\sqrt{Var[y_{t+1}|y^t,\theta,\mathcal{M}]}$.

3 Results: Black Swan Risk and Uncertainty Fluctuations

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

[9] More details are presented in the Appendix B.

from skewness and parameter learning. We begin by showing that neither parameter updating nor skewness alone produces the large uncertainty fluctuations. Instead, most of the effect arises from the interaction of these two forces. Then, we proceed to explain how this interaction effect works.

model:	unc/vol	normal ($\mathcal{M}_2$)	skewed ($\mathcal{M}_1$)
Mean	U_t	4.20%	4.53%
	V_t	3.45%	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
		Cyclical properties	
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%	2.73%	2.27%
Mean $\|F\,Err\|$	1.87%	2.25%	2.51%
Std forecast	2.25%	1.17%	0.64%
Std $\|F\,Err\|$	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 mean posterior beliefs at the end of the sample for the parameter learning models.

What effects can parameter estimation alone explain? Column 1 of table 1 reveals that, without skewness (or any other higher moment in play), parameter revisions generate small uncertainty shocks. This happens when, for example, the forecaster sees an outlier observation and revises up the estimated variance of one or both innovations. With known parameters these revisions do not take place and stdev(U_t) = 0. When parameters are updated every period, stdev(U_t) = 0.48.

Column 1 also exposes two aspects of linear-normal model 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 normal model does not explain the low average forecasts of GDP observed in the professional forecaster data. The true average of GDP growth over 1968:Q4-2013:Q4 is 2.68%. The average professional forecast of GDP growth is 2.24%, almost half a percentage point lower.[10] This model fails to explain that gap.

What part of the results can skewness alone explain? One reason that uncertainty varies so little with a normal forecasting model is that the normal distribution has the unusual properties that the conditional variance is the same irrespective of the conditional mean. 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. The skewed forecasting model does have a conditional variance that depends on the mean. Even when parameters of the model are known, changes in the estimated state move the conditional standard deviation of the forecast. This raises the question of whether most of our variation in uncertainty comes from skewness alone.

In table 1, column 2, the rows labelled V_t report the moments of the model without parameter uncertainty or parameter revisions. Indeed, even without the parameter revisions, uncertainty does vary. But that effect is tiny. It is less than 5% of the size of the fluctuations in the full model.

Figure 1 plots the time series of our uncertainty estimates, breaking out the fluctuations that come from parameter updating or skewness alone. Column "skewed" of Table 1 shows that updating beliefs about the skewness of the GDP growth distribution has a large 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.

[10]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.

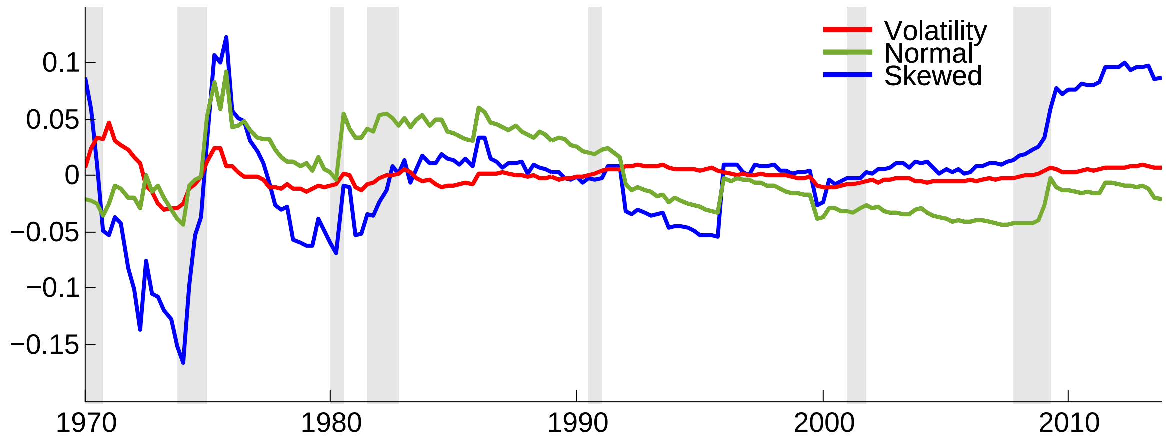

Figure 1: Uncertainty implied by normal ($\mathcal{M}_2$) and skewed ($\mathcal{M}_1$) models (detrended)

.

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:

$$\tilde{U}_t \equiv ln(U_t) - ln(U_t^{trend}) \tag{6}$$

The resulting series, plotted in figure 1, 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. In addition, uncertainty is very persistent here. That persistence also declines once we introduce stochastic volatility. These series are not yet a complete picture of macroeconomic uncertainty. Instead, they are a look at what part of uncertainty is missed when we just measure volatility.

3.1 Skewness 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 us consider the probability of a particular negative growth event. The historical 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 us call this rare event a black swan.

$$Black\,Swan\,Risk_t = Prob[y_{t+1} \leq -6.8\% | \mathcal{I}_t]. \tag{7}$$

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 3.1 plots the estimated black swan probability each period. The black swan probability varies

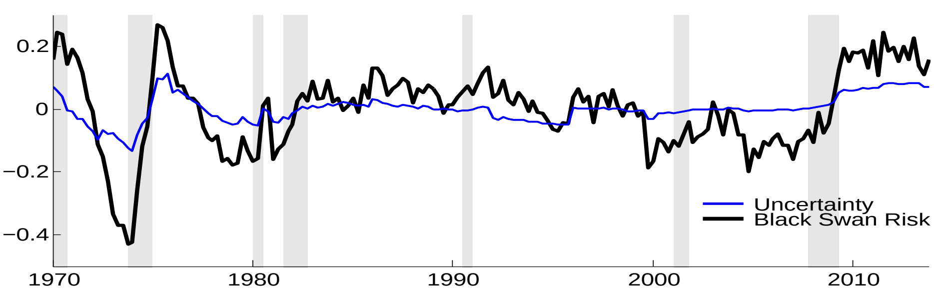

Figure 2: When the probability of a black swan event is high, uncertainty is high. *Black Swan Risk* is defined in (7). Both black swan risk and the uncertainty series are exponentially detrended.

considerably. Leading up to the 2008 financial crisis, the black swan probability rose from 3.5% in 2007:Q1 to over 4.6% in 2009:Q3.

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 belief. 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 1980s. 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 followed by a mildly negative observation. 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 1970s as can be seen in Figure 3. A sequence of positive growth realizations causes a rise and then a fall in uncertainty. But the persistence of the high estimated skewness sets the stage for the large rise in uncertainty in the second half of the 1970s. This mechanism provides one explanation for why uncertainty seems to rise particularly at the end of long spells of consistently positive growth.

3.2 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 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

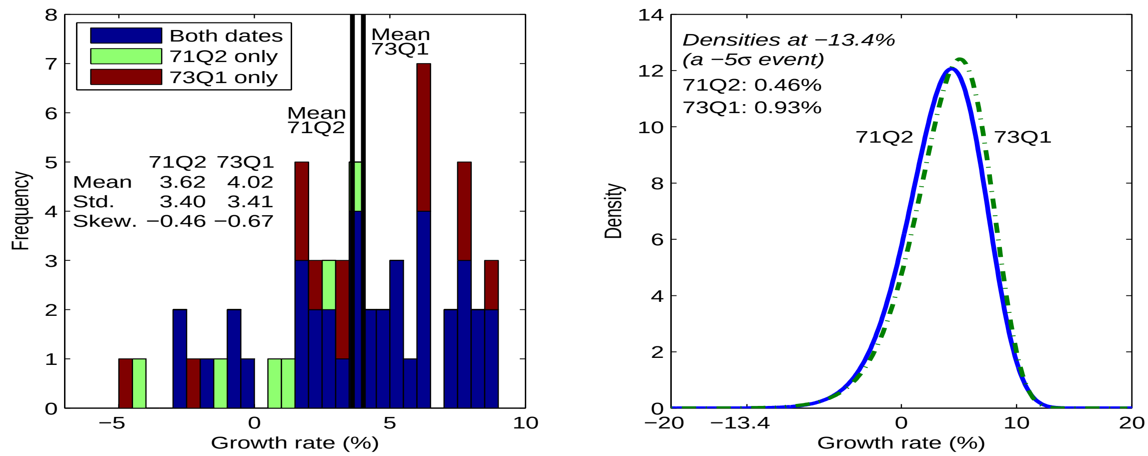

Figure 3: An example of a positive growth episode that increased the estimated mean, skewness and black swan probability.

regulates the skewness of outcomes.[11] 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 (2015) for a similar approach.) The following result shows that a concave transformation of a variable with a normal probability density results in a variable whose distribution has negative skewness. For proof see Appendix A.

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 GDP growth, we consistently find that the coefficient b is negative, meaning that the transformation is increasing and concave. A concave transformation of a normal variable

[11]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.

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.

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

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,

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

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

The conditional 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 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 4 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 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

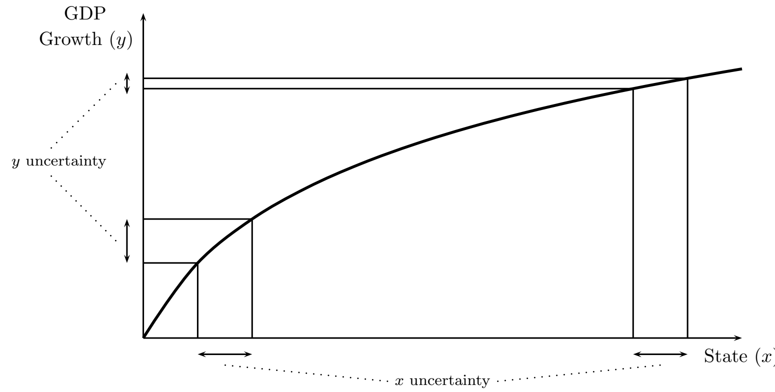

Figure 4: Nonlinear change of measure and counter-cyclical uncertainty. A given amount of uncertainty about x creates more uncertainty about y when x is low than it does when x is high.

counter-cyclical uncertainty is related to Straub and Ulbricht (2013), except that in their 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 skewed model in Table 1.

3.3 Why Skewness and Parameter 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.[12] These forecasts are puzzling because the average GDP growth rate is 2.68%. It cannot be that over 70 years of post-war history, forecasters have not figured out that the sample mean is 0.4% higher than their

[12] Various studies prior to ours document a downward (pessimistic) forecast bias. Elliott and Timmermann (2008) argue that stock analysts over-estimate earnings growth and the Federal Reserve under-estimates GDP growth. Wieland and Wolters (2013) document the bias in the Greenbook forecasts of the Federal Reserve both for the GDP growth and inflation forecasts.

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. This is an application of the Box (1971) result that Bayesian estimates of parameters in non-linear functions are typically biased.

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 $h(\mu'|\sigma')$ and $k(\sigma')$, with means μ and σ. The forecaster uses these parameter densities to construct a forecast: $\hat{y} \equiv \int\int\int y\, f(y|\mu',\sigma')\, h(\mu'|\sigma')\, k(\sigma')\, dy\, d\mu'\, d\sigma'$. 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 mean 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 5, left panel). It must

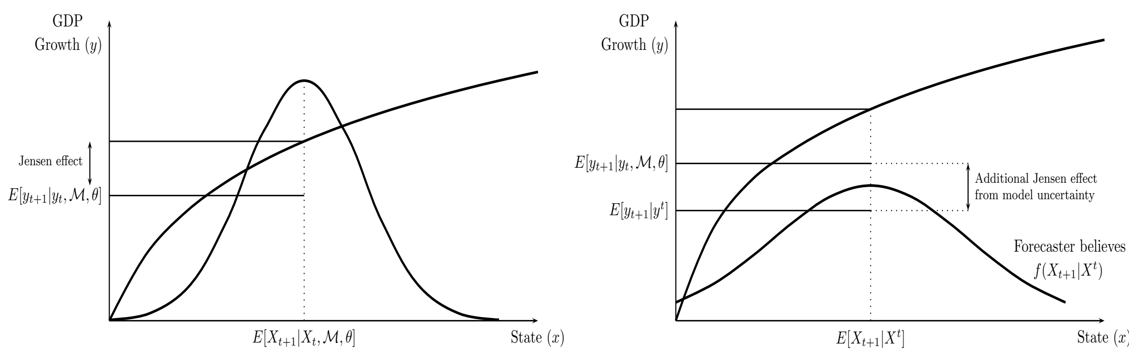

Figure 5: Explaining why average forecasts are lower than mean GDP growth. The result has two key ingredients: The forecaster faces more uncertainty than he would if he knew the true distribution of outcomes, and a Jensen inequality effect from the concave change of measure.

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 5, 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.

This downward bias in beliefs is the kind of bias that is typically only seen in models of ambiguity aversion or robust control. Those models use a particular form of risk preferences to make agents act as if they believed that systematically bad outcomes would arise. Such models with non-linear transformations of preferences are typically solved as if they had simple preferences with twisted probabilities. Our framework generates similar beliefs because the non-linear functions of normal variables that we introduce to capture skewness are similar to the non-linear functions robustness/ambiguity solution methods employ to "twist" their probabilities.

This parallel is useful because it suggests that results from ambiguity aversion theories could be reproduced in Bayesian settings with standard preferences. We could replace the min-max preferences of ambiguity with a skewed distribution of outcomes and agents who are imperfectly informed about the distribution's parameters. This could be a useful step forward for this literature simply because the data disciplines econometric estimates of probability distributions more precisely than it does preference specifications.

3.4 Introducing Additional Signals to Reduce Forecast Error

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.

Suppose that each period, each forecaster i observes an additional signal z_{it} that is the next period's GDP growth, with common signal noise and idiosyncratic signal noise:

$$z_{it} = y_{t+1} + \eta_t + \epsilon_{it} \tag{8}$$

where $\eta_t \sim N(0, \sigma_\eta^2)$ is common to all forecasters and $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ is i.i.d. across forecasters. The two signal noise variances σ_η^2 and σ_ϵ^2 can be calibrated to match the average dispersion of forecasts and the average forecast error, $1/T \sum_t FE_t$. Kozeniauskas, Orlik, and Veldkamp (2014) embeds this non-normal forecasting model with additional forecasting information (signals) in a business cycle model to show how micro uncertainty and higher-order uncertainty can all arise from the same mechanism. The results in that paper show that these additional signals remedy the forecast accuracy problem, without compromising the large, counter-cyclical uncertainty shocks.

3.5 Convergence and the Downward Trend in Uncertainty

Since the parameters in this model are constant, eventually agents will learn them if the model is correctly specified. Even in our 45-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 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 simply because GDP growth 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 prevent 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. One example of a model with time-varying parameters is a stochastic volatility model. We turn to these results next.

4 Uncertainty Shocks with 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.

Recall that variance is itself a hidden state that can take on one of two values $\sigma(S_t) \in \{\sigma(H), \sigma(L)\}$. State changes are governed by a Markov transition matrix whose entries are also estimated by our forecaster.

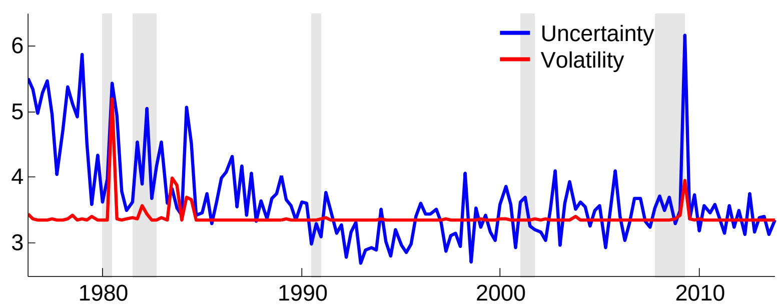

Figure 6: Uncertainty U_t and volatility V_t in the skewed model with stochastic volatility.

Figure 6 plots the uncertainty that results with parameter learning and stochastic volatility in the skewed model. 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 1970s 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 the fundamental volatility over time. The forecaster with the stochastic volatility model revises her beliefs by increasing 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	4.29%	3.43%
Std deviation	2.00%	0.34%
Autocorrelation	0.83	0.22
Mean forecast	2.05%	2.35%
Mean $\lvert F\,Err \rvert$	2.61%	2.42%
Std forecast	0.54%	0.09%
Std $\lvert F\,Err \rvert$	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 and might be a better description of reality. But such a rich model is obviously difficult to estimate. The hope is that this simple first step in that direction might give us some insight about how time-varying parameters and parameter learning might interact more generally.

5 Data Used to Proxy for Uncertainty

Our model generates a measure of economic uncertainty. In this section, we describe the commonly used proxies of uncertainty, analyze their theoretical relationship with conditional variance and then compare their statical properties to those of our measure.

Forecast Dispersion Some authors use forecast dispersion as a measure of uncertainty[13] often because it is 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:

$$y_{t+1} = E[y_{t+1}|\mathcal{I}_{it}] + \eta_t + e_{it} \tag{9}$$

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

$$D_t \equiv \frac{1}{N}\sum_i (E[y_{t+1}|\mathcal{I}_{it}] - \bar{E}[y_{t+1}])^2 = \frac{1}{N}\sum_i e_{it}^2 \tag{10}$$

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 $\sqrt{E[(\eta_t + e_{it})^2|\mathcal{I}_{it}]}$ and depends on both sources of noise. Thus, whether dispersion accurately reflects uncertainty depends on the private or public nature of information.

Mean-Squared Forecast Error A measure that captures both private and common forecast errors is the forecast mean-squared error.

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

$$MSE_{t+1} = \sqrt{\frac{\sum_{i \in I_t}(E[y_{t+1}|\mathcal{I}_{it}] - y_{t+1})^2}{N_t}}. \tag{11}$$

If forecast errors were completely idiosyncratic, with no common component, then

[13] See e.g. Diether, Malloy, and Scherbina (2002), and Johnson (2004).

dispersion in forecasts and mean-squared forecasting errors would be equal. [14] We use this insight to measure how much variation in mean-squared errors (MSE) comes from changes in the accuracy of average forecasts and how much comes from changes in dispersion. Using SPF data, we regress MSE^2 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. This teaches us that most of the fluctuations in MSE come from changes in average forecast errors. It implies that using forecast dispersion as a proxy for uncertainty will miss an important source of variation.

Volatility and Confidence Measures Jurado, Ludvigson, and Ng (2015) offer a state-of-the-art macro volatility measure. It uses a rich set of time series, computes conditional volatility of the unforecastable component of the future value of each of these series, and then aggregates these individual conditional volatilities into a macro uncertainty index. Other proxy variables for uncertainty are informative, but have a less clear connection to a conditional variance definition of uncertainty. 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. It would require a complex model to link macroeconomic uncertainty to the VIX. Nevertheless, we compare its statistical properties to those of our uncertainty measure in Figure 7.

Another commonly cited measure of uncertainty is business or consumer confidence. The 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." These questions are about the direction of future changes and not about any variance or uncertainty. They may be correlated with uncertainty because uncertainty is counter-cyclical.

Finally, Baker, Bloom, and Davis (2015) use newspaper text analysis, the number of expiring tax laws, and forecast dispersion to create a policy uncertainty index. While

[14] 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: $E[y_{t+1}|\mathcal{I}_{jt}] - \bar{E}_t[y_{t+1}])^2 + (\bar{E}_t[y_{t+1}] - y_{t+1})^2$.

the qualitative nature of the data precludes any theoretical comparison, we include it for comparison as an influential alternative.

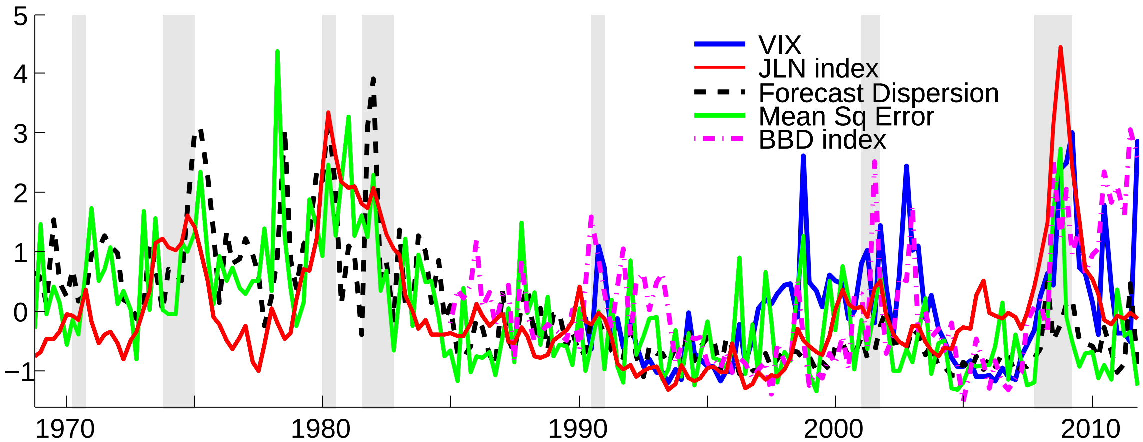

Figure 7: Comparing variables used to measure uncertainty in the literature. See Table 3 for definitions and sources.

	Mean	Standard deviation	autocorr	correlation with y_{t+1}	correlation with $\tilde{U}_t$
JLN index	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%
BBD index	105.95	31.79	0.65	-0.41	60.0%

Table 3: **Properties of uncertainty measures used in the literature.** JLN index is the uncertainty measure from Jurado, Ludvigson, and Ng (2015). Forecast MSE and dispersion are defined in (11) and (10) and use data from 1968:Q4-2011:Q4. 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 1990:Q1-2011:Q4. BBD index is the uncertainty measure from Baker, Bloom, and Davis (2015). $\tilde{U}_t$ is uncertainty from our skewed model, measured as the log deviation from trend (eq. 6).

Comparing Uncertainty Proxies to Model-Generated Uncertainty Figure 7 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 Baker, Bloom, and Davis (2015) policy uncertainty index, the volatility index (VIX) and the Jurado, Ludvigson, and Ng (2015) stochastic volatility measure.

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 SPF asks about the probability that GDP growth exceeds 6%, is between 5-5.9%, between 4-4.9%, . . . , and below -2%. The survey averages across all forecasters and reports a single average probability for each bin. Since this data does not completely describe a conditional distribution, computing the conditional variance requires approximation. The most obvious approximation is to assume that these are probabilities of ten discrete growth rates, each corresponding to the mid-point of a bin.[15]

The resulting conditional variance series is not very informative. It hardly varies (range is $[0.0072, 0.0099]$). It does not spike in the financial crisis. In fact, the SPF-implied variance suggests that uncertainty in 2008 was roughly the same as it was in 2003. The problem is that the growth rates are top- and bottom-coded. All extremely bad GDP events are grouped in the bin "growth less than 2%." If there is a very high probability of growth below 2%, then since most of the probability is concentrated in one bin, variance and, thus, uncertainty is low.

The main point of our paper is that most uncertainty shocks come from changes in the probabilities of extreme events. This survey truncates extremes and, therefore, fails to capture most changes in uncertainty.

[15]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 $var[y|\mathcal{I}] = \sum_{m\epsilon M} p_m (m - \bar{y})^2$.

6 Conclusions

Theories based on news shocks, uncertainty shocks, higher-order uncertainty shocks, tail risk shocks, and belief shocks generally have been influential in macroeconomics. But they leave unanswered the question: Why do beliefs fluctuate in this way? Just like output arises from feeding inputs into a technology, beliefs arise from feeding information sets into a belief-formation procedure. Just like a complete theory explains why the inputs and output change, it should also tell us why beliefs change.

In this paper, we consider a Bayesian belief-formation mechanism that allows for estimation of tail risk. We feed in an information set that is simply the real-time available GDP history and a reference forecasting model that the forecaster estimates in the real time just like an econometrician. We find that these simple ingredients produce large, countercyclical fluctuations in tail risk and uncertainty. Furthermore, without any preference assumptions, they produce a downward bias in mean beliefs that resembles ambiguity or robustness.

This theory of the origins of belief shocks suggests a change in our approach to measurement. Most economic uncertainty measures 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 volatility with uncertainty, the econometrician assumes 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 has focused on the belief formation process. In our approach disciplined by the data we uncovered the mechanisms that make uncertainty fluctuate over time. As such, this paper is a foundation on which other theories can build. Kozeniauskas, Orlik, and Veldkamp (2014) show how a similar mechanism can be embedded in a production economy with heterogeneous information, forecast dispersion and heterogeneous firm outputs. Our mechanism could also be used to model default risk. Since "black swan" probabilities could be interpreted as default probabilities, the model would then tell us what kinds of data realizations trigger high default premia and debt crises. In another project, our mechanism could be embedded in a consumption-based asset pricing model. We know that a well-engineered stochastic process for time-varying rare event probabilities can match many features of equity returns. Our tools could be used to estimate these rare event probabilities and assess whether the estimates explain asset return puzzles.

References

Albagli, E., C. Hellwig, and A. Tsyvinski (2015): "A Theory of Asset Pricing Based on Heterogeneous Information," Yale University working paper.

Amador, M., and P.-O. Weill (2010): "Learning from prices: Public communication and welfare," *Journal of Political Economy*, 118, 866–907.

Angeletos, G., and J. La'O (2013): "Sentiments," *Econometrica*, 81(2), 739–780.

Bacchetta, P., C. Tille, and E. van Wincoop (2012): "Self-Fulfilling Risk Panics," *American Economic Review*, 102(7), 3674–3700.

Bachmann, R., and C. Bayer (2013): "Wait-and-See Business Cycles?," *Journal of Monetary Economics*, 60(6), 704–719.

——— (2014): "Investment Dispersion and the Business Cycle," *American Economic Review*, 104(4), 1392–1416.

Bachmann, R., S. Elstner, and E. Sims (2013): "Uncertainty and Economic Activity: Evidence from Business Survey Data," *American Economic Journal: Macroeconomics*, 5(2), 217–249.

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

Baker, S., N. Bloom, and S. Davis (2015): "Measuring Economic Policy Uncertainty," Stanford University working paper.

Baley, I., and A. Blanco (2015): "Menu Costs, Uncertainty Cycles and the Propagation of Nominal Shocks," New York 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," *The Quarterly Journal of Economics*, 121(3), 823–866.

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

Bidder, R., and M. Smith (2012): "Robust Animal Spirits," *Journal of Monetary Economics*, 59(8), 738–750.

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., and J. Pfeifer (2012): "A practical guide to volatility forecasting through calm and storm," *The Journal of Risk*, 3-22, 14.

——— (2014): "Policy risk and the business cycle," *Journal of Monetary Economics*, 68, 68–85.

Box, M. (1971): "Bias in Nonlinear Estimation," *The Journal of the Royal Statistical Society*, 171-201, 33(2).

Breon-Drish, B. (2015): "On Existence and Uniqueness of Equilibrium in a Class of Noisy Rational Expectations Models," *Review of Economic Studies*, forthcoming.

Bruno, V., and H. Shin (2015): "Capital Flows, Cross-Border Banking and Global Liquidity," *Journal of Monetary Economics*, 71, 119–132.

Chabakauri, G., K. Zachariadis, and K. Yuan (2015): "Multi-Asset Noisy Rational Expectations Equilibrium with Contingent Claims," LSE 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 (2014): "Risk Shocks," *American Economic Review*, 104(1), 27–65.

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.

Del Negro, M., and F. Schorfheide (2011): "Bayesian Macroeconometrics," in *The Oxford Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. van Dijk. Oxford University Press.

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 (2014): "Uncertainty Traps," NBER Working Paper No. 19973.

Gao, G., and Z. Song (2015): "Rare Disaster Concerns Everywhere," Cornell working paper.

Giacomini, R., and B. Rossi (2013): "Forecasting in Macroeconomics,," in *Handbook of Research Methods and Applications on Empirical Macroeconomics*, ed. by N. Hashimzade, and M. Thornton. Edward Elgar.

Giglio, S., B. Kelly, and S. Pruitt (2015): "Systemic Risk and the Macroeconomy: An Empirical Evaluation," *Journal of Financial Economics*, forthcoming.

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.

Headrick, T. (2010): *Statistical Simulation: Power Method Polynomials and Other Transformations*. CRC Press: Carbondale, IL, first edn.

Hoaglin, D., F. Mosteller, and J. Tukey (1985): *Exploring Data Tables, Trends and Shapes*. John Wiley & Sons: New York, NY, first edn.

Ilut, C., and M. Schneider (2014): "Ambiguous Business Cycles," *American Economic Review*, 104(8), 2368—2399.

Johannes, M., L. Lochstoer, and Y. Mou (forthcoming): "Learning About Consumption Dynamics," *Journal of Finance*.

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 (2015): "Measuring Uncertainty," *American Economic Review*, 105(3), 1177–1216.

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

Kacperczyk, M., J. Nosal, and L. Stevens (2015): "Investor Sophistication and Capital Income Inequality," working paper.

Kelly, B., and H. Jiang (2014): "Tail Risk and Asset Prices," *Review of Financial Studies*, forthcoming.

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

Maćkowiak, B., and M. Wiederholt (2009): "Optimal sticky prices under rational inattention," *American Economic Review*, 99 (3), 769–803.

Matejka, F., and A. McKay (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105(1), 272–98.

Nimark, K. (2014): "Man-Bites-Dog Business Cycles," *American Economic Review*, 104(8), 2320–2367.

Pastor, L., and P. Veronesi (2012): "Uncertainty about Government Policy and Stock Prices," *Journal of Finance*, 67(4), 1219–1264.

Reis, R. (2006): "Inattentive producers," *Review of Economic Studies*, 73(3), 793–821.

Rietz, T. (1988): "The Equity Risk Premium: A Solution," *Journal of Monetary Economics*, 22(1), 117–131.

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.

Stock, J., and M. Watson (2012): "Disentangling the Channels of the 2007-2009 Recession," *Brookings Papers on Economic Activity*, pp. 81–135.

Straub, L., and R. Ulbricht (2013): "Credit Crunches, Information Failures, and the Persistence of Pessimism," Toulouse School of Economics working paper.

Taleb, N. N. (2010): *The Black Swan: The Impact of the Highly Improbable*. Random House.

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.

Wieland, V., and M. H. Wolters (2013): "Forecasting and policy making," in *Handbook of economic forecasting*, ed. by G. Elliott, and A. Timmermann. North-Holland Amsterdam.

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

$$E[(y - E[y])^3] = \int (y - E[y])^3 \phi(g^{-1}(y))dy \tag{12}$$

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

$$E[(g(x) - E[g(x)])^3] = \int (g(x) - E[g(x)])^3 \frac{\partial g}{\partial x} \phi(x)dx \tag{13}$$

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 13 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 $\hat{y} \equiv \int \int \int y f(y|\mu', \sigma') g(\mu'|\sigma') h(\sigma') dy d\mu' d\sigma'$, we can substitute $g(x)$ for y and substitute $x = g^{-1}(y)$ into $\phi((g^{-1}(y) - \mu)/\sigma) = f(y)$ to get

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

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

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

Note that the inside integral evaluated at $\mu' = \mu$ and $\sigma' = \sigma$ is the true mean of y: $\bar{y} \equiv \int y f(y|\mu, \sigma) dy = \int g(\mu + \sigma\tilde{x}) \phi(\tilde{x}) dg(x)$. 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 $t = 1, 2, .., T$. [16]

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 or multiplicative random walk proposals in case of variance parameters[17].

The standard deviations of the shocks in the random walk proposals 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 proposals are independent of each other and symmetric in all the cases, we have $q\left(\theta|\theta'\right) = q\left(\theta'|\theta\right)$, 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

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

[16] We drop here the dependence on $\mathcal{M}$ hoping that no confusion arises.

[17] In the case of the transition probability matrix for the hidden state in the skewed stochastic volatility model, 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

$$\log \omega'_{ij} = \log \omega_{ij} + \tau_\omega \xi_\omega \tag{14}$$

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}.

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).

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 obtained through the Kalman filter.

Estimating Prior Beliefs To discipline the priors, we use historical data, i.e. the vintage of the data as of 1968:Q3 (1947:Q2-1968:Q2). We use uniform priors on all the parameters, and estimate respective models using Bayesian techniques described above. The mean and standard deviations of the posterior parameter distributions as of 1968:Q3 become the moments of the prior distributions for respective parameters that will be used in the real-time estimation from 1968:Q4 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. Figure 8 plots the priors and the evolution of parameter beliefs over the sample.

	Normal		Skewed	
Parameter	Mean	Stdev	Mean	Stdev
c	2.35	0.68	41.27	6.97
ρ	0.47	0.12	0.05	0.07
$\bar{\sigma}^2$	4.89	3.45	0.02	0.01
σ_s^2	15.92	4.47	0.005	0.007

Table 4: Moments of the prior distributions in the linear-normal and skewed models.

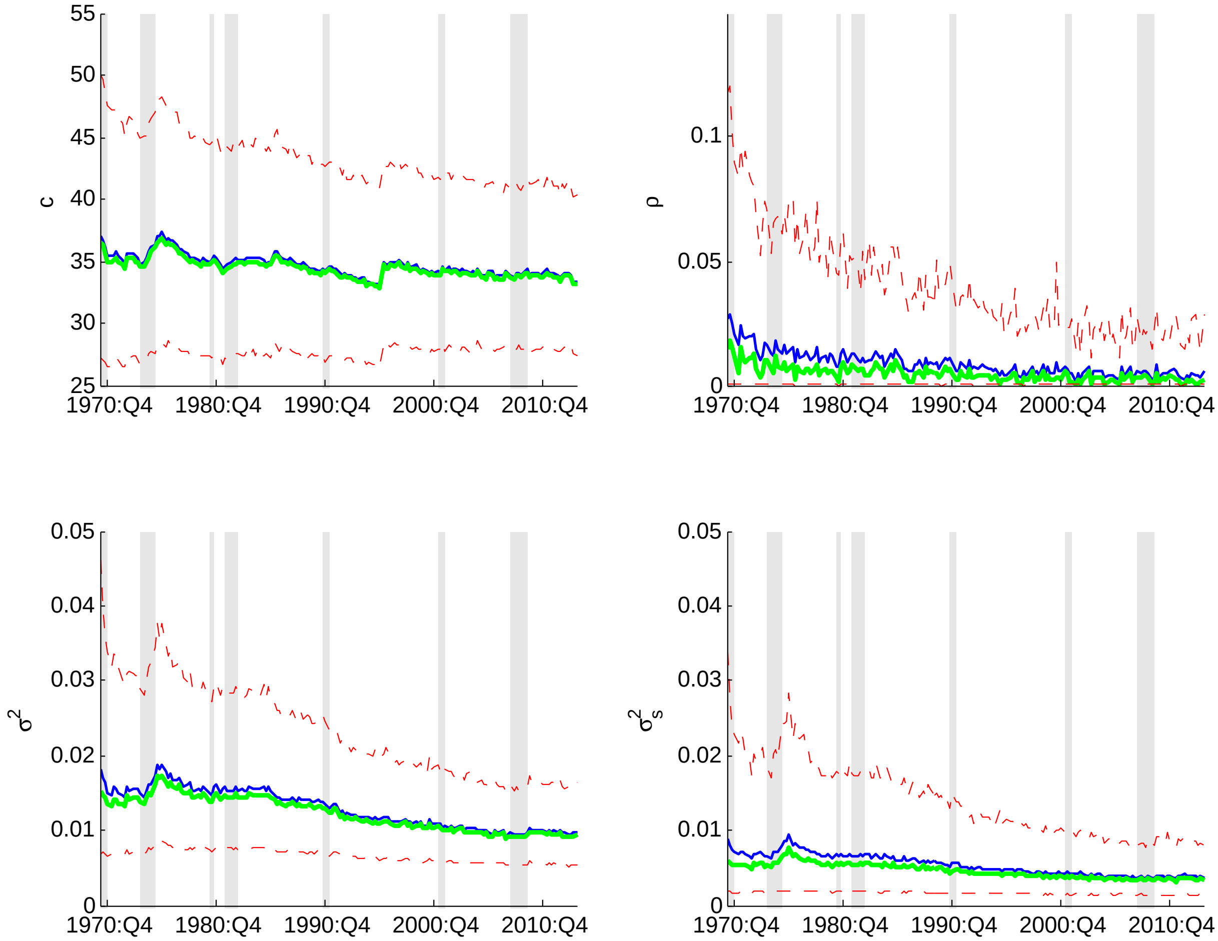

Figure 8: Skewed Model ($\mathcal{M}_1$) Parameters: Posterior Means, Medians, and 95% Credible Sets.