Uncertainty shocks as second-moment news shocks

David Berger, Ian Dew-Becker, and Stefano Giglio*

June 8, 2017

Abstract

This paper provides new empirical evidence on the relationship between aggregate uncertainty and the macroeconomy. We identify uncertainty shocks using methods from the literature on news shocks, following the observation that second-moment news is a shock to uncertainty. The key distinction we draw is between realized volatility – the realization of large shocks – and forward-looking uncertainty – the expectation that future shocks will be large. According to a wide range of VAR specifications, shocks to realized stock market volatility are contractionary, while shocks to uncertainty have no significant effect on the economy. In line with those findings, investors have historically paid large premia to hedge shocks to realized volatility, but the premia associated with shocks to uncertainty have not been statistically different from zero. We argue that these facts, and the VAR identification, are consistent with a simple model in which output growth is skewed left. Aggregate volatility matters, but it is the *realization* of volatility, rather than uncertainty about the future, that seems to be associated with declines.

*Berger: Northwestern University and NBER. Dew-Becker: Northwestern University and NBER. Giglio: University of Chicago and NBER. This paper was previously circulated and presented under the title "Contractionary volatility or volatile contractions?" We appreciate helpful comments and discussions from Robert Barsky, Nick Bloom, Gideon Bornstein, Brent Bundick, Nicolas Crouzet, Larry Christiano, Guido Lorenzoni, Aytek Malkhozov, Amir Yaron, and seminar participants at Booth, CITE, SITE, the Federal Reserve Banks of Chicago and Boston, Northwestern, the BI-SHoF finance conference, LBS, the NBER Summer Institute, UCL Workshop on Uncertainty, the SED, Arizona State, Yale, Texas A&M, and the SFS Cavalcade.

1 Introduction

A growing literature in macroeconomics studies the effects of news shocks on the economy. Models with rational forward-looking agents imply that pure changes in expectations about the future – news shocks – can induce a response in the aggregate economy. The existing literature has focused on *first-moment* news shocks: news about the average future path of the economy. For example, the literature on total factor productivity (TFP) and real business cycles identifies two types of TFP shocks: surprise innovations in TFP, and news about the future level of TFP that has no effects on TFP on impact. Empirically, the literature has documented important differences in how the economy responds to the two shocks (Beaudry and Portier (2006), Barsky and Sims (2011), and Barsky, Basu, and Lee (2015)).

This paper contributes to the news shock literature by extending the estimation to *second-moment* news shocks. Whereas the work described so far studies changes in the expected future growth rates, we study changes in expected future *squared* growth rates. News about the expectation of future squared growth rates represents a change in the conditional variance – that is, it is an *uncertainty shock*. Beaudry and Portier (2014) in fact suggest precisely this conceptualization of uncertainty shocks, and we use it to obtain estimates of the real effects of uncertainty shocks. Our goal is to use a news shock identification scheme to study the effects of uncertainty on the real economy.[1]

The analogy to news shocks is central to our analysis. In studies of first-moment news shocks, a distinction is made between innovations in the current level of TFP and innovations in expected future growth rates that are orthogonal to current TFP (i.e. shocks on date t that affect $E_t TFP_{t+1}$ but have no effect on TFP_t). That is, the aim is to identify responses to pure news shocks that affect expectations of future growth rates but have no effect on TFP on impact.

In the context of second-moment news shocks, similarly, we must distinguish between current squared growth rates and news about future squared growth rates, i.e. between $(\Delta TFP_t)^2$ and $E_t\left[(\Delta TFP_{t+1})^2\right]$. For reasons discussed below, we measure second moments using aggregate stock returns instead of TFP. So our uncertainty shock is an increase in the variance of the conditional distribution of future stock prices, while the analog to the first-moment impact shock is the surprise in the size of the squared change in stock prices – realized volatility – in the current period.

The key distinction between the two shocks is that realized volatility – squared stock returns in period t – is not the same as uncertainty about the future, which is instead equal to the expectation of *future* squared stock returns.[2] Models of the effects of uncertainty, such as those with wait-and-

[1] For past work on the relationship of uncertainty and the business cycle, see, among others, Alexopoulos and Cohen (2009), Bachmann and Bayer (2013), Bachmann, Elstner, and Sims (2013), Bachmann and Moscarini (2012), Baker and Bloom (2013), Baker, Bloom, and Davis (2015), Basu and Bundick (2015), Bloom (2009), Born and Pfeifer (2014), Caldara et al. (2016), Fernandez-Villaverde et al. (2011) , Fernandez-Villaverde et al. (2013), Leduc and Liu (2016), and Ludvigson, Ma, and Ng (2016). The theoretical literature has developed numerous mechanisms through which uncertainty about the future could affect the economy, such as precautionary saving demand among households (e.g. Basu and Bundick (2015) and wait-and-see behavior in firm investment (Bloom (2009)) or labor markets (Leduc and Liu (2016)).

[2] Distinguishing realizations and expectations is particularly important in light of the existing empirical literature,

see or precautionary saving effects, are driven by variation in agents' subjective distributions of future shocks, as opposed to the realization of volatility itself. The importance of that distinction is part of the basic message of this paper; to the best of our knowledge, we are the first to highlight the importance of distinguishing the two shocks when conducting empirical studies of the effects of uncertainty shocks.

The analysis focuses on the effects of uncertainty about the aggregate stock market (the S&P 500 index). Our concept of uncertainty therefore refers to the aggregate value of the largest firms in the US economy. Stock returns have the advantage that they should be affected by any shock that affects the value of firms. Furthermore, the availability of high-frequency data allows accurate calculation of realized volatility, and option prices provide information about expected volatility. Finally, measures of stock market volatility such as the VIX have been widely used in past research on uncertainty shocks, making it easy to compare our work to the existing literature. Other data sources typically only allow one to calculate either uncertainty (e.g. the surveys studied in Bachmann, Elstner, and Sims (2013)) or realized volatility (e.g. the cross-sectional variance of income growth in Storesletten, Telmer, and Yaron (2007)), but not both.

Technically, we use the identification scheme of Barsky, Basu, and Lee (2015), which identifies a news shock in a VAR as the rotation of the reduced-form shocks that predicts the future level of TFP (in our case, the sum of squared future stock returns) and is also orthogonal to current TFP (in our case, orthogonal to contemporaneous squared stock returns). In order for the identification to have any power, the VAR must include data that contains information about future volatility. We therefore include measures of option-implied volatility in the VAR. Unlike in past work, we do not necessarily assume that options directly measure expectations of future volatility. Rather, the identification just requires that they contain information about expectations; they can also have other components, such as time-varying risk premia.

Across a range of VAR specifications, increases in contemporaneous realized volatility are associated with declines in output, consumption, investment, and employment, consistent with the empirical findings in Bloom (2009), Basu and Bundick (2015), and Leduc and Liu (2016). But whereas those papers argue that their findings are driven by uncertainty, our robust finding is that it is *realized volatility* that is robustly followed by contractions. In fact, the uncertainty/second-moment news shock is, across a range of specifications, estimated to have no significant effect on the real economy, even though it accounts for more than a third of the variation in overall uncertainty and is strongly correlated with declines in stock prices. In some specifications uncertainty shocks are mildly contractionary, in others they are actually expansionary, but in no case are they statistically significant. In other words, there is no evidence in the data under our identification scheme that a second-moment news shock has any negative effect on the economy. And the difference between the responses of the economy to the realized and expected volatility shocks is itself statistically

which has effectively ignored their difference. While theoretical models are purely about forward-looking uncertainty – the variance of the conditional distribution of future outcomes – the data that has been studied is frequently about *realizations* of volatility. Bloom (2009), in fact, uses realized volatility in a VAR as a proxy for forward-looking uncertainty when the VIX is unavailable.

significant in our benchmark specification, indicating that the failure to find news shocks to be contractionary is not simply due to low statistical power.

The news shocks are also not small. They have statistically significant forecasting power for future stock market volatility at horizons of one to two years (which is typical for stock market volatility and similar to the length of the uncertainty shocks measured by Bloom (2009)), they account for 30–60 percent of the total variance of uncertainty, and we show in regressions that option-implied volatility contributes as much to variation in expectations of future volatility as lags of volatility itself do. In other words, option market investors appear to have economically meaningful information about future uncertainty that is not contained in the time series of past realized volatility. It is that information that drives our identification.

If uncertainty shocks are not actually contractionary, then realized volatility should not affect the economy through an uncertainty channel. What then explains why realized volatility is associated with contractions? There are two theoretical possibilities: realized volatility might have negative effects through some other channel, or contractions and realized volatility might be jointly caused by a third factor. We provide evidence in support of the second possibility.

The last section of the paper presents a simple model in which fluctuations in economic activity are negatively skewed and stochastically volatile.[3] When we estimate the same VAR in the model that we estimate in the data, we find highly similar results – identified uncertainty shocks have trivial effects on output, while the realized volatility shocks are contractionary, with a similar magnitude to what is observed empirically. Moreover, the identified shocks in the simulated VAR are strongly correlated with the simulated structural shocks, providing theoretical support for our identification scheme (Basu and Bundick (2015) use a similar argument in support of their identification scheme).

There are two important further pieces of evidence in favor of the skewness hypothesis. First, changes in a wide variety of measures of real activity are negatively skewed in the data, as are stock returns.[4] Second, investors have historically paid large premia for insurance against high realized volatility and extreme negative stock returns in the last 30 years, whereas the premium paid for protection against increases in *expected* volatility (uncertainty) has historically been near zero or even positive.[5] Those findings are consistent with uncertainty having no effects on the economy in equilibrium. We show that the model qualitatively matches both the empirical left skewness and the large premium on realized volatility compared to shocks to volatility expectations (though the

[3]Skewness in equilibrium quantities could arise because the fundamental shocks are skewed, or because symmetrical shocks are transmitted to the economy asymmetrically (perhaps because constraints, such as financial frictions, bind more tightly in bad times; see Kocherlakota (2000), or because firms respond to shocks in a concave manner, as in Ilut, Kehrig, and Schneider (2016)).

[4]Recent empirical work on skewness includes Morley and Piger (2012), Giglio, Kelly and Pruitt (2016), Adrian, Boyarchenko, and Giannone (2016), Salgado, Guvenen, and Bloom (2016), and Ilut, Kehrig, and Schneider (2017).

[5]See for example Egloff, Leippold, and Wu (2010); Ait-Sahalia et al. (2015); Dew-Becker et al. (2017). A large literature in finance studies the pricing of realized and expected future volatility. See, among many others, Adrian and Rosenberg (2008), Bollerslev et al. (2009), Heston (1993), Ang et al. (2006), Carr and Wu (2009), Bakshi and Kapadia (2003), Egloff, Leippold, and Wu (2010), and Ait-Sahalia, Karaman, and Mancini (2013) (see Dew-Becker et al. (2017) for a review).

magnitude of the risk premia is somewhat smaller in the model than the data, as is common in models of the business cycle).

To be clear, our claim is not that uncertainty does not matter for the economy. There is an incredibly wide range of different measures of uncertainty. We focus on stock market volatility for the reasons discussed above, but other types of uncertainty certainly might matter, such as policy, interest rate, or idiosyncratic uncertainty, and applying the ideas in this paper to those concepts would represent an interesting extension to the present work.

In addition to the macroeconomic studies discussed above, our work is also closely related to an important strand of research in finance. It has long been understood in the asset pricing literature that expected and realized volatility, while correlated, have important differences (e.g. Andersen, Bollerslev, and Diebold (2007)). A jump in stock prices, such as a crash or the response to a particularly bad macro data announcement, mechanically generates high realized volatility. On the other hand, news about future uncertainty, such as an approaching presidential election, increases expected volatility (Kelly, Pastor, and Veronesi (2016)). Shocks to realized and expected future volatility are correlated, but they are not as strongly correlated as one might expect – in our sample, 60–70 percent. This means that it is possible to identify in the data shocks to expectations that are orthogonal to realizations, and it is well established that they are priced differently by investors (Broadie, Chernov, and Johannes (2007)).

Our work is related to a large empirical literature that studies the relationship between aggregate volatility and the macroeconomy noted above. A wide range of measures of volatility in financial markets and the real economy have been found to be countercyclical.[6] To identify causal effects, a number of papers use VARs, often with recursive identification, to measure the effects of volatility shocks on the economy.[7] Ludvigson, Ma, and Ng (2015), like us, distinguish between different types of uncertainty. They show that variation in uncertainty about macro variables is largely an endogenous response to business cycles, whereas shocks to financial uncertainty cause recessions.[8] Similarly, Caldara et al. (2016) use a penalty-function based identification scheme to distinguish between the effects of uncertainty and financial conditions. A key distinction between our work and those two papers is that we focus on the distinction between uncertainty expectations and realizations. Moreover, unlike most past work (Ludvigson, Ma, and Ng (2015) and Caldara et al.

[6] Gilchrist, Sim, and Zakrajsek (2014) use the same fact as a starting point for an analysis of volatility, irreversible investment, and financial frictions. See, among others, Campbell et al. (2001) (equity volatility at the index, industry, and firm level is countercyclical); Storesletten, Telmer, and Yaron (2004) and Guvenen, Ozkan, and Song (2014) (household income risk is countercyclical); Eisfeldt and Rampini (2006) (dispersion in industry TFP growth rates is countercyclical); Alexopoulos and Cohen (2009) and Baker, Bloom, and Davis (2015) (news sources use uncertainty-related language countercyclically).

[7] See Bloom (2009) and Basu and Bundick (2015), who study the VIX; and Baker, Bloom, and Davis (2015) and Alexopoulos and Cohen (2009), who study news-based measures of uncertainty. Jurado, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (2015) measure uncertainty based on squared forecast errors for a large panel of macroeconomic time series (using a two-sided filter to extract a latent volatility factor). Baker and Bloom (2013) use cross-country evidence to argue that there is causal and negative relationship between uncertainty and growth.

[8] Other papers arguing that causality could run from real activity to volatility and uncertainty include Decker, D'erasmo and Boedo (2016), Berger and Vavra (2013), Ilut, Kehrig and Schneider (2015), and Kozlowski, Veldkamp, and Venkateswaran (2016).

(2016) excepted), our identification scheme builds on the news shock literature, rather than using a more restrictive recursive setup.

2 Inference

This section describes how we measure second-moment news shocks in the data. The feature of the data that we want to measure is the variance of the flow of aggregate shocks that hit the economy. Empirically, then, we focus on uncertainty about the future level of the stock market. Equity prices are useful for summarizing information about the future path of the economy. Furthermore, the analysis eschews a strict structural interpretation. The goal is simply to estimate the dynamics of the economy *following* an unexpected increase in uncertainty; that is, we have the conservative goal of estimating reduced-form impulse responses, which represent changes in conditional means (e.g. Hamilton (1994), Jordà (2005)).

2.1 Conditional variances

Denote the log of the total return stock index as s_t. Uncertainty about the future value of the stock market relative to its value today is measured in this paper as the conditional variance, $Var_t\left[s_{t+n}\right]$.

The one-period log stock return is $r_t \equiv s_t - s_{t-1}$. If returns are uncorrelated over time and time periods are sufficiently short that $E_t r_{t+1} \approx 0$, we have:

$$Var_t\left[s_{t+n}\right] = E_t\left[\sum_{j=1}^{n-1} r_{t+j}^2\right] - \sum_{j=1}^{n-1} E_t\left[r_{t+j}\right]^2 \tag{1}$$

$$\approx E_t\left[\sum_{j=1}^{n-1} r_{t+j}^2\right] \tag{2}$$

When returns are serially uncorrelated (which is very nearly true empirically, especially at short horizons), the conditional variance of stock prices on some future date is equivalent to the expected total variance of returns over that same period.[9] Under standard conditions on the returns process, as the length of a time period approaches zero, the second line becomes an equality. Writing the conditional variance in (2) as an expectation directly connects the analysis to the news shock literature, which studies changes in expectations.

Whereas the literature on news about TFP studies $E_t\left[\sum_{j=1}^{n} \Delta tfp_{t+j}\right]$ where Δtfp is the first difference of log TFP, here we study second-moment expectations: the expectation of future squared returns ($r^2 = (\Delta s)^2$), which is simply the conditional variance of future stock prices. Second-moment news shocks – uncertainty shocks – are shifts in expected future squared returns.

[9] In practice, we work with daily returns where the zero-mean approximation holds strongly, as documented in the literature. In the notation of continuous time models, $E_t\left[r_{t+1}^2\right]$ is $O\left(\Delta t\right)$, while $E_t\left[r_{t+1}\right]^2$ is $O\left(\Delta t^2\right)$, where Δt is the length of a time period. So as the time period shrinks, the terms involving squared expected returns become negligible.

In the literature on TFP news shocks, there is also the contemporaneous innovation in TFP, $tfp_t - E_{t-1}tfp_t$. The analog here is the innovation in *realized volatility,* $r_t^2 - E_{t-1}\left[r_t^2\right]$. The conditional variance of future stock prices, $Var_t\left[s_{t+n}\right]$, is equal (when returns are calculated at high frequency) to cumulative expected future realized volatility.

So the analysis parallels the first-moment news shock literature closely. Anywhere past work talks about Δtfp, it is replaced here with r^2, both when looking at realization shocks and at news. First-moment news shocks are changes in the expectation of future values of Δtfp, holding constant the current innovation in Δtfp. Second-moment news shocks are changes in the expectation of future values of r^2, holding constant the current innovation in r^2 (current realized volatility).

One last minor issue is that we have data on daily stock returns, but data on real activity only at the monthly level. We therefore aggregate volatility to a monthly frequency. Specifically, we define realized volatility in month t, RV_t, as

$$RV_t \equiv \sum_{days \in t} r_i^2 \tag{3}$$

We then have

$$Var_t\left[s_{t+n}\right] \approx E_t\left[\sum_{j=1}^{n} RV_{t+j}\right] \tag{4}$$

Again, the approximation is only due to discreteness – if we had truly continuous data instead of sampling only at the daily level, (2) and (4) would hold exactly. Given how small average daily stock returns are (less than 0.05 percent), the approximation errors here are quantitatively unimportant.

2.2 Vector autoregressions

We now discuss how we measure second-moment news shocks using a vector autoregression (VAR) structure similar to the existing first-moment news literature. The purpose of the VARs is to help identify innovations in uncertainty and then contruct reduced-form impulse response functions.

We estimate VARs of the form

$$\begin{bmatrix} RV_t \\ Y_t \end{bmatrix} = C + F\left(L\right)\begin{bmatrix} RV_{t-1} \\ Y_{t-1} \end{bmatrix} + A\varepsilon_t \tag{5}$$

where RV_t is realized volatility from (3), Y_t is a vector including measures of real activity, variables that help forecast future values of realized volatility, and other controls, C is a vector of constants, and $F\left(L\right)$ is a matrix polynomial in the lag operator, L. ε_t is a vector of uncorrelated innovations with unit variances and A is the lower-triangular Cholesky factorization of the variance matrix of the reduced-form innovations. (5) can be estimated by ordinary least squares. The VAR has a moving average (MA) representation,

$$\begin{bmatrix} RV_t \\ Y_t \end{bmatrix} = (I - F(1))^{-1} C + B(L) A\varepsilon_t \tag{6}$$

$$\text{where } B(L) = \sum_{j=0}^{\infty} B_j L^j = (I - F(L))^{-1} \tag{7}$$

We do not claim to identify structural shocks. Rather, our goal is simply to estimate the average dynamics of the economy *following* changes in realized volatility and uncertainty shocks. The impulse responses to the realized volatility shock are measured as

$$\frac{\partial E\left[Y_{t+j} \mid RV_t, RV^{t-1}, Y^{t-1}\right]}{\partial RV_t} = B_j A_{(:,1)} \tag{8}$$

where $A_{(:,1)}$ denotes the first column of A. E denotes the expectation operator conditional on the VAR (5) and RV^{t-1} and Y^{t-1} are the histories of RV and Y up to date $t-1$. As in Hamilton (1994), the impulse response is defined as the average change in expectations for the future following a unit innovation in realized volatility. It tells us *on average* how the economy (output, employment, etc.) changes when realized volatility changes – it does not identify a structural shock.

The second estimated shock is the residual innovation in uncertainty over the following n periods, $Var_t\left[s_{t+n}\right]$ or, equivalently, the residual innovation in expectations of future volatility. It is the component of $E_t \sum_{j=1}^{n} RV_{t+j} - E_{t-1} \sum_{j=1}^{n} RV_{t+j}$ that is orthogonal to $RV_t - E_{t-1}RV_t$. The total change in cumulative expected volatility up to time $t+n$ is constructed as

$$E_t \sum_{j=1}^{n} RV_{t+j} - E_{t-1} \sum_{j=1}^{n} RV_{t+j} = \left(e_1 \sum_{j=1}^{n} B_j\right) A\varepsilon_t \tag{9}$$

where $e_1 = [1, 0, ...]$. The parameter n determines the horizon over which the news shock is calculated. Cumulative expected volatility depends on the sum of the first rows of the MA matrices up to lag n. The innovation to expectations over horizon n, i.e. the news about future volatility, is then simply the linear combination of shocks represented by $e_1 \sum_{j=1}^{n} B_j A$. As in Barsky, Basu, and Lee (BBL; 2014) and Barsky and Sims (2011), we then orthogonalize that linear combination with respect to the innovation to RV_t so that the impact shock to RV_t is uncorrelated with the news shock.[10] The BBL method is only partially identified in that it identifies two shocks and is unstructured otherwise.

So the uncertainty shock, denoted u_t, is equal to the component of $\left(e_1 \sum_{j=1}^{n} B_j\right) A\varepsilon_t$ that is orthogonal to $e_1 A\varepsilon_t$, and the impulse responses to the uncertainty shock are defined as

$$\partial E\left[Y_{t+j} \mid u_t, RV_t, RV^{t-1}, Y^{t-1}\right] / \partial u_t$$

[10] The orthogonalized innovation is simply $e_1 \sum_{j=1}^{n} B_j \bar{A}\varepsilon_t$, where $\bar{A}$ is equal to A but with the first column set to zero.

Again, the VAR yields the average behavior of the economy following a unit increase in uncertainty (conditional on RV_t), as measured by u_t. The measure of uncertainty here thus obviously depends on the variables included in the VAR. Our measure of volatility news is, strictly, the change in expectations of future realized volatility conditional on the vector $[RV_t, Y_t']'$ and the estimated VAR coefficients.

One might naturally ask whether it is even possible for there to be variation in uncertainty that is orthogonal to realized volatility. Shouldn't an increase in uncertainty drive asset prices down, thus generating realized volatility? In general that is true, but what is important to note is that the rate of change in uncertainty will determine the extent to which it appears in realized volatility. Suppose the conditional variance of future stock returns rises by 10 percent over a month. If that rise comes all in a single day, it will be associated with high realized volatility. But if the rise comes smoothly over the month, then realized volatility will be much lower, due to the fact that it is calculated based on daily squared returns. So one source of independent variation in realized and expected volatility (though certainly not the only one) comes from whether shocks come as "jumps" or as slower diffusive changes over time.

Another important concern with the identification scheme used here is that it could be prone to overfitting volatility news. In our most general model, we will have six variables and four lags, meaning that future volatility is being forecast with 24 degrees of freedom. We will thus argue that it is important to restrict the model somewhat to alleviate overfitting and ensure that the results are consistent with economic priors.

The two impulse responses defined here are only identified up to some normalization. We rescale them so that they have the same effect on uncertainty. Specifically, denote responses to unit shocks as $g_{j,k,s}$, where $g_{j,k,s}$ is the response of variable j to shock k at horizon s. We report normalized IRFs of the form

$$\hat{g}_{j,k,s} \equiv \frac{g_{j,k,s}}{\sum_{m=1}^{n} g_{1,k,m}} \tag{10}$$

The scaling factor in the denominator is the cumulative expected effect of shock k on future RV_t up to horizon $t+n$. In this way, the IRFs we report are scaled so that they all have a unit effect on uncertainty about the level of stock prices in period $t+n$: they contain the same amount of news about future uncertainty. In our empirical work, we set $n = 24$ months, which is the horizon over which we examine IRFs (past work finds that volatility shocks have half-lives of 6–12 months, so 24 months represents the point at which the average shock has dissipated by 75 percent or more). A larger value of n reduces power, since long-term effects are relatively difficult to estimate, while a smaller value identifies shocks that may be less relevant for economic decisions. We find similar results for n ranging between 12 and 60 months.

2.3 Information about uncertainty

Obviously in order to identify a news shock, the vector of state variables in the VAR, Y_t, must contain information that can reveal expectations of future volatility beyond what is contained in current RV_t (u_t must have a component independent of the innovation to RV_t). We therefore include information from financial markets. First, we use $V_{1,t}$, the option-implied volatility of stock returns over the next month (we define $V_{1,t}$ very similarly to the VIX). Since $V_{1,t}$ may not include all the available financial information about uncertainty, we also include the six-month implied volatility, $V_{6,t}$, in some specifications.

Importantly, there is no assumption here that risk premia are zero or constant or that the option-implied volatility is measured without error. The only assumption that we need to calculate the impulse responses defined above is that some elements of Y_t contain information about future values of RV beyond the innovation to RV_t itself. We include option-implied volatilities because we would expect them to contain such information, but they are obviously also contaminated by risk premia and potential measurement error (e.g. due to stale prices or bid/ask spreads). Below we examine a number of other variables that past work has found can also help forecast volatility, but we find that their predictive power is subsumed by that in lagged RV and V_1.

Duffee (2011) shows that in standard linear term structure models, except for in knife-edge cases (even allowing for time-varying risk premia), investor expectations of the future can be extracted from observed asset prices. In our setting, that result corresponds to the view that the current state of the term structure of option-implied volatilities and RV_t should, together, encode all available information about future values of RV and option-implied volatility. We show that idea is in fact consistent with our data (even though Duffee (2011) argues it appears to be violated in the bond market), which will allow us to impose useful coefficient restrictions on the VAR and further reduce overfitting.

3 Data

3.1 Macroeconomic data

We focus on monthly data to maximize statistical power, especially since fluctuations in both expected and realized volatility are sometimes short-lived (e.g. Bloom (2009)). We measure real activity using the Federal Reserve's measure of industrial production (IP) for the manufacturing sector. Employment and hours worked are measured as those of the total private non-farm economy.

3.2 Financial data

We obtain data on daily stock returns of the S&P 500 index from the CRSP database and use it to construct RV_t at the monthly frequency. Option-implied volatilities, $V_{n,t}$, are constructed using prices of S&P 500 options obtained from the Chicago Mercantile Exchange (CME), with traded maturities from one to at least six months since 1983 (we thus have a substantially longer sample of implied voaltility than has been used in past work). Given that shocks to stock market volatility are typically short-lived, with half lives often estimated to be on the order of six to nine months (see Bloom (2009) and Drechsler and Yaron (2011)), one- to six-month options will contain information about the dominant shocks to stock market uncertainty.

Using results from Bakshi, Kapadia, and Madan (2003) it is straightforward to show that the variance of the index under the pricing measure Q can be written as a function of option prices,[11]

$$V_{n,t} \equiv Var_t^Q[s_{t+n}] \tag{11}$$

$$= 2\int_0^\infty \frac{1-\log\left(\frac{K}{e^{rt}S_t}\right)}{B_t(n)K^2}O(K)\,dK - \left(e^{rt}\int_0^\infty \frac{O(K)}{B_t(n)K^2}dK\right)^2 \tag{12}$$

Note that this formula holds generally, requiring only the existence of a well-behaved pricing measure; there is no need to assume a particular specification for the returns process. $Var_t^Q[s_{t+n}]$ is calculated as an integral over option prices, where K denotes strikes, $O_t(n, K)$ is the price of an out-of-the-money option with strike K and maturity n, and $B_t(n)$ is the price at time t of a bond paying one dollar at time $t+n$. $V_{n,t}$ is equal to the option-implied variance of log stock prices n months in the future. Computing $V_{n,t}$ with real-world data requires several steps; the appendix provides a description of our calculation methods and analyzes the accuracy of the data.

We use the measure of $Var_t^Q[s_{t+n}]$ in (11) because it maps directly to our object of interest, the conditional variance of log stock prices. In practice, it is extremely close to the VIX calculated by the CBOE and other related model-free implied volatility measures (the VIX is also a measure of variance under the pricing measure, but requires stronger assumptions about the returns process).

Finally, in the remainder of the paper we focus on the logs of realized and option-implied volatility ($rv_t \equiv \log RV_t$, $v_{n,t} \equiv V_{n,t}$). Given the high skewness of realized volatility, the log transformation makes the results less dependent on the occasional volatility spikes. We nevertheless also show in section 5.2.3 that our results are robust to performing the analysis in levels.

3.3 The time series of uncertainty and realized volatility

Figure 1 plots the history of realized volatility along with 1-month option-implied volatility in annualized standard deviation terms. Both realized and option-implied volatility vary considerably over the sample. The two most notable jumps in volatility are the financial crisis in 2008 and the 1987 market crash, which both involved realized volatility above 75 annualized percentage points and rises of $V_{1,t}$ to 65 percent. At lower frequencies, the periods 1997–2003 and 2008–2012 are associated with persistently high uncertainty, while it is lower in other periods, especially the early 1980's, early 1990's, and mid-2000's. There are also distinct spikes in uncertainty in the summers of 2010 and 2011, likely due to concerns about the stability of the Euro zone and the willingness of the United States government to continue to pay its debts.

In terms of raw levels, volatility is countercyclical overall, though somewhat weakly. Realized

[11] The pricing measure, Q, is equal to the true (or physical) pricing measure multiplied by $M_{t+1}/E_t[M_{t+1}]$, where M_{t+1} is the pricing kernel. The result for $Var_t^Q[s_{t+n}]$ is obtained from equation 3 in Bakshi, Kapadia, and Madan (2003) by first setting $H(S) = \log(S)$ to obtain $E_t^Q[\log S_{t+n}]$ and then defining $G(S) = \left(\log(S) - E_t^Q[\log S_{t+n}]\right)^2$ and inserting it into equation 3 in place of H.

and option-implied volatility are both positively correlated with the unemployment rate (~0.3), negatively correlated with capacity utilization (~-0.44), and negatively correlated with returns on the S&P 500 (~-0.31).[12]

4 Second-moment forecasting regressions

Since identification of the second-moment news shock depends on using the variables in the VAR to forecast future realized volatility, a natural first question is which of those variables, if any, has forecasting power for rv. Table 1 reports results of regressions of $\sum_{j=1}^{6} rv_{t+j}$ on various predictors. The first column reports results from a regression on rv_t and $v_{1,t}$. We see that $v_{1,t}$ has a substantially larger t-statistic, indicating that it has greater explanatory power. The marginal R^2 of $v_{1,t}$ is 0.0554, while that of rv_t is smaller by a factor of eight at 0.0069. rv_t is in fact only marginally significant at the ten-percent level. In other words, option-implied volatility is a substantially stronger predictor of future realized volatility than the current level of realized volatility is. Given that asset prices typically aggregate information efficiently, it is not surprising that $v_{1,t}$ nearly drives rv_t out in a forecasting regression.

The second column of table 1 shows that we obtain a similar result when we include a lagged value of rv. In fact, if we include 6 lags of rv, their combined marginal R^2 is still only half the marginal R^2 of $v_{1,t}$ (0.0274). So in simple forecasting regressions, we find that option prices yield substantially better predictions of future volatility than the past history of volatility does. If one thought that realized volatility followed a simple autoregressive process, then the current and lagged values would yield a sufficient statistic for expectations about the future, and v_1 would have no marginal predictive power. The results reported in table 1 show that option prices contain information about uncertainty – expected future realized volatility – above and beyond what is contained in the history of stock market volatility, and that the predictive power from v_1 in terms of marginal R^2 is in fact *far larger* than that of rv.[13]

The third column of table 1 adds the six-month implied volatility, $v_{6,t}$, and finds that it adds no incremental information. We obtain similar results using a principal component of the term structure of implied volatilities.

The fourth and fifth columns of table 1 add macroeconomic and financial variables to the regressions. None of them are individually statistically significant, nor are they jointly significant. The fifth column includes principal components from the large set of financial and macroeconomic time series collected by Ludvigson and Ng (2007), as well as the market return and the default spread (difference between the yields of Baa and Aaa corporate bonds). None of them has statistically significant forecasting power after controlling for rv_t and $v_{1,t}$, so we exclude them from the remainder of the analysis.

[12] The numbers reported here are for implied volatility and after detrending the unemployment and capacity utilization rates with a one-sided HP filter with a smoothing parameter of 1.296×10^7.

[13] Note that this does not mean that *uncertainty* cannot follow a univariate process. A simple model consistent with these regressions is that uncertainty (expected rv) follows an AR(1), and rv is equal to uncertainty plus noise.

The R^2s are similar across all the specifications, and always 0.46 or less. The majority of the variation in six-month realized stock market volatility is thus unpredictable, even given information available at the beginning of the period.

It is perhaps notable that we do not find any variables beyond lagged rv and v_1 to be significant in predicting future volatility. Table A.1 in the appendix shows that when rv and v_1 are excluded from the regression, a number of the macroeconomic and financial variables become significant predictors of future volatility. In other words, the macro and financial time series on their own can help predict future volatility, but their forecasting power is subsumed by current realized and option-implied volatility.

5 VAR results

We now report our main VAR. For all the VARs that we run, we include four lags, as suggested by the Akaike information criterion for our main specification. In the main results, the vector of variables included in the VAR is $[rv_t, v_{1,t}, FFR_t, ip_t, emp_t]$, where the latter three variables are the Fed Funds rate, log industrial production, and log employment, respectively.

The benchmark specification imposes the restriction that the coefficients on the lags of FFR, ip, and emp in the equations for rv and v_1 are equal to zero, consistent with the predictive regression results above. Table A.2 in the appendix shows, in regressions analogous to those in table 1, that v_1 is predicted only by its own lags and those of rv_t (again consistent with the theoretical predictions of Duffee (2011) for linear factor pricing models). In addition to their economic motivation, the restrictions help keep the news shock from being overfit, which we show can be a problem in section 5.3.2. These restrictions together imply that in the benchmark specification the uncertainty shock is equivalent to the reduced-form shock to v_1, orthogonalized with respect to rv. We relax the restrictions below and show that our results remain similar.

5.1 Coefficient estimates

Before reporting impulse responses it is useful to examine the coefficients in the VAR. In our benchmark specification, the sums of the coefficients on lags of rv_t and $v_{1,t}$ in the equations for employment and industrial production are

	Sum of coefficients			
Dep. var.:	rv	v_1	Diff.	p-value
Employment	-0.0017	0.0012	0.0028	0.063
Indus. prod.	-0.0034	0.0024	0.0057	0.285

The coefficients on rv are negative for both variables, while the coefficients on v_1 are actually *positive*. That basic result appears consistently across the specifications that we estimate, and it gives the first indication of the different effects of realized and expected volatility on real outcomes.

5.2 Benchmark specification

We now examine impulse response functions (IRFs), which describe the average responses of the variables in the economy to the two innovations. As discussed above, the IRFs are scaled so that the two shocks – current rv and the identified uncertainty shock – have the same cumulative effect on volatility expectations 2–24 months in the future (i.e. not counting the impact period). That is, they are scaled so as to have the same impact on uncertainty about the level of stock prices two years in the future.

In the benchmark specification the vector of state variables is $[rv, v_1, FFR, ip, emp]$ and the coefficients on lagged FFR, ip, and emp in the equations for rv and v_1 are restricted to zero. A χ^2 test of the validity of those restrictions yields a p-value of 0.61, implying that they are consistent with the data.

Figure 2 has three columns for the responses of rv, employment, and industrial production to the shocks. Each panel shows the point estimate for the IRFs along with 1-standard deviation (68-percent) and 90-percent bootstrapped confidence intervals. The first row shows the response of the economy to the identified rv shock. It shows that a shock to realized volatility is very short-lived: the IRF falls by half within two months, and by three-fourths within five months, showing that realized volatility has a highly transitory component.

Those transitory increases in realized volatility are associated with statistically and economically significant declines in both employment and industrial production. So, consistent with past work, we find a significant negative relationship between *volatility* and real activity. However, this result does not allow us to conclude that an uncertainty shock is contractionary. The reason is that the realized volatility shock is a combination of an uncertainty shock (we can see from the first panel that the shock does predict future rv after impact, so it contains news about future volatility) with a shock to *current* realized volatility; by observing how the economy reacts to this combination of shocks we cannot draw conclusions about how it responds to a pure uncertainty shock.

The second row of panels in figure 2 plots IRFs for the identified uncertainty shock, which has no contemporaneous effect on rv, but is meant to capture uncertainty about the future. First, as we would expect from equation (4), the news shock forecasts high realized volatility in the future at a high level of statistical significance (p-values testing whether the IRF is positive range between 0.001 and 0.05). That result alone is important: it says that the identified news shock contains statistically significant news about uncertainty. That is, the market-implied conditional variance contains information about uncertainty even after controlling for current and past realized volatility.

Surprisingly, the second-moment news shocks are associated with no significant change in either employment or industrial production; the IRFs both stay very close to zero. Furthermore, the confidence bands are reasonably narrow: at almost all horizons, the point estimate for the responses of employment and industrial production to the rv shock are outside the 90-percent confidence bands for the uncertainty shock.

To further examine the magnitudes, the bottom row of panels in figure 2 reports the difference in the IRFs for the uncertainty and realized volatility shocks along with confidence bands. The two shocks have the same cumulative impacts on the future path of realized volatility (by construction, due to the scaling of the IRFs). But they obviously have different effects on rv on impact due to the identifying assumptions.

The two other panels in the bottom row of figure 2 plot the difference between the IRFs for industrial production and employment. We see that the difference is statistically significant for employment, weakly so for industrial production. So innovations in rv are followed by statistically significant declines in real activity, while uncertainty shocks are not, and that difference itself is statistically significant in one case.

An alternative way to interpret the bottom row of IRFs is that it represents the response of the economy to a pure shock to realized volatility that has no net effect on forward-looking uncertainty (compared to the second row, which is a shock to uncertainty that has no effect on realized volatility on impact). By construction, the shock in the third row has a positive effect on rv on impact, but the sum of the IRF for rv over the following 24 periods sums to zero. The figure therefore shows that such a shock has negative effects on the economy.

Overall, then, figure 2 shows that under our baseline identification scheme, shocks to rv are associated with statistically significant subsequent declines in output, while uncertainty shocks are not, and the difference between those two results is itself statistically significant. That result is notable given that past work (e.g. Bloom (2009), Basu and Bundick (2016), Leduc and Liu (2016)) has found that increases in option-implied volatility are followed by declines in output (a result we also obtain in our data; see appendix figure A.5). The results here show that it is actually *realizations* of volatility that seem to drive such effects, as opposed to shocks to uncertainty about the future. However, since realized volatility and option-implied volatility are positively correlated, using either of the two variables alone in a VAR as in past studies will lead the researcher to conclude that each shock is contractionary. To uncover the differential macroeconomic effects of realization and uncertainty shocks it is cricital to include both in the VAR.

5.2.1 Volatility and uncertainty

To further understand the importance of the uncertainty and rv shocks, figure 3 reports forecast error variance decompositions (FEVDs). As in figure 2, we report the effect of the rv shock, the uncertainty shock, and their difference.

The first column reports variance decompositions for $E_t\left[\sum_{j=1}^{24} rv_{t+j}\right]$, which is our measure of uncertainty. The figures show that at the point estimates the rv shock accounts for 65 percent of the variance of uncertainty, while the uncertainty shock accounts for the remaining 35 percent. The confidence bands are wide, though: at most horizons, we cannot reject the hypothesis that the rv and news shocks account for the same fraction of the variance of uncertainty at even the 32-percent significance level. Since in the benchmark model the only variable that is assumed to have predictive power for uncertainty beyond rv_t itself is $v_{1,t}$, this is our most conservative specification and provides a lower bound on the fraction of variation in uncertainty coming from news shocks. Even in this case, though, the news shock accounts for a substantial fraction of the total variation in uncertainty.

To further examine the relative importance of the two shocks for uncertainty, figure 4 plots fitted uncertainty from the VAR along with the parts driven by the rv and identified uncertainty shocks. Total uncertainty is calculated from the VAR as $E_t\left[\sum_{j=1}^{24} rv_{t+j}\right]$. Note that in the VAR, there exists a vector B such that

$$E_t\left[\sum_{j=1}^{24} rv_{t+j}\right] = B\left[\tilde{Y}_t', \tilde{Y}_{t-1}', ..., \tilde{Y}_{t-4}'\right]' \tag{13}$$

$$\text{where } \tilde{Y}_t \equiv \left[Y_t', RV_t\right]' \tag{14}$$

To find the part of uncertainty driven by the rv shock, then, we construct a vector time series $\tilde{Y}^{rv}$ using the VAR structure setting all the estimated shocks to zero except for the rv shock.[14] Similarly, $\tilde{Y}^{news}$ is constructed by setting all estimated shocks except the news shock to zero. The parts of uncertainty coming from rv and news are then $B\tilde{Y}^{rv}$ and $B\tilde{Y}^{news}$, respectively. Ignoring constants, total uncertainty equal to the sum of those two parts. Figure 4 plots total uncertainty and the parts from rv and news (all demeaned).

The standard deviation of the part of uncertainty coming from news is only 22 percent smaller than that coming from rv, which is visible from the similar overall variation in the two series. The quarters with the largest news shocks are associated with events that are clearly associated with uncertainty: the 1987 and 2008 market crashes, and the third quarter of 2011 (debt ceiling and Euro crisis). At lower frequencies, the news-driven component of uncertainty is high around the 1990 recession (i.e. around the First Gulf War), in the late 1990's (Asian financial crisis, Russian default, LTCM), and following the financial crisis (debt ceiling debates, Euro crisis).

To further evaluate the plausibility of the time series of uncertainty implied by the model, the table below reports the correlation between the three series in figure 4 and three other commonly used measures of uncertainty: the Economic Policy Uncertainty index of Baker, Bloom, and Davis (EPU; 2015), a measure of uncertainty from the Michigan Consumer Survey (used by Leduc and Liu (2016)), and forecast uncertainty measure from Jurado, Ludvigson, and Ng (JLN; 2015).

Correlations:

	EPU	Michigan	JLN
Total uncertainty	0.41	0.26	0.51
rv-driven uncertainty	0.40	0.22	0.76
News-driven uncertainty	0.09	0.12	-0.25

Our overall uncertainty measure is positively correlated with all three of the other measures of

[14] That is, if $\tilde{Y}_t = C + F(L)\tilde{Y}_{t-1} + A\varepsilon_t$, we define ε^{rv} to be equal to the fitted values of ε but with all elements except for the first set to zero. $\tilde{Y}^{rv}$ is then constructed as $\tilde{Y}^{rv} = C + F(L)\tilde{Y}_{t-1}^{rv} + A\varepsilon_t^{rv}$.

uncertainty, and that correlation is stronger for the rv-driven part. The pure uncertainty shocks that we identify here, that are separate from realized volatility, are essentially uncorrelated with the other uncertainty measures and therefore seem to capture feature of the data that is independent of them.

That said, and crucially for our analysis, the news shock is also not just noise. Again, it forecasts realized volatility significantly into the future, accounting for a substantial fraction of the variation in total uncertainty. Moreover, the identified news shock is strongly correlated with returns on the S&P 500. Specifically, the correlations of the identified rv and uncertainty shocks with S&P 500 returns are:

Correlations with the S&P 500:	
rv shock:	-0.412
Uncertainty shock:	-0.405

That is, the correlations with stock returns are almost exactly the same. Both rv and the uncertainty shock (which are uncorrelated with each other by construction) have substantial negative correlations with stock returns. So not only is the uncertainty shock significantly associated with future volatility, but it is also associated with substantial declines in stock returns. These facts make it all the more surprising that the uncertainty shock is not associated with any significant change in real activity.

5.2.2 Forecast error variance decompositions for employment and IP

In addition to the FEVDs for uncertainty, figure 3 also reports FEVDs for the two real outcomes, employment and industrial production. The realized volatility shock explains 25 percent of the variance of employment and 10 percent of the variance of industrial production at the two-year horizon, while the point estimates for the fraction of the variance accounted for by second-moment news are two percent or less, and the upper ends of the 90-percent confidence intervals are near 5 percent for the first year. The upper end of the 90-percent confidence interval for the rv shock, though, reaches as high as 45 percent for employment and 30 percent for industrial production 24 months ahead, indicating that realized volatility can potentially be an important driver of the real economy (though this is *not* a causal statement).

Note also that the lack of importance of the uncertainty shock for real outcomes is not simply due to its size. Again, the pure uncertainty shock in the second row accounts for 35 percent of the total variance of uncertainty. So if we scale up the variance decompositions for employment and industrial production by multiplying them by 3, the point estimates still say that all uncertainty variation accounts for less than 5 percent of the total variance of employment and industrial production, with upper ends of the 90-percent confidence intervals at approximately 25 percent. And since the point estimates for the IRFs are near zero, about half of that probability mass is associated with scenarios in which uncertainty shocks are actually *expansionary*.

Looking at the third row, we see that the behavior of the rv and news shocks is again significantly different. For employment, the variance accounted for by the rv shock is larger at the 90-percent level, while for IP, they differ at only a 1-standard-deviation level.

In the end, then, figure 3 shows that pure uncertainty shocks account for a substantial fraction of the total variation in uncertainty about future stock returns, but they account for only a trivial fraction of the variation in real activity.

5.2.3 Robustness

Figures 5 and 6 report impulse responses for employment and IP to the two identified shocks and their difference across a number of perturbations of the benchmark specification, and figure 7 shows the corresponding forecast error variance decomposition.

First, we consider alternative orderings of the variables in the VAR. The effects of the ordering depend ultimately on the correlation matrix of the innovations, which we report below:

	rv	v_1	Fed Funds	Empl.	IP
rv	1				
v_1	0.73	1			
Fed Funds	0.01	-0.06	1		
Empl.	0.01	0.05	0.05	1	
IP	-0.04	0.01	0.04	0.54	1

The shocks to rv and v_1 are correlated with each other, but almost completely uncorrelated with those in the other variables, implying that if the news shock were orthogonalized not just to the shock to rv_t but to all the macro variables also, its IRF would remain unchanged.[15] The top panels of figures 5 and 6 show that when rv and v_1 are ordered last in the VAR, we obtain results that are nearly identical to our benchmark specification.

The second row of panels in figures 5 and 6 reports results when we substitute v_6 for v_1. The six-month option-implied volatility seems like potentially a more natural variable to include since it might represent a more realistic economic decision horizon. As the figure shows, in that case we find results even less favorable to uncertainty shocks, with these shocks now having slightly expansionary effects on employment and industrial production.

The goal of the main analysis is to identify a pure shock to uncertainty that has no contemporaneous impact on realized volatility. A natural question is what happens if we reverse the ordering of the identification so that the first shock is the *entire* shock to uncertainty, while the second captures the residual variation in rv. In this case, then, the first shock is a combined shock to uncertainty and rv, while the second shock is a pure shock to realized volatility that has no net effect on uncertainty on impact. Appendix figure A.6 reports results from such a specification. It shows that both these shocks have essentially the same effect on employment and industrial production, exactly as

[15]The correlation between the innovations to rv and v_1 may be worrisomely high. We also report results using v_6 in place of v_1, and then the correlation in the innovations is reduced to 0.66.

should be expected from our main analysis. Recall that in the main analysis, the identified rv and uncertainty shocks have similar effects on future volatility, but they are identified so that only the rv shock affects rv on impact. In figure A.6, on the other hand, the shocks have very *similar* effects on rv on impact, and they are now distinguished by having *different* effects on future volatility.

What figures 2 and A.6 show together is that when two shocks have the same initial effect on rv, they have the same effects on output (figure A.6) whereas when they have different effects on rv, they have different effects on output (figure 2). In other words, figures 2 and A.6 together show that it is the impact of a shock on contemporaneous realized volatility, not its effect on uncertainty, that determines how it affects output.

Figures A.9 and A.10 in the appendix report results from four additional robustness tests:

1. Detrending all variables with a one-sided HP filter
2. Limiting the sample to 1988–2006 to remove the jumps in volatility in 1987 and 2008.
3. Controlling for the S&P 500 in the VAR before the identified shocks.
4. Using RV and V_1 (i.e. in levels rather than logs).

The results of those robustness tests are qualitatively and quantitatively consistent with our baseline results.

Appendix figures A.7 and A.8 report results from a quarterly VAR similar to what is estimated in Basu and Bundick (2016) with the set of variables now rv, v_1, GDP, aggregate consumption, aggregate investment (all in real terms), the GDP deflator, and the Fed funds rate. The large number of variables and smaller number of time series observations gives this VAR specification much lower power overall than our benchmark monthly specification. As in our benchmark results, output, consumption, and investment all fall following an increase in rv. However, in this case the confidence bands around the uncertainty shock are wide enough to render the estimates essentially uninformative. When we use v_1 to help measure uncertainty, the shock appears slightly contractionary, while using v_6 makes it appear actually expansionary. We thus conclude that the quarterly data does not yield sufficient power to draw firm conclusions, but it in no way conflicts with our main results, and it reinforces the robustness of the findings for the contractionary effects of realized volatility shocks.

Finally, we also estimate a factor-augmented VAR using the first three principal components from the set of macroeconomic time series studied by Ludvigson and Ng (2007). The advantage of the FAVAR specification is that incorporates information from an extremely broad range of variables, instead of just using employment and industrial production (we use the setup of Bernanke, Boivin, and Eliasz (2005)). Figures A.11 and A.12 report results from that estimation, which are again highly similar to the benchmark.

5.3 Allowing more variables to predict second moments

The benchmark VAR has two key restrictions that this subsection relaxes. First, it assumes that the vector $[rv_t, v_{1,t}]'$ is driven only by its own lags (that is, the coefficients in those rows of the VAR on the other variables are set to zero). Second, it uses information only from a single point on the term structure of implied volatilities. Those restrictions are both consistent with the data in terms of the regressions in table 1 and a χ^2 test in the VAR, but it is reasonable to ask what happens when they are relaxed. This section first uses lasso as an alternative method to shrink coefficients towards zero and reduce overfitting, and second allows v_6 to enter the VAR and removes all the coefficient restrictions. In that case, the uncertainty series appears to be substantially overfit, emphasizing why it is desirable to discipline how the news shock is constructed.

5.3.1 Lasso

Lasso (Tibshirani (1996)) is a regularization method for regressions. Instead of choosing the set of coefficients in the VAR to minimize the residual variance, the objective function in lasso also includes a penalty on the sum of the absolute values of the coefficients. Because the penalization function is not differentiable at zero (with a positive slope on both sides of zero), it implies that coefficients that are sufficiently small are optimally set to zero. The advantage of using lasso for our purposes is that it is economically agnostic and driven by statistical considerations. It thus yields restrictions on the VAR to help reduce overfitting similarly to our benchmark specification, but without imposing the set of restrictions based purely on theory.

The appendix describes the details of our implementation of lasso. We choose the magnitude of the penalty on the coefficients based on a cross-validation criterion. In our implementation, lasso restricts many of the coefficients on the macro variables in the equations for rv and v_1 to zero, but not all of them. It also restricts the coefficients on some of the lags of rv and v_1 to be zero (they were not restricted in the benchmark). So it reduces the restrictions on the macro variables but increases them on the lags of the volatility measures.

The third row in figures 5, 6 and 7 reports results when we use lasso for the estimation instead of the benchmark restrictions. We see that the results remain highly similar to the benchmark case, showing that the restrictions applied in the benchmark model do not drive the results alone.

5.3.2 Completely unrestricted model

Finally, the bottom rows in figures 5, 6 and 7 report results using the completely unrestricted identification scheme from BBL for the news shock and include $v_{6,t}$ as an additional predictor of volatility. These are thus our most general and least constrained results. Shocks to realized volatility continue to be contractionary, and we again find no evidence that the news shock has negative effects on output, but the confidence bands for the news shock in this case become so wide as to be uninformative. The difference between the IRFs for rv and news is now statistically insignificant. We obtain similar results when we include other predictors of future volatility instead of $v_{6,t}$ such as the default spread.

The unrestricted specification has low power for identifying the effect of news shocks because the news shock itself is difficult to identify. There are, in this case, 24 potential predictors, and noise in the estimated coefficients on them is inherited by the identified news shock. That is why the specifications with restrictions and using lasso, which both have fewer coefficients to estimate, are much more stable.

Beyond power, though, the unrestricted model is potentially problematic since the news shock may be overfit. That can be seen partly from the fact that the news shock in this case actually accounts for a substantially *larger* amount of the variance of uncertainty than the rv shock does – up to 60 percent of the total variance at short horizons.

As further evidence of overfitting, the bottom panel of figure 4 plots total uncertainty, $E_t\left[\sum_{j=1}^{24} rv_{t+j}\right]$, from the benchmark and completely unrestricted specifications. We see that in the unrestricted specification, uncertainty varies much more – its standard deviation is higher by a factor of 1.6. Moreover, it rises well in advance of the 2008 financial crisis, even though realized and option-implied volatility were low. That is, the unrestricted model seems to have seen the crisis coming before investors, consistent with overfitting. Uncertainty in the restricted model leads the benchmark specifications in a number of other episodes. This lead-lag relationship can also be confirmed by examining cross-correlations. Finally, whereas uncertainty in the restricted model is positively correlated with alternative measures of uncertainty, the Baker–Bloom–Davis index and the Michigan index, uncertainty in the unrestricted model is actually *negatively* correlated with those series (with coefficients of -0.25 and -0.30), again emphasizing the implausibility of this specification.

In the end, then, when we use the most general specification and add no information to the model to help control the predictive coefficients, we continue to find no evidence that uncertainty shocks are contractionary, but we no longer have power to statistically distinguish their effects from those of rv shocks. The most conservative interpretation of our empirical analysis is therefore simply that *realized* volatility is followed by contractions and that it is critical to separate realized volatility from uncertainty when trying to estimate the effects of uncertainty shocks. In the specifications that we view as most reasonable, though, which add information to the VAR either through economic priors or through statistical regularization, we can go a step further and argue that there is in fact some affirmative evidence that uncertainty shocks have effects on the economy that are quantitatively close to zero, both in terms of impulse responses and forecast error variance decompositions.

6 Evidence from risk premia

The results above use data from stock options to provide information about second-moment expectations. If risk premia were zero or constant, option-implied volatility would be an incredibly valuable data source since it would give a direct measure of volatility expectations. That is not the case, though, which is why we are forced to use the news shock identification scheme. So in the VARs, risk premia are a contaminant that make identification more difficult.

But risk premia actually contain information that can be useful to complement and support the previous macroeconomic analysis. Risk premia reveal how much investors are willing to pay to hedge against certain risks. By looking at risk premia of volatility-related securities (variance swaps and portfolios of options) we now show that investors have historically paid large premia for insurance against increases in realized volatility, but not for insurance against increases in market-implied uncertainty. That suggests that investors do not view periods in which uncertainty rises as having high marginal utility (i.e. as being bad times), consistent with our VAR results.

A one-month variance swap is an asset whose final payoff is the sum of daily squared log returns of the underlying index (the S&P 500, in our case) over the next month. That asset gives the buyer protection against a surprise in equity return volatility (rv) over the next month. If investors are averse to periods of high realized volatility, then, we would expect to see negative average returns on one-month variance swaps, reflecting the cost of buying that insurance. A simple way to see that is to note that in general the Sharpe ratio of an asset, the ratio of its expected excess return to its standard deviation, is

$$\frac{E_t\left[R_{t+1}-R_{f,t+1}\right]}{SD_t\left[R_{t+1}\right]}=-corr_t\left(R_{t+1},MU_{t+1}\right)\times std\left(MU_{t+1}\right) \tag{15}$$

for any return R_{t+1}, where $R_{f,t+1}$ is the risk-free rate and MU_{t+1} denotes the marginal utility of consumption on date $t+1$. Assets that covary positively with marginal utility, and hence are hedges, earn negative average returns. So if realized volatility is high in high marginal utility states (in most models, bad times), then one-month variance swaps will earn high Sharpe ratios.

The first point on the left in the left-hand panel of Figure 8 plots average annualized Sharpe ratios on 1-month S&P 500 variance swaps between 1996 and 2014.[16] The average Sharpe ratio is -1.4, approximately three times larger (with the opposite sign) than the Sharpe ratio on the aggregate equity market in that period. In other words, investors have been willing to pay extraordinarily large premia for protection against periods of high realized volatility, suggesting that they view those times as particularly bad (or as having very high marginal utility).

Now consider a j-month variance forward, whose payoff, instead of being the sum of squared returns over the next month ($t+1$), is the sum of squared returns in month $t+j$ (so then the one-month variance swap above can also be called a 1-month variance forward). If an investor buys a j-month variance forward and holds it for a single month, selling it in month $t+1$, then the variance forward protects her over that period against *news* about volatility in month $t+j$. If between t and $t+1$ investors receive news that volatility will be higher in the future – i.e. if uncertainty rises – the holding period return on that j-period forward will increase. The left-hand panel of figure 8 also plots one-month holding period Sharpe ratios for variance forwards with maturities from 2 to 12 months. We see that for all maturities higher than 2 months, the Sharpe ratios are near zero, and in fact the sample point estimates are positive. The Sharpe ratios are also all statistically significantly closer to zero than the Sharpe ratio on the one-month variance swap.

The left panel of Figure 8 therefore shows that there is something special about the surprise in realized volatility compared to news about volatility going forward (uncertainty). Investors

[16] The data is described in Dew-Becker et al. (2017); it is obtained from a large asset manager and Markit, but may be closely approximated by portfolios of options, for which prices are widely available (e.g. from Optionmetrics).

have paid large premia for protection against surprises in realized volatility, but news about future uncertainty has had a premium that is indistinguishable from zero, and may even be positive. Realized volatility thus appears to have a large positive correlation with marginal utility, while shocks to expected volatility have a correlation that is close to zero or negative.

Using the options data described above, it is possible to extend those results further, back to 1983. The right-hand panel of figure 8 reports the average shape of the term structure of variance forward prices constructed from data on S&P 500 options (we study the term structure with this data because it is estimated more accurately than returns). The variance forwards are constructed from synthetic variance swaps, a calculation almost identical to our calculation of $Var_t^Q[s_{t+n}]$. The term structure reported here is directly informative about risk premia. The average return on an n-month variance claim is:

$$\underbrace{E\left[\frac{F_{n-1,t}-F_{n,t-1}}{F_{n,t-1}}\right]}_{\text{Average return of the forward}} \approx \frac{\overbrace{E\left[F_{n-1}\right]-E\left[F_n\right]}^{\text{Average slope of the forward curve}}}{E\left[F_n\right]} \tag{16}$$

where $F_{n,t}$ is the price on date t of an n-maturity volatility forward. The slope of the average term structure thus indicates the average risk premium on news about volatility n months forward. If the term structure is upward sloping, then the prices of the variance claims fall on average as their maturities approach, indicating that they have negative average returns. If it slopes down, then average returns are positive.

The right-hand panel of Figure 8 plots the average term structure of variance forward prices for the period 1983–2013. The term structure is strongly upward sloping for the first two months, again indicating that investors have paid large premia for assets that are exposed to realized variance and expected variance one month in the future. But the curve quickly flattens, indicating that the risk premia for exposure to fluctuations in expected variance further in the future have been much smaller.

The asset return data says that investors appear to have been highly averse to news about high realized volatility, while shocks to expected volatility do not seem to have been related to marginal utility. The confidence intervals that we obtain are sufficiently wide that we cannot claim that shocks to expected future volatility do not earn an economically meaningfully negative risk premium. What we can say, though, is that investors seem to have cared over our sample *much more* about surprises in realized volatility than in uncertainty. Figure 8 therefore confirms and complements the results from our VAR, that show that shocks to rv are associated with recessions but uncertainty shocks are not.

7 Equilibrium model and further evidence

The paper thus far has provided empirical evidence on two basic points: first, surprises in realized volatility in the stock market are associated with future declines in real activity, while uncertainty shocks, identified as second-moment news, are not; second, investors have historically paid large premia to hedge shocks to realized volatility, but have paid premia that have averaged to nearly zero to hedge shocks to uncertainty.

In this section we present a simple stylized structural model that is consistent with those features of the data, essentially a small extension of the classic RBC model. Our goal is to develop the simplest possible model consistent with the data and the VAR. We show that they key ingredient is asymmetry in fundamental shocks. Intuitively, when output (or technology) growth is skewed to the left, large shocks, which are associated with high realized volatility, tend to be negative. That is simply the definition of left skewness: the squared innovation is negatively correlated with the level of the innovation.

We first discuss evidence that there is left skewness in economic activity and then describe the model and show that estimated impulse responses to uncertainty and realized volatility shocks in the model match what we have found in the data.

7.1 Skewness

A potential source of negative correlation between output and realized volatility is negatively skewed shocks. Specifically, if some shock ε is negatively skewed, then $E\left[\varepsilon^3\right] < 0 \Rightarrow cov\left(\varepsilon, \varepsilon^2\right) < 0$. That is, negative skewness implies a negative correlation between ε^2 and ε itself. So high realized volatility (ε^2) should be associated with downturns. The obvious question, then, is whether shocks to output and asset returns are actually skewed to the left. There are large literatures studying skewness in both aggregate stock returns and economic growth. We therefore provide just a brief overview of the literature and the basic evidence.

Table 2 reports the skewness of monthly and quarterly changes in a range of measures of economic activity. Nearly all the variables that we examine are negatively skewed, at both the monthly and quarterly levels. One major exception is monthly growth in industrial production, but that result appears to be due to some large fluctuations in the 1950s. When the sample is cut off at 1960, the results for industrial production are consistent with those for other variables.

In addition to real variables, table 2 also reports realized and option-implied skewness for S&P 500 returns.[17] The implied and realized skewness of monthly stock returns is substantially negative, and in fact surprisingly similar to the skewness of capacity utilization. The realized skewness of stock returns is less negative than option-implied skewness, which is consistent with investors demanding a risk premium on assets that have negative returns in periods when realized skewness is especially negative (i.e. that covary positively with skewness).

[17] We obtain option-implied skewness from the CBOE's time series of its SKEW index, which is defined as $SKEW = 100 - 10 \times Skew\left(R\right)$. We thus report $10 - SKEW/10$.

In addition to the basic evidence reported here, there is a large literature providing much more sophisticated analyses of asymmetries in the distributions of output and stock returns. Morley and Piger (2012) provide an extensive analysis of asymmetries in the business cycle and review the large literature. They estimate a wide range of models, including symmetrical ARMA specifications, regime-switching models, and frameworks that allow nonlinearity. The models that fit aggregate output best have explicit non-linearity and negative skewness. Even after averaging across models using a measure of posterior probability, which puts substantial weight on purely symmetrical models, Morley and Piger find that their measure of the business cycle is substantially skewed to the left, consistent with the results reported in table 2. More recently, Salgado, Guvenen, and Bloom (2016) provide evidence that left skewness is a robust feature of business cycles, at both the macro and micro levels and across many countries.

The finance literature has also long recognized that there is skewness in aggregate equity returns and in option-implied return distributions (see Campbell and Hentschel (1992), Ait-Sahalia and Lo (1998), and Bakshi, Kapadia, and Madan (2003), for recent analyses and reviews). The skewness that we measure here appears to be pervasive and has existed in returns reaching back even to the 19th century (Campbell and Hentschel (1992)).

Taken as a whole, then, across a range of data sources and estimation methods, there is a substantial body of evidence that fluctuations in the economy are negatively skewed. In a world of negative skewness, it is not surprising that measures of realized volatility are correlated with declines in activity, simply because skewness is related to the third moment: $E\left[\varepsilon^3\right] = E\left[\varepsilon \cdot \varepsilon^2\right]$.

7.2 An equilibrium model

The empirical evidence presented thus far is consistent with the view that shocks to aggregate realized volatility are associated with significant declines in macroeconomic activity while shocks to expected aggregate volatility are not. In this section we present a stylized equilibrium model that is both consistent with our evidence and close to the workhorse RBC model. We deliberately keep the model simple in order to highlight the economic channels that are at work; the model is not rich enough to provide a tight quantitative fit to the economy. But despite its simplicity, the model is qualitatively consistent with a wide variety of real and asset pricing facts, and illustrates what features a model can have to match the volatility patterns we document.

Our model is an RBC model where aggregate TFP growth is heteroskedastic and skewed to the left. We want it to be consistent with the three facts presented in this paper: (1) shocks to realized volatility are associated with declines in real activity, while shocks to expected volatility are not; (2) Sharpe ratios on short-term claims to volatility are much more negative than those on longer-term claims; and (3) output growth and equity returns are negatively skewed.

7.3 Model structure

Firms produce output with technology, A_t, capital, K_t, and labor, N_t,

$$Y_t = A_t^{1-\alpha} K_{t-1}^{\alpha} N_t^{1-\alpha} \tag{17}$$

We set $\alpha = 0.33$, consistent with capital's share of income. Capital is produced subject to adjustment costs according to the production function

$$K_t = (1-\delta) K_{t-1} + K_{t-1} \left(\frac{I_t}{K_{t-1}^{agg}} - \frac{\zeta}{2} \left(\frac{I_t}{K_{t-1}^{agg}} - \overline{I/K} \right)^2 \right) \tag{18}$$

where I_t is gross investment, K_t^{agg} is the aggregate capital stock (which is external to individual firm decisions), ζ is a parameter determining the magnitude of adjustment costs and $\overline{I/K}$ is the steady-state investment/capital ratio. We set $\delta = 0.08/12$ (corresponding to a monthly calibration) and $\zeta = 0.5$.[18] Given the structure for adjustment costs and production, the equilibrium price and return on a unit of installed capital are

$$P_{K,t} = \frac{1}{1 - \zeta \left(\frac{I_t / K_{t-1}}{\overline{I/K}} - 1 \right)} \tag{19}$$

$$R_{K,t} = \frac{\alpha A_t^{1-\alpha} K_{t-1}^{\alpha-1} N_t^{1-\alpha} + (1-\delta) P_{K,t}}{P_{K,t-1}} \tag{20}$$

We assume there is a representative agent who maximizes Epstein–Zin (1991) preferences over consumption and leisure,

$$v_t = \arg\max_{C,N} \log \left(C_{t+j} - b C_{t+j-1}^{agg} \right) - \theta \frac{N_{t+j}^{1+\chi}}{1+\chi} + \beta \log E_t \exp\left(-\gamma v_{t+1}\right) \tag{21}$$

where C^{agg} is aggregate consumption, subject to the budget constraint

$$C_t + I_t \leq Y_t \tag{22}$$

Agents have log utility over consumption minus an external habit. We set the magnitude of the habit to $b = 0.8$ to help generate smoothness in consumption. β is set to $0.99^{1/12}$, χ to 1/3 for a Frisch elasticity of 3, and θ to generate steady-state employment of 1/3. The coefficient of relative risk aversion, γ, is set to 6 to generate a Sharpe ratio on a claim to capital of 0.35, similar to what is observed for US equities. Note, though, that the habit will induce variation in effective risk aversion over time. Epstein–Zin preferences are a simple way to generate realistically large risk premia.

[18] See, e.g., Cummins, Hassett, and Hubbard (1994) for estimates of adjustment costs similar to this value. Jermann (1998) use a similar values. $\zeta = 0.5$ is on the lower end of estimates based on aggregate data and more consistent with micro evidence, but our results are not sensitive to the choice of this parameter.

The model is closed by the Euler equation and the optimization condition for labor,

$$1 = E_t\left[\beta\frac{\exp\left(\left(1-\alpha\right)v_{t+1}\right)}{E_t\exp\left(\left(1-\alpha\right)v_{t+1}\right)}\frac{C_t-bC_{t-1}^{agg}}{C_{t+1}-bC_t^{agg}}R_{K,t+1}\right] \tag{23}$$

$$\frac{\theta N_t^{\gamma}}{C_t-bC_{t-1}} = \left(1-\alpha\right)A_t^{-\alpha}K_{t-1}^{\alpha}N_t^{-\alpha} \tag{24}$$

We model realized volatility as in the empirical analysis as the squared excess return on capital,

$$RV_t = \left(R_{K,t+1}-R_{f,t}\right)^2 \tag{25}$$

where $R_{f,t}$ is the risk-free rate. It is then straightforward to construct prices on claims to future realized volatility. The price of a claim to RV_{t+j} on date t is denoted $P_{V,j,t}$.

The only exogenous variable in the model is technology, A_t, which follows the process

$$\Delta\log A_t = \sigma_{t-1}\bar{\sigma}_{\varepsilon}\varepsilon_t - J\left(\nu_t-\bar{p}\right)+\mu \tag{26}$$

$$\log\sigma_t = \phi_{\sigma}\log\sigma_{t-1}+\sigma_{\sigma}\eta_t+\kappa_{\sigma,A}\Delta\log A_t \tag{27}$$

$$\varepsilon_t,\eta_t \sim N\left(0,1\right) \tag{28}$$

$$\nu_t \sim Bernoulli\left(\bar{p}\right) \tag{29}$$

Technology follows a random walk in logs with drift μ, set to 2 percent per year. ε_t is a normally distributed innovation that affects technology in each period, while ν_t is a shock that is equal to zero in most periods but equal to 1 with probability $\bar{p}$ – that is, it induces downward jumps in technology, with J determining the size of the jump and $\bar{p}$ the average frequency. σ_t determines the volatility of normally distributed shocks to technology. It is itself driven by two shocks: an independent shock η_t (volatility news) and also the innovations to technology in period t. A positive technology shock may feed into lower volatility in the future, yielding both countercyclical volatility and ARCH-type effects (Engle (1982)). The volatility process thus has two features that will be important in matching the data: it has news shocks, and it is countercyclical for $\kappa_{\sigma,A}<0$.

ϕ_{σ} and σ_{σ} are calibrated so that $\log\sigma_t$ has a standard deviation of 0.35 and a one-month autocorrelation of 0.91, consistent with the behavior of the VIX. $\kappa_{\sigma,A}$ is set to -4.37, which implies that a jump in technology, $J\nu_t$, increases σ_t by 1.5 standard deviations, generating countercyclical volatility. $\bar{\sigma}_{\varepsilon}$ is set so that normally distributed shocks on average generate a standard deviation of output growth close to the value of 1.92 we observe empirically. Jumps on average reduce technology by 8 percent (which is 3.2 times $\bar{\sigma}_{\varepsilon}$, the average standard deviation of the Gaussian TFP shocks) and are calibrated to occur once every 10 years on average. We thus think of them as representing small disasters or relatively large recessions (consistent Backus, Chernov, and Martin (2011) and with the view of skewed recessions in Salgado, Guvenen, and Bloom (2016)), rather than depression-type disasters.[19] While the size of the jump seems large initially, recall that the

[19] A realistic extension of the model would be to allow for jumps to be drawn from a distribution, rather than all

model is calibrated to match the standard deviation of output growth.

We solve the model by projecting the decision rule for consumption on a set of Chebyshev polynomials up to the 6th order (a so-called global solution) to ensure accuracy not only for real dynamics but also for asset prices and realized volatility. Integrals are calculated using Gaussian quadrature with 20 points. Euler equation errors are less than $10^{-5.0}$ across the range of the state space that the simulation explores and have an average absolute value of $10^{-5.3}$. The use of a global solution method allows for high accuracy in the solution, but also makes it infeasible to search over many parameters or estimate the model, which is why we explore just a single calibration here.

7.4 Simulation results

We examine three sets of implications of the model: VAR estimates, risk premia, and skewness. All results are population statistics calculated from a simulation lasting 10,000 years.

Table 3 reports basic moments of returns on capital and growth rates of output, consumption, and investment. The model generates negative skewness in all four variables in the table, consistent with the data, but the skewness is much larger than is observed empirically. Mean growth rates of real variables are similar to the data, though mean stock returns are much smaller since we do not assume leverage. In terms of the Sharpe ratio – the mean excess return divided by the standard deviation – the model matches the data. The standard deviation of output is almost identical to the data, while consumption is less and investment more volatile; the gap between the two is smaller than observed empirically, however. Table 3 thus suggests that the model generates moments that are broadly consistent with the data, in particular generating comovement among aggregate variables (the three growth rate series in table 3 have correlations between 0.53 and 0.88) and volatilities that are empirically reasonable.

Figure 9 plots the Sharpe ratios of volatility claims in the model that correspond to the forward volatility claims examined in Figure 8. As in Figure 8 (left panel), the Sharpe ratio of the one-month asset, which is a claim to realized volatility, is far more negative than the Sharpe ratios for the claims with longer maturities. Intuitively, this is because shocks to volatility expectations, η_t, have relatively small effects on consumption, hence earning a small risk premium. Shocks to realized volatility, on the other hand, tend to isolate the jumps, ν_t, (as we will show below), so they earn larger premia. The magnitudes are somewhat smaller than observed in the data, though, which is a common problem in matching the prices of assets that pay out in extremely bad states of the world (e.g. Schreindorfer (2014)).

Finally, the solid lines in figure 10 summarize the results of VARs estimated from simulations of the model. The VAR in the simulations replicates the one used in the main analysis above. In particular, it includes realized volatility, option-implied expected volatility ($P_{V,1,t}$), and the level of output, using the same news shock identification as above.

We see, as in the main results, that a unit standard deviation shock to realized volatility has a

having the same size. See, e.g., Barro and Jin (2011).

highly transitory effect on realized volatility and a negative effect on output of a similar magnitude to what is observed empirically. The second-moment news shock has a predictive effect on future realized volatility, but only a quantitatively trivial effect on output. The bottom row shows the difference in the IRFs, and we see that the RV shock has substantially more negative effects on output than the uncertainty shock.

The VAR results are notable because they replicate the results observed empirically even though there is no structural "realized volatility shock" in the model. Rather, the identified RV comes from the jump in TFP in the model ($J\nu_t$). To see that, we report the correlations between the VAR-identified shocks and the structural shocks in the model in the bottom section of table 3.

The RV shock is correlated nearly exclusively with $J\nu_t$, the jump shock in the model. So the VAR successfully identifies the jumps as realized volatility shocks, which are then structurally, but obviously not causally, related to declines in output. The identified uncertainty shock, as we would hope, is, similarly, almost purely correlated with η_t, the volatility news shock (even though option-implied volatility in the model is not an unbiased predictor of future volatility, since there is a volatility risk premium that varies over time, as in the data). Finally, the third shock from the VAR – which is simply a residual unexplained by the RV and uncertainty shocks – is primarily correlated with ε_t, the small shock to technology. So our main VAR specification does a good job in this setting – a non-linear production model – of actually identifying true structural shocks and also fitting the qualitative behavior of our empirical VAR analysis.

The fact that the shocks identified by the VAR are very similar to the structural shocks in the model suggests that the impulse responses estimated by the VAR should be very similar to the effects of the shocks in the structural model itself. Figure 10 therefore plots, in addition to the IRFs estimated from the simulation, the responses of realized volatility, expected volatility, and output, to shocks to $J\nu_t$ and η_t – the structural jump shock and volatility news shock, respectively. The structural IRFs are the dashed lines.

We see that the response of realized and expected volatility to the VAR-estimated RV and uncertainty shocks are nearly identical to the responses to the true structural shocks $J\nu_t$ and η_t (recall that all shocks are scaled to have the same effect on expected future volatility, which is why they match closely in the middle panel on the left side of the figure). Most importantly, the response of output to the estimated shocks is rather similar to the response to the structural shocks. Output falls by 0.4 percent following the estimated RV shock, while it falls by 0.6 percent following the structural shock. The second row shows that there is essentially zero response to both the estimated uncertainty shock and to η_t, even though both do increase uncertainty and future realized volatility.

This section thus shows that a simple production model can match the basic features of the data that we have estimated in this paper: output responds negatively to shocks to realized volatility but not to shocks to uncertainty, there is a much larger risk premium for realized than expected volatility, and economic activity and stock returns are both skewed to the left.

8 Conclusion

The key distinction that this paper draws is between realized volatility and uncertainty. Volatility matters for output, but it is the realized part that is robustly followed by downturns. Changes in expected volatility – uncertainty shocks – appear to have no significant negative effects. Evidence from asset prices and risk premia is consistent with these findings, and we develop a simple model that can rationalize the data and also justifies our identification scheme.

The evidence we present favors the view that bad times are volatile times, not that uncertainty causes bad times. A leading hypothesized explanation for the slow recovery from the 2008 financial crisis has been that uncertainty about the aggregate economy (e.g. due to policy uncertainty) since then has been high. Our evidence suggests that aggregate uncertainty may not have been the driving force, and that economists should search elsewhere for an explanation to the slow recovery puzzle.

More constructively, this paper aims to lay out a specific view of the joint behavior of stock market volatility and the real economy. There appear to be negative shocks to the stock market that occur at business cycle frequencies, are associated with high realized volatility and declines in output, and are priced strongly by investors. The simple idea that fundamentals are skewed left can explain our VAR evidence, the pricing of volatility risk, and the negative unconditional correlation between economic activity and volatility.

References

Adrian, Tobias and Joshua Rosenberg, "Stock Returns and Volatility: Pricing the Short-Run and Long-run Components of Market Risk," *The Journal of Finance*, 2008, *63(6)*, 2997–3030.

____ **, Nina Boyarchenko, and Domenico Giannone**, "Vulnerable growth," 2016.

Ait-Sahalia, Yacine and Andrew W Lo, "Nonparametric estimation of state-price densities implicit in financial asset prices," *The Journal of Finance*, 1998, *53* (2), 499–547.

____ **, Mustafa Karaman, and Loriano Mancini**, "The Term Structure of Variance Swaps, Risk Premia and the Expectations Hypothesis," 2015. Working paper.

Alexopoulos, Michelle and Jon Cohen, "Uncertain Times, Uncertain Measures," 2009. Working paper.

Andersen, Torben G. and Oleg Bondarenko, "Construction and Interpretation of Model-Free Implied Volatility," Working Paper, National Bureau of Economic Research September 2007.

Andersen, Torben G, Tim Bollerslev, and Francis X Diebold, "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility," *The Review of Economics and Statistics*, 2007, *89* (4), 701–720.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, "The Cross-Section of Volatility and Expected Returns," *The Journal of Finance*, 2006, *61(1)*, 259–299.

Bachmann, Rudiger and Christian Bayer, ""Wait-and-See" business cycles?," *Journal of Monetary Economics*, 2013, *60* (6), 704–719.

___ **and Giuseppe Moscarini**, "Business Cycles and Endogenous Uncertainty," 2012. Working paper.

Bachmann, Rüdiger, Steffen Elstner, and Eric R Sims, "Uncertainty and economic activity: Evidence from business survey data," *American Economic Journal: Macroeconomics*, 2013, *5* (2), 217–249.

Baker, Scott and Nicholas Bloom, "The Effects of Uncertainty Shocks on International Economic Growth," 2013. Working paper.

Bakshi, Gurdip and Nikunj Kapadia, "Delta-Hedge Gains and the Negative Market Volatility Risk Premium," *The Review of Financial Studies*, 2003, *16(2)*, 527–566.

___**,** ___**, and Dilip Madan**, "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *Review of Financial Studies*, 2003, *16* (1), 101–143.

Barro, Robert J and Tao Jin, "On the size distribution of macroeconomic disasters," *Econometrica*, 2011, *79* (5), 1567–1589.

Barsky, Robert B and Eric R Sims, "News shocks and business cycles," *Journal of Monetary Economics*, 2011, *58* (3), 273–289.

___**, Susanto Basu, and Keyoung Lee**, "Whither News Shocks?," Technical Report, National Bureau of Economic Research 2014.

Beaudry, Paul and Franck Portier, "Stock Prices, News, and Economic Fluctuations," *The American Economic Review*, 2006, *96* (4), 1293–1307.

___ **and** ___, "News-driven business cycles: Insights and challenges," *Journal of Economic Literature*, 2014, *52* (4), 993–1074.

Berger, David and Joseph Vavra, "Volatility and Pass-through," *NBER Working Paper Serie No. 19651*, 2013.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz, "Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach.," *Quarterly Journal of Economics*, 2005, *120* (1), 387–422.

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

——, **Scott Baker, and Steven J. Davis**, "Measuring Economic Policy Uncertainty," 2015. Working paper.

Bollerslev, Tim, George Tauchen, and Hao Zhou, "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 2009, *22(11)*, 4463–4492.

Born, Benjamin and Johannes Pfeifer, "Policy Risk and the Business Cycle," *Journal of Monetary Economics*, 2014, *68*, 65–85.

Broadie, Mark and Jerome Detemple, "American option valuation: new bounds, approximations, and a comparison of existing methods," *Review of Financial Studies*, 1996, *9* (4), 1211–1250.

——, **Mikhail Chernov, and Michael Johannes**, "Model Specification and Risk Premia: Evidence from Futures Options," *The Journal of Finance*, 2007, *62(3)*, 1453–1490.

Caldara, Dario, Cristina Fuentes-Albero, Simon Gilchrist, and Egon Zakrajšek, "The macroeconomic impact of financial and uncertainty shocks," *European Economic Review*, 2016.

Campbell, John Y. and Ludger Hentschel, "No news is good news: An asymmetric model of changing volatility in stock returns," *Journal of Financial Economics*, 1992, *31* (3), 281–318.

——, **Martin Lettau, Burton G. Malkiel, and Yexiao Xu**, "Have Individual Stocks Become More Volatile? an Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, 2001, *56* (1), 1–43.

Carr, Peter and Liuren Wu, "Variance Risk Premiums," *Review of Financial Studies*, 2009, *22(3)*, 1311–1341.

Cummins, Jason G., Kevin A. Hassett, and R. Glenn Hubbard, "A Reconsideration of Investment Behavior Using Tax Reforms as Natural Experiments," *Brookings Papers on Economic Activity*, 1994, *1994* (2), 1–74.

Decker, Ryan, Pablo N. D'Erasmo, and Hernan Moscoso Boedo, "Market exposure and endogenous firm volatility over the business cycle," *American Economic Journal: Macroeconomics*, 2016, *8* (1), 148–198.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, "The Price of Variance Risk," 2016. Working paper.

Drechsler, Itamar and Amir Yaron, "What's Vol Got to Do with it?," *The Review of Financial Studies*, 2011, *24(1)*, 1–45.

Duffee, Gregory R., "Information in (and not in) the Term Structure," *Review of Financial Studies*, 2011, *24(9)*, 2895–2934.

Egloff, Daniel, Markus Leippold, and Liuren Wu, "The Term Structure of Variance Swap Rates and Optimal Variance Swap Investments," *Journal of Financial and Quantitative Analysis*, 2010, *45(5)*, 1279–1310.

Eisfeldt, Andrea L. and Adriano A. Rampini, "Capital Reallocation and Liquidity," *Journal of Monetary Economics*, 2006, *53* (3), 369–399.

Engle, Robert F, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the Econometric Society*, 1982, pp. 987–1007.

Epstein, Larry G. and Stanley E. Zin, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *The Journal of Political Economy*, 1991, *99(2)*, 263–286.

Fernandez-Villaverde, Jesus, Pablo Guerron, Juan F. Rubio-Ramirez, and Martin Uribe, "Risk Matters: The Real Effects of Volatility Shocks," *American Economic Review*, 2011, *101(6)*, 2530–2561.

___ **, Pablo Guerron-Quintana, Keith Kuester, and Juan Rubio-Ramirez**, "Fiscal Volatility Shocks and Economic Activity," 2013. Working paper.

Gatheral, Jim and Antoine Jacquier, "Convergence of Heston to SVI," *Quantitative Finance*, 2011, *11* (8), 1129–1132.

___ **and** ___ , "Arbitrage-free SVI volatility surfaces," *Quantitative Finance*, 2014, *14* (1), 59–71.

Giglio, Stefano, Bryan Kelly, and Seth Pruitt, "Systemic risk and the macroeconomy: An empirical evaluation," *Journal of Financial Economics*, 2016, *119* (3), 457–471.

Guvenen, Fatih, Serdar Ozkan, and Jae Song, "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 2014, *122* (3), 621–660.

Hamilton, James Douglas, *Time series analysis*, Vol. 2, Princeton university press Princeton, 1994.

Homescu, Cristian, "Implied Volatility Surface: Construction Methodologies and Characteristics," 2011. Working paper.

Ilut, Cosmin, Matthias Kehrig, and Martin Schneider, "Slow to hire, quick to fire: Employment dynamics with asymmetric responses to news," Technical Report, National Bureau of Economic Research 2014.

___ **,** ___ **, and** ___ , "Slow to Hire, Quick to Fire: Employment Dynamics with Asymmetric Responses to News," *NBER Working Paper Series*, 2015.

Jiang, George J. and Yisong S. Tian, "Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index," *The Journal of Derivatives*, 2007, *14* (3), 35–60.

Jordà, Òscar, "Estimation and inference of impulse responses by local projections," *American economic review*, 2005, pp. 161–182.

Jurado, Kyle, Sydney Ludvigson, and Serena Ng, "Measuring Uncertainty," *American Economic Review*, 2015, *105* (3), 1177–1216.

Kelly, Bryan, Lúboš Pástor, and Pietro Veronesi, "The price of political uncertainty: Theory and evidence from the option market," *The Journal of Finance*, 2016.

Kocherlakota, Narayana, "Creating business cycles through credit constraints," *Federal Reserve Bank of Minneapolis Quarterly Review*, 2000, *24* (3), 2–10.

Kozlowski, Julian, Laura Veldkamp, and Venky Venkateswaran, "The Tail that Wags the Economy: Belief-Driven Business Cycles and Persistent Stagnation," 2016. Working paper.

Leduc, Sylvain and Zheng Liu, "Uncertainty Shocks are Aggregate Demand Shocks," 2015. Working paper.

Lee, Roger W., "The Moment Formula for Implied Volatility at Extreme Strikes," *Mathematical Finance*, 2004, *14* (3), 469–480.

Ludvigson, Sydney C. and Serena Ng, "The Empirical Risk-Return Relation: A Factor Analysis Approach," *Journal of Financial Economics*, 2007, *83(1)*, 171–222.

Ludvigson, Sydney C, Sai Ma, and Serena Ng, "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?," Technical Report, National Bureau of Economic Research 2015.

Morley, James and Jeremy Piger, "The asymmetric business cycle," *Review of Economics and Statistics*, 2012, *94* (1), 208–221.

Salgado, Sergio, Fatih Guvenen, and Nicholas Bloom, "Skewed Business Cycles." Working paper.

Schreindorfer, David, "Tails, fears, and equilibrium option prices," 2014.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron, "Cyclical Dynamics in Idiosyncratic Labor Market Risk," *Journal of Political Economy*, 2004, *112* (3), 695–717.

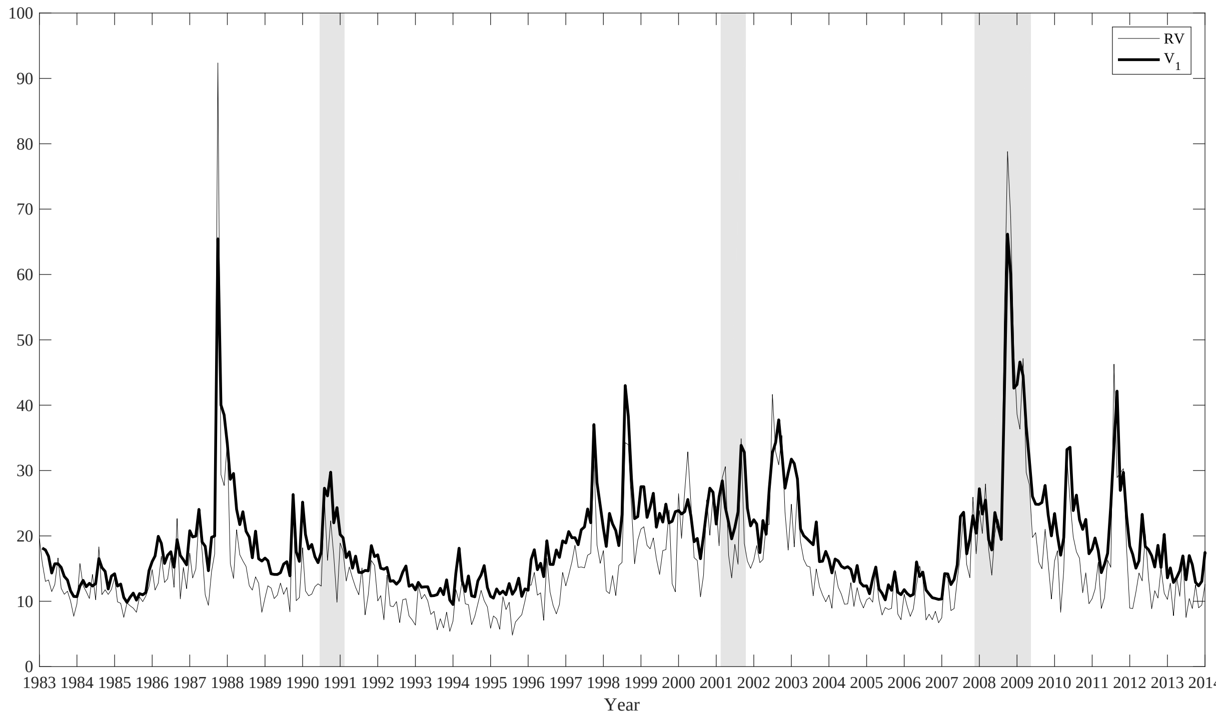
___ **, Christopher I. Telmer, and Amir Yaron**, "Asset pricing with idiosyncratic risk and overlapping generations," *Review of Economic Dynamics*, 2007, *10* (4), 519 – 548.

Systems, Zeliade, "Quasi-Explicit Calibration of Gatheral's SVI model," Technical Report, Zeliade Systems 2009.

Taylor, Stephen J., Pradeep K. Yadav, and Yuanyuan Zhang, "The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks," *Journal of Banking & Finance*, 2010, *34* (4), 871–881.

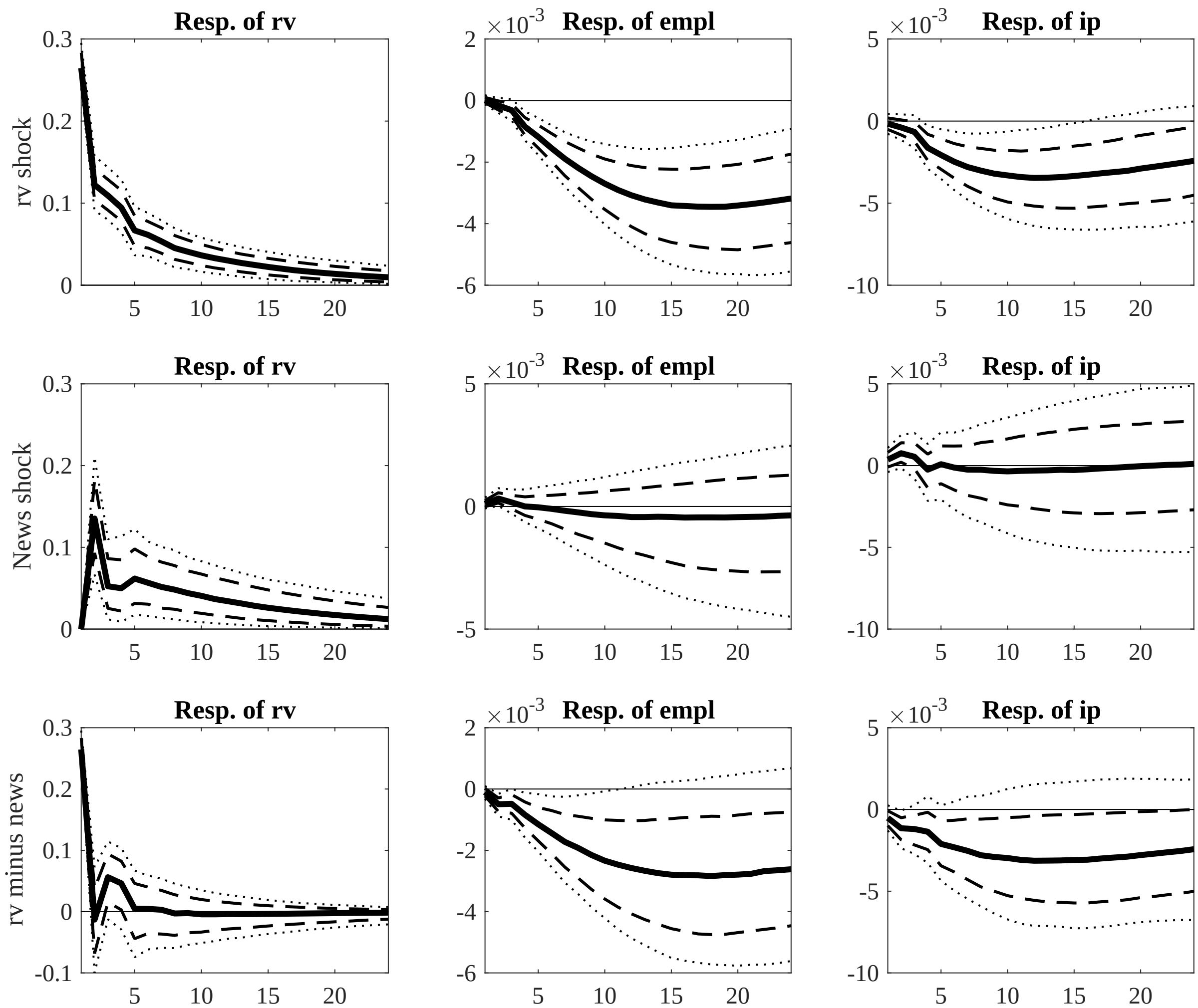
Tibshirani, Robert, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1996, pp. 267–288.

Figure 1: Time series of realized volatility and expectations



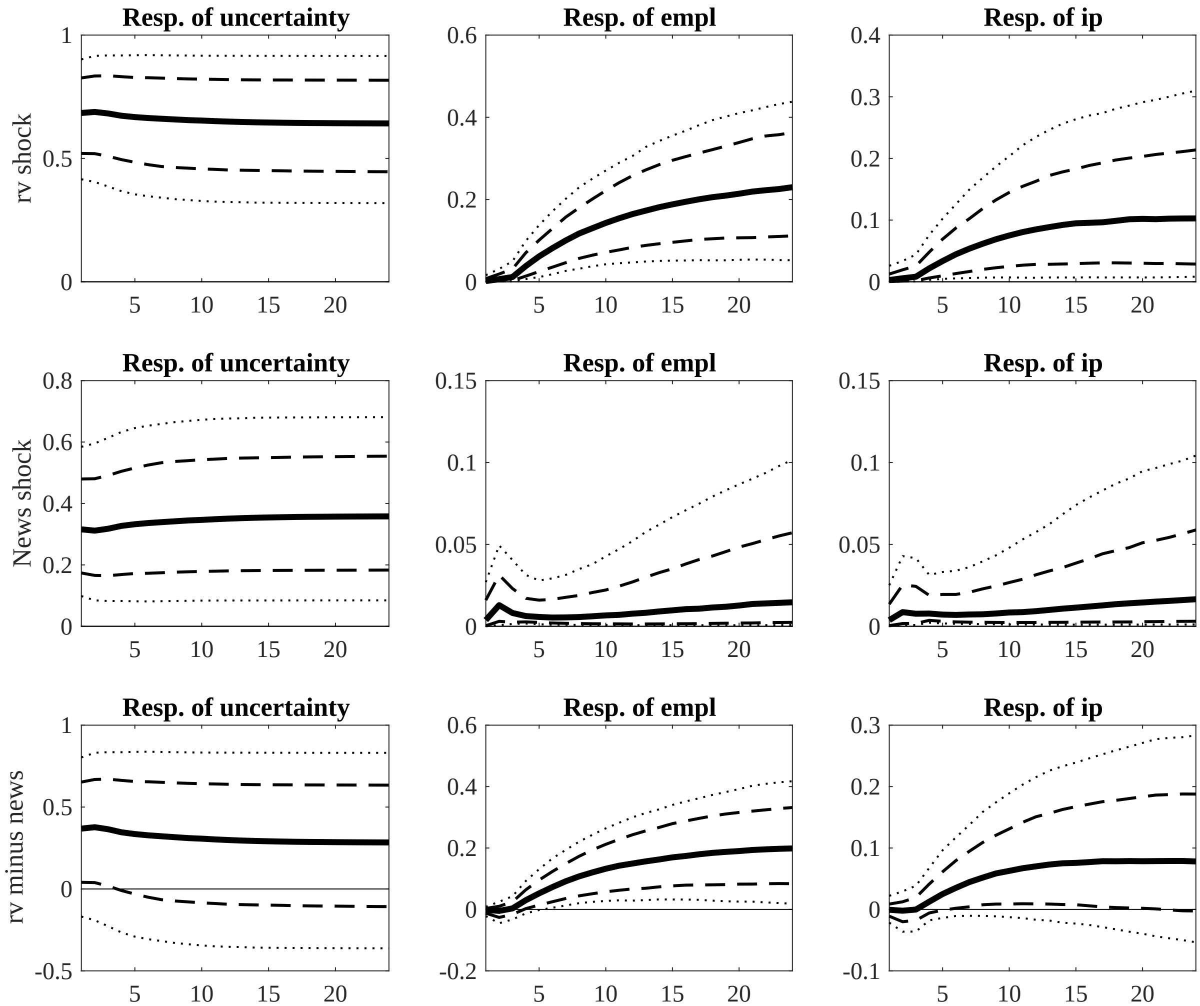
Note: Time series of realized volatility (RV), and 1-month option-implied volatility (V_1), in annualized standard deviation units. Grey bars indicate NBER recessions.

Figure 2: Impulse response functions from benchmark VAR



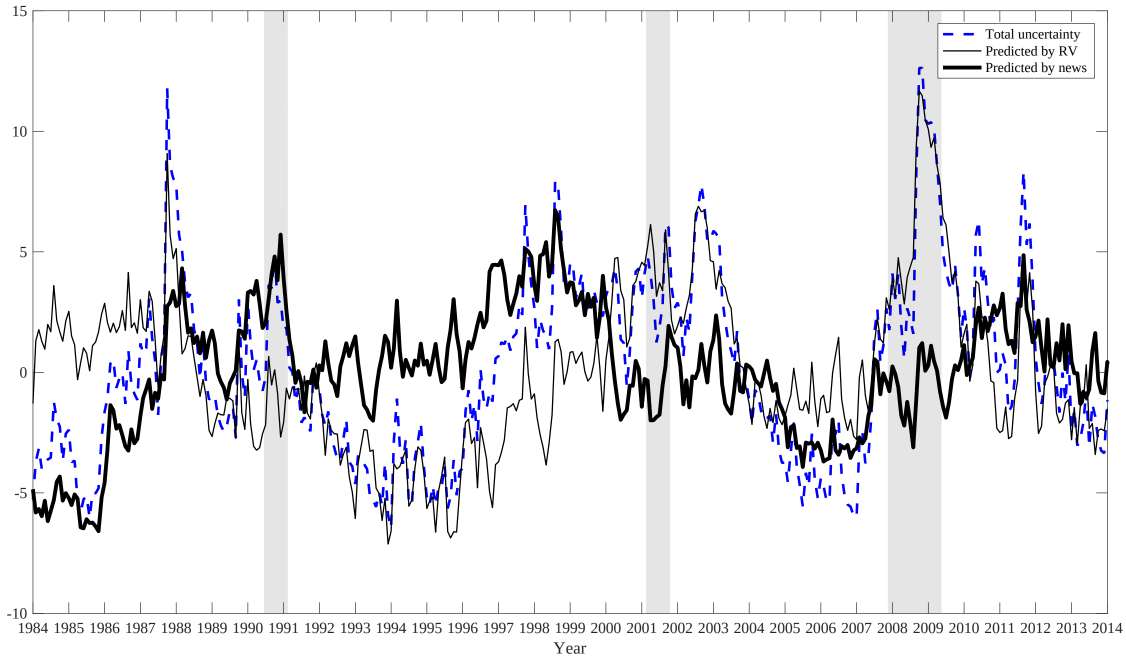
Note: Responses of rv, employment, and industrial production to shocks to rv and the identified uncertainty shock, in a VAR with rv, v_1, federal funds rate, log employment, and log industrial production. The IRFs are scaled so that the two shocks have equal cumulative effects on rv over months 2–24 following the shock. The sample period is 1983–2014. The dotted lines are 68% and 90% confidence intervals.

Figure 3: Forecast error variance decomposition

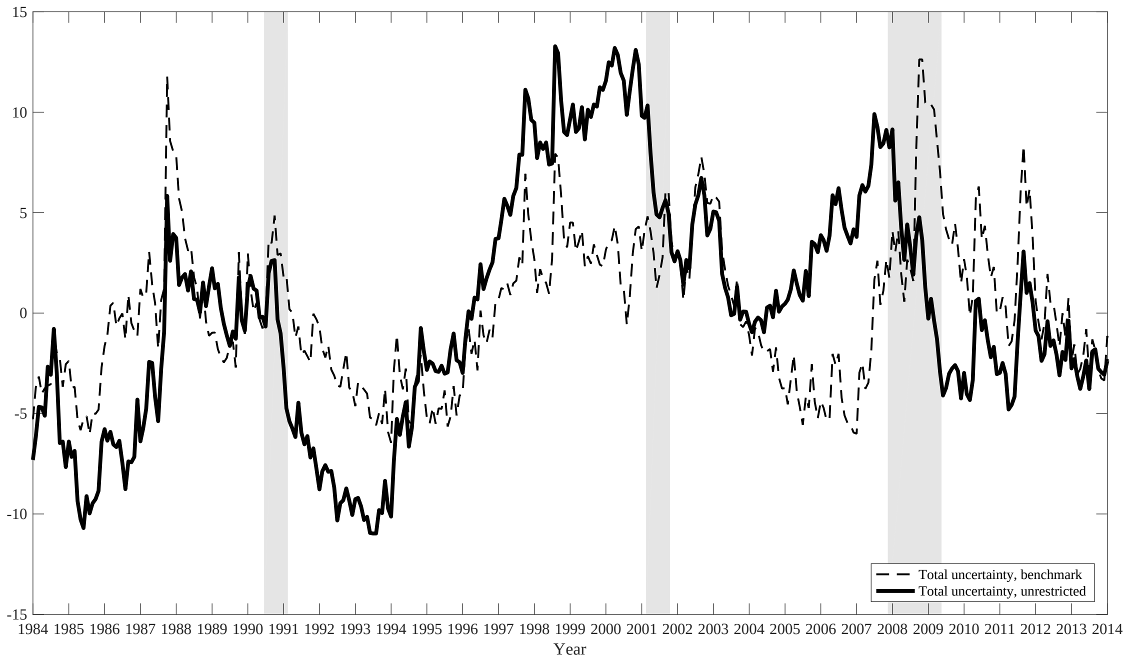


Note: Fraction of the forecast error variance of uncertainty (the expected sum of rv over the next 24 months), employment, and industrial production to shocks to rv and uncertainty in the VAR of figure 2.

Figure 4: Fitted uncertainty



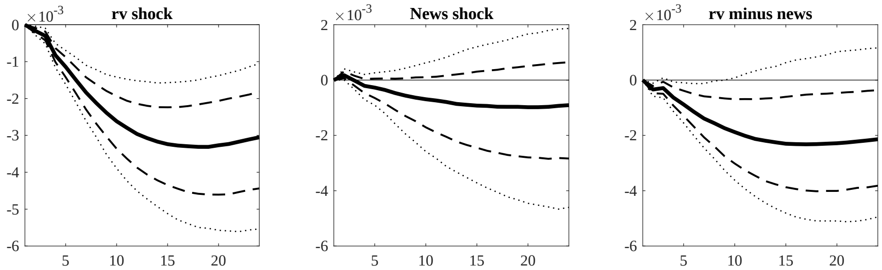
(a) Decomposition of total uncertainty in the benchmark specification



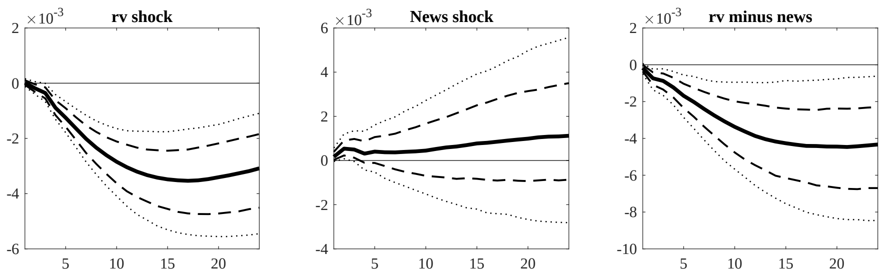
(b) Total uncertainty in benchmark vs. unrestricted specification

Note: The top panel reports a decomposition of total uncertainty (the conditional expectation of the sum of rv over the next 24 months) between the component driven by the rv shock and the component driven by the uncertainty shock in the benchmark model. The bottom panel reports the total uncertainty in the benchmark model and in the unrestricted specification.

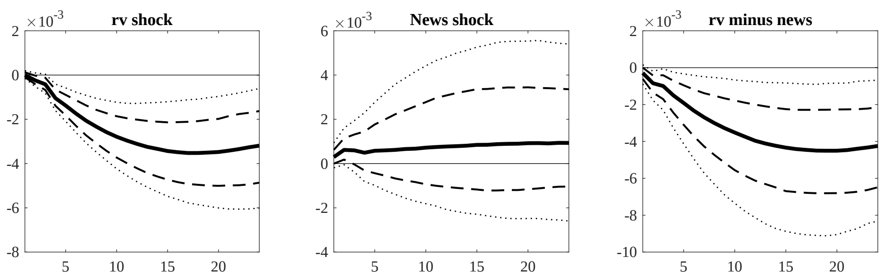
Figure 5: Response of employment to rv and uncertainty shocks across specifications



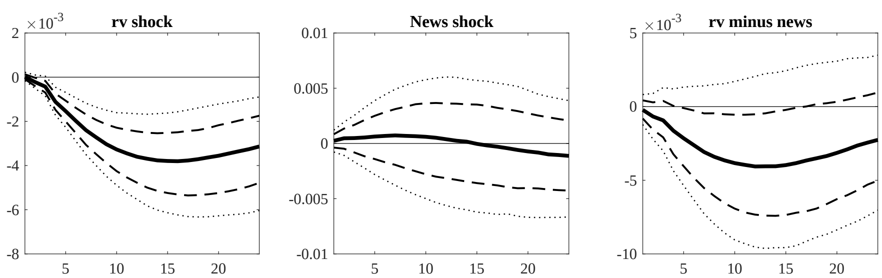
(a) rv and v_1 ordered last



(b) Replacing v_1 with v_6



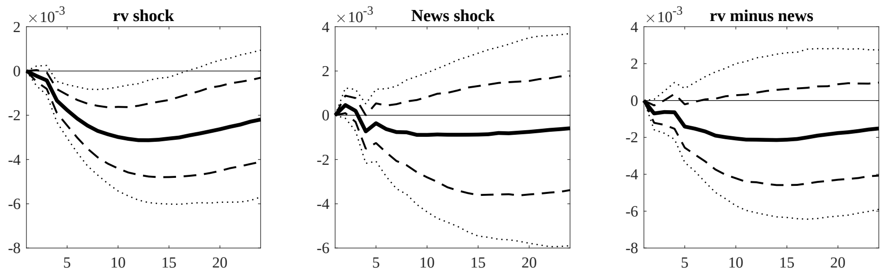
(c) Lasso



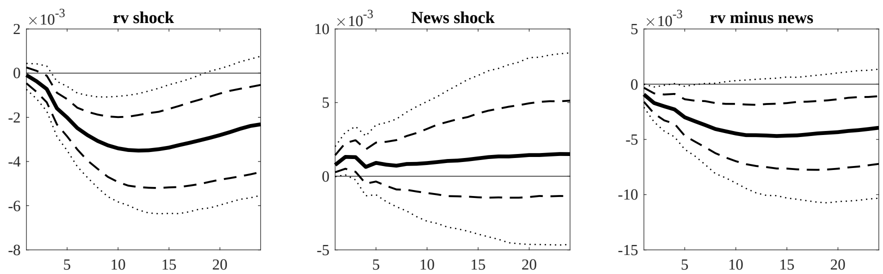
(d) Unrestricted

Note: Response of employment to RV shocks (left panels) and news shocks (middle panels) with the difference in the right panel and different model specifications in each row. Row (a) orders rv and v_1 last. Row (b) uses v_6 instead of v_1. Row (c) uses lasso to estimate the VAR (see section A.2 for details). Row (d) estimates the benchmark VAR without any coefficient restrictions.

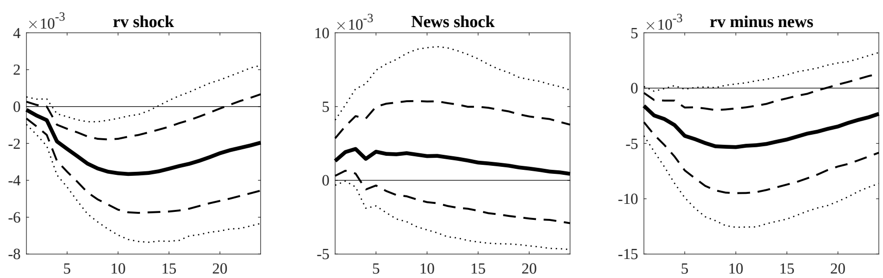
Figure 6: Response of IP to rv and uncertainty shocks across specifications



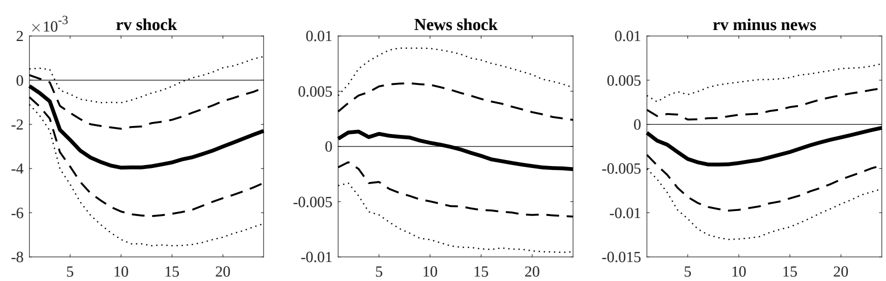
(a) rv and $v1$ ordered last



(b) Replacing v_1 with v_6



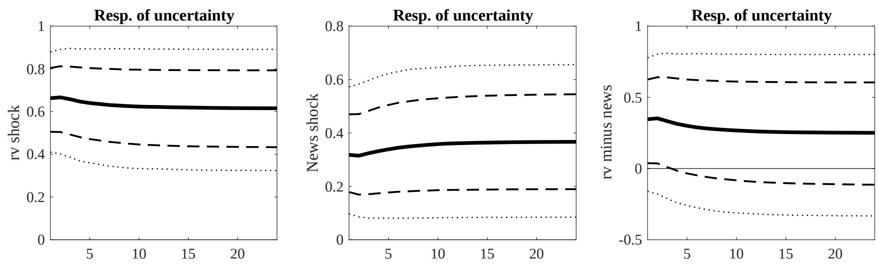
(c) Lasso



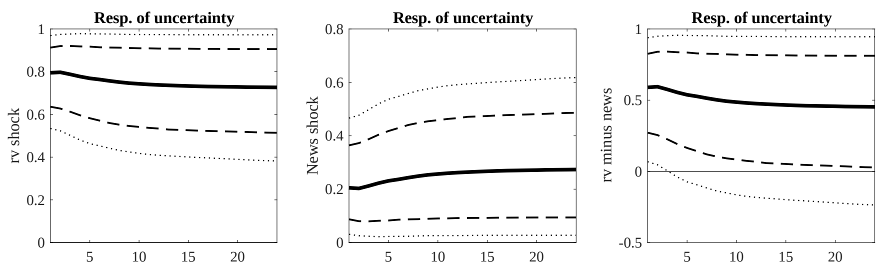
(d) Unrestricted

Note: See figure 5

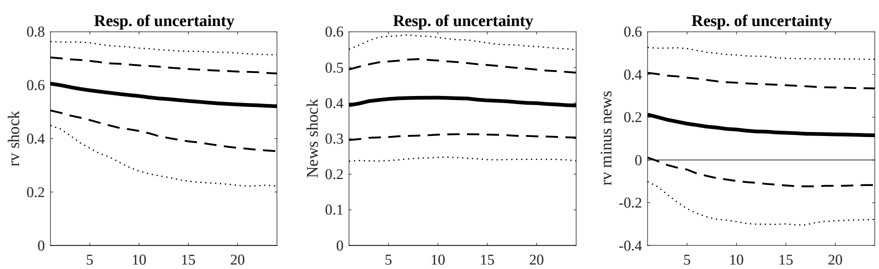
Figure 7: FEVD decomposition of rv across models



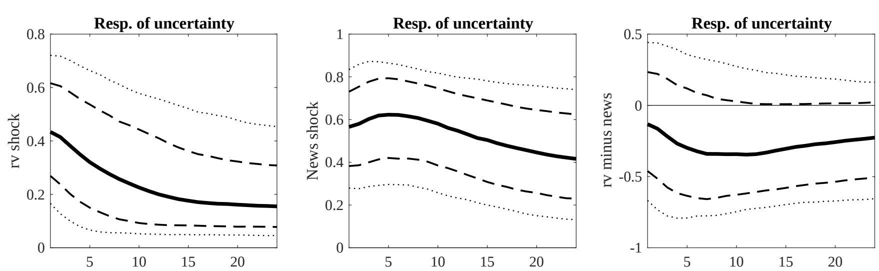
(a) rv and $v1$ ordered last



(b) Replacing v_1 with v_6



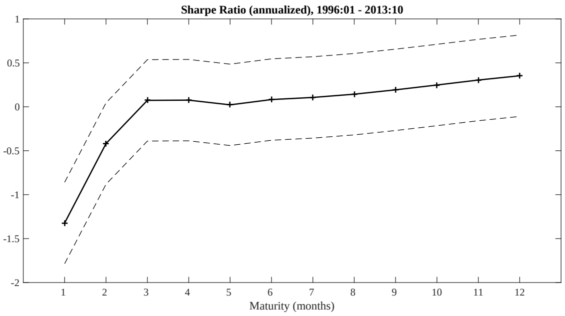
(c) Lasso



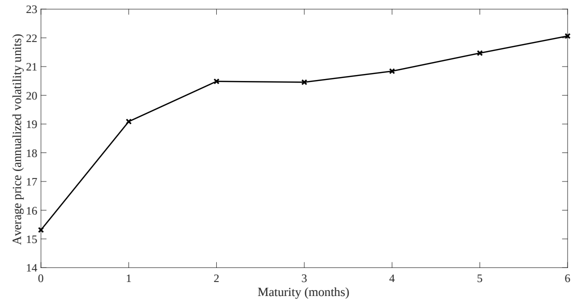
(d) Unrestricted

Note: See figure 5.

Figure 8: Forward variance claims: returns and prices



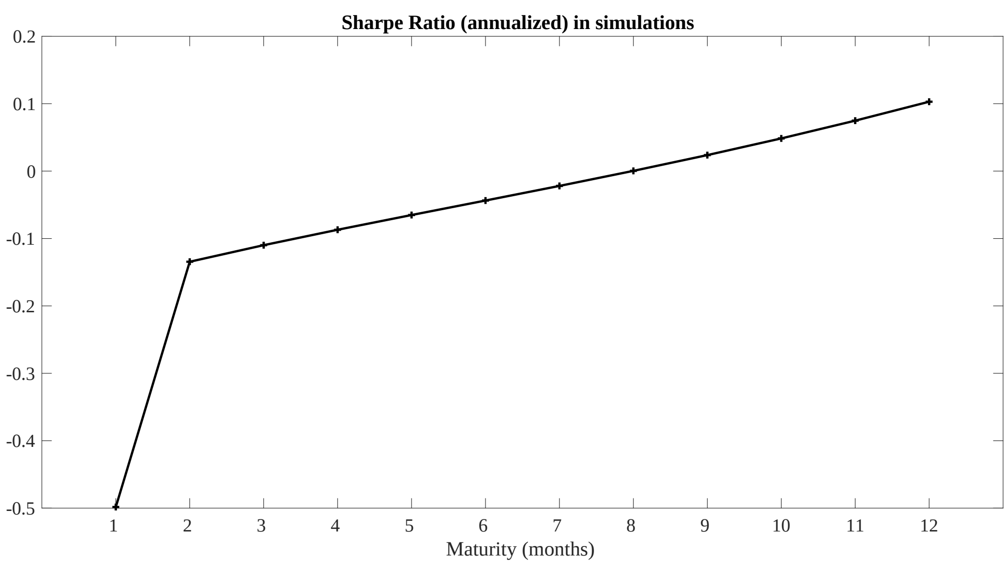
(a) Sharpe ratios



(b) Average prices

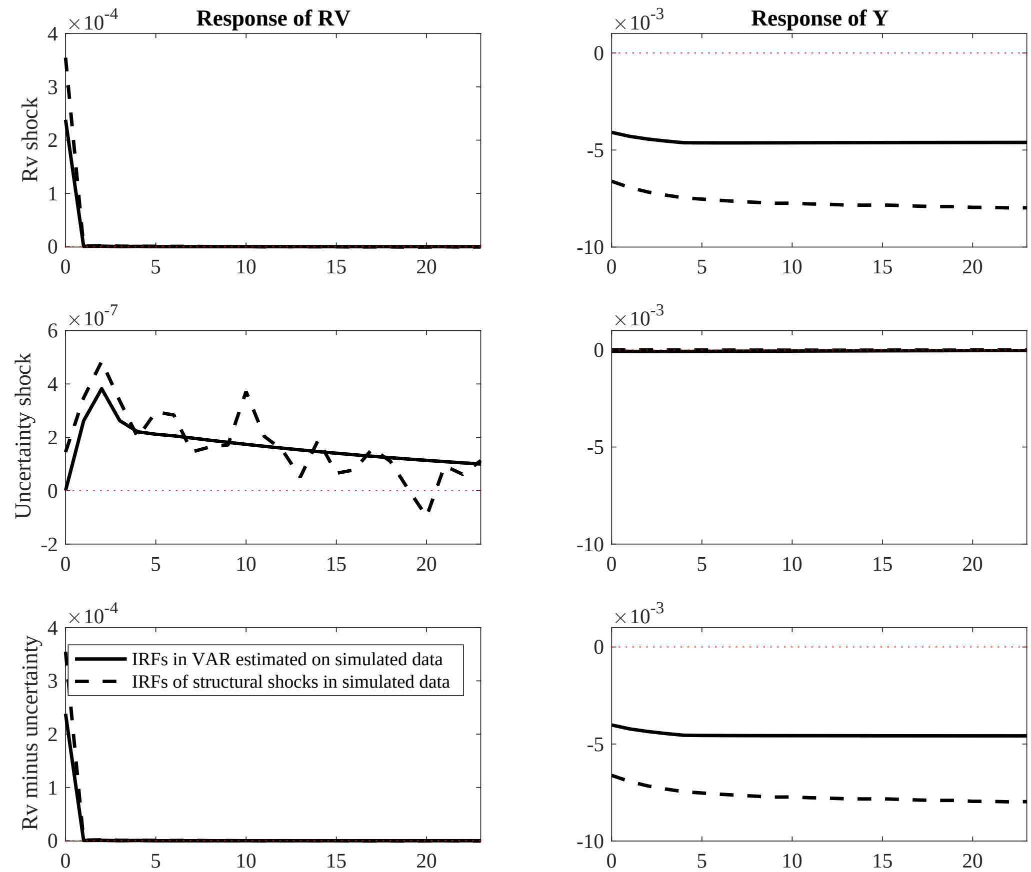
Note: Panel A shows the annualized Sharpe ratio for the forward variance claims, constructed using variance swaps. The returns are calculated assuming that the investment in an n-month variance claim is rolled over each month. Dotted lines represent 95% confidence intervals. All tests for the difference in Sharpe ratio between the 1-month variance swap and any other maturity confirm that they are statistically different with a p-value of 0.03 (for the second month) and < 0.01 (for all other maturities). The sample used is 1996–2013. For more information on the data sources, see Dew-Becker et al. (2017). Panel B shows the average prices across maturities of synthetic forward variance claims constructed from option prices for the period 1983–2014. All prices are reported in annualized standard deviation units. Maturity zero corresponds to average realized volatility.

Figure 9: Annual Sharpe ratios on forward claims (simulated structural model)



Note: Annual Sharpe ratios on forward variance claims in the simulated model of section 7. The Sharpe ratios are constructed as in Figure 8.

Figure 10: IRFs from structural model



Note: Impulse response functions from data simulated from the model in Section 7. Solid lines correspond to IRFs estimated using our VAR methodology as in Figure 2. Dashed lines correspond to IRFs for the two structural shocks $J\nu_t$ and η_t.

Table 1: Predictability of 6-month rv

Predictors	(1)	(2)	(3)	(4)	(5)
rv_t	0.16*	0.15	0.16*	0.14	0.09
	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)
$v_{1,t}$	0.57***	0.46***	0.60**	0.56***	0.60***
	(0.14)	(0.16)	(0.29)	(0.14)	(0.15)
$v_{6,t}$			-0.03		
			(0.30)		
rv_{t-1}		0.12**			
		(0.05)			
FFR				0.01	
				(0.02)	
$\Delta empl$				-2.93	
				(10.92)	
Δip				-5.74	
				(3.52)	
PC1					-0.010
					(0.009)
PC2					-0.017
					(0.012)
PC3					-0.011
					(0.007)
$R_{S\&P}$					0.12
					(0.41)
Default spread					0.07
					(0.09)
Adj. R^2	0.44	0.45	0.44	0.45	0.46

Note: Results of linear predictive regressions of realized volatility over the next six months on lagged rv, option-implied volatility, and various macroeconomic variables, with Hansen–Hodrick (1980) standard errors using a 6-month lag window. PC1–3 are principal components from the data set used in Ludvigson and Ng (2007). The default spread is the difference in yields on Baa and Aaa bonds. The sample is 1983–2014.

Table 2: Skewness

Panel A: real economic activity	Monthly	Quarterly	Start of sample (year)
Employment	-0.41	-0.41	1948
Capacity Utilization	-1.02	-1.30	1967
IP	0.17	-0.16	1948
IP, starting 1960	-0.93	-1.28	1960
Y		-0.11	1947
C		-0.28	1947
I		-0.03	1947

Panel B: skewness of S&P 500 monthly returns	
Implied (since 1990)	-1.81
Realized (since 1926)	0.36
Realized (since 1948)	-0.42
Realized (since 1990)	-0.61

Note: Panel A reports the skewness of changes of employment, capacity utilization, industrial production (beginning both in 1948 and in 1960), GDP, consumption and investments. The first column reports the skewness of monthly changes, the second column the skewness of quarterly changes. Panel B reports the realized skewness of S&P 500 monthly returns in different periods, as well as the implied skewness computed by the CBOE using option prices.

Table 3: Model Calibration

Panel A: Moments	Model			Data		
	Mean	Std.	Skewness	Mean	Std.	Skewness
Returns	1.24	2.37	-3.37	7.46	14.77	-0.48
Output	1.99	1.90	-2.00	1.33	1.92	-0.11
Investment	1.99	3.50	-1.97	2.35	7.43	-0.03
Consumption	1.99	1.55	-1.53	1.20	1.06	-0.28

Panel B: Corr. of VAR and structural shocks		Structural shocks		
		Jv_t	η_t	ε_t
	RV	0.96	0.01	0.00
VAR identified shocks	Uncertainty	-0.01	0.89	-0.14

Note: Panel A reports the mean, standard deviation, and skewness of financial and macroeconomic variables in the data and in the model. Panel B shows the correlation between the structural shocks in the model and the shocks identified in the VAR.

A.1 Construction of option-implied volatility, V_n

In this section we describe the details of the procedure we use to construct model implied uncertainty at different horizons, starting from our dataset of end-of-day prices for American options on S&P 500 futures from the CME.

A.1.1 Main steps of construction of V_n

A first step in constructing the model-free implied volatility is to obtain implied volatilities corresponding to the observed option prices. We do so using a binomial model.[1] For the most recent years, CME itself provides the implied volatility together with the option price. For this part of the sample, the IV we estimate with the binomial model and the CME's IV have a correlation of 99%, which provides an external validation on our implementation of the binomial model.

Once we have estimated these implied volatilities, we could in theory simply invert them to yield implied prices of European options on forwards. These can then be used to compute V_n directly as described in equation (11).

In practice, however, an extra step is required before inverting for the European option prices and integrating to obtain the model-free implied volatility. The model-free implied volatility defined in equation (11) depends on the integral of option prices over *all strikes*, but option prices are only observed at discrete strikes. We are therefore forced to interpolate option prices between available strikes and also extrapolate beyond the bounds of observed strikes.[2] Following the literature, we fit a parametric model to the Black–Scholes implied volatilities of the options and use the model to then interpolate and extrapolate across all strikes (see, for example, Jiang and Tian (2007), Carr and Wu (2009), Taylor, Yadav, and Zhang (2010), and references therein). Only after this extra interpolation-extrapolation step, the fitted implied volatilities are then inverted to yield option prices and compute V_n according to equation (11). To interpolate and extrapolate the implied volatility curve, we use the SVI (stochastic volatility inspired) model of Gatheral and Jacquier (2014).

In the next sections, we describe in more detail the interpolation-extrapolation step of the procedure (SVI fitting) as well as our construction of V_n after fitting the SVI curve. Finally, we report a description of the data we use and some examples and diagnostics on the SVI fitting method.

A.1.2 SVI interpolation: theory

There are numerous methods for fitting implied volatilities across strikes. Homescu (2011) provides a thorough review. We obtained the most success using Gatheral's SVI model (see Gatheral and Jacquier 2014). SVI is widely used in financial institutions because it is parsimonious but also known

[1] See for example Broadie and Detemple (1996) and Bakshi, Kapadia, and Madan (2003), among others.

[2] See Jiang and Tian (2007) for a discussion of biases arising from the failure to interpolate and extrapolate.

to approximate well the behavior of implied volatility in fully specified option pricing models (e.g. Gatheral and Jacquier (2011)); SVI also satisfies the limiting results for implied volatilities at very high and low strikes in Lee (2004), and, importantly, ensures that no-arbitrage conditions are not violated.

The SVI model simply assumes a hyperbolic relationship between implied variance (the square of the Black–Scholes implied volatility) and the log moneyness of the option, k (log strike/forward price).

$$\sigma_{BS}^2(k) = a + b\left(\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2}\right)$$

where $\sigma_{BS}^2(k)$ is the implied variance under the Black–Scholes model at log moneyness k. SVI has five parameters: a, b, ρ, m, and σ. The parameter ρ controls asymmetry in the variances across strikes. Because the behavior of options at high strikes has minimal impact on the calculation of model-free implied volatilities, and because we generally observe few strikes far above the spot, we set $\rho = 0$ (in simulations with calculating the VIX for the S&P 500 – for which we observe a wide range of options – we have found that including or excluding ρ has minimal impact on the result).

We fit the parameters of SVI by minimizing the sum of squared fitting errors for the observed implied volatilities. Because the fitted values are non-linear in the parameters, the optimization must be performed numerically. We follow the methodology in Zeliade (2009) to analytically concentrate a and b out of the optimization. We then only need to optimize numerically over σ and m (as mentioned above, we set $\rho = 0$). We optimize with a grid search over $\sigma \times m = [0.001, 10] \times [-1, 1]$ followed by the simplex algorithm.

For many date/firm/maturity triplets, we do not have a sufficient number of contract observations to fit the implied volatility curve (i.e. sometimes fewer than four). We therefore include strike/implied volatility data from the two neighboring maturities and dates in the estimation. The parameters of SVI are obtained by minimizing squared fitting errors. We reweight the observations from the neighboring dates and maturities so that they carry the same amount of weight as the observations from the date and maturity of interest. Adding data in this way encourages smoothness in the estimates over time and across maturities but it does not induce a systematic upward or downward bias. We drop all date/firm/maturity triplets for which we have fewer than four total options with $k < 0$ or fewer than two options at the actual date/firm/maturity (i.e. ignoring the data from the neighboring dates and maturities).

When we estimate the parameters of the SVI model, we impose conditions that guarantee the absence of arbitrage. In particular, we assume that $b \le \frac{4}{(1+|\rho|)T}$, which when we assume $\rho = 0$, simplifies to $b \le \frac{4}{T}$. We also assume that $\sigma > 0.0001$ in order to ensure that the estimation is well defined. Those conditions do not necessarily guarantee, though, that the integral determining the model-free implied volatility is convergent (the absence of arbitrage implies that a risk-neutral probability density exists – it does not guarantee that it has a finite variance). We therefore eliminate observations where the integral determining the model-free implied volatility fails to converge numerically. Specifically, we eliminate observations where the argument of the integral does not approach zero as the log strike rises above two standard deviations from the spot or falls more than five standard deviations below the strike (measured based on the at-the-money implied volatility).

A.1.3 Construction of V_n from the SVI fitted curve

After fitting the SVI curve for each date and maturity, we compute the integral in equation (11) numerically, over a range of strikes from -5 to +2 standard deviations away from the spot price.[3] We then have V_n for every firm/date/maturity observation. The model-free implied volatilities are then interpolated (but not extrapolated) to construct V_n at maturities from 1–6 months for each firm/date pair.

A.1.4 Data description and diagnostics of SVI fitting

Our dataset consists of 2.3 million end-of-day prices for all American options on S&P 500 futures from the CME.

When more than one option (e.g. a call and a put) is available at any strike, we compute IV at that strike as the average of the observed IVs. We keep only IVs greater than zero, at maturities higher than 9 days and lower than 2 years, for a total of 1.9 million IVs. The number of available options has increased over time, as demonstrated by Figure A.2 (top panel), which plots the number of options available for V_n estimation in each year.

The maturity structure of observed options has also expanded over time, with options being introduced at higher maturities and for more intermediate maturities. Figure A.1 (top panel) reports the cross-sectional distribution of available maturities in each year to estimate the term structure of the model-free implied volatility. The average maturity of available options over our sample was 4 months, and was relatively stable. The maximum maturity observed ranged from 9 to 24 months and varied substantially over time.

Crucial to compute the model-free implied volatility is the availability of IVs at low strikes, since options with low strikes receive a high weight in the construction of V_n. The bottom panel of Figure A.1 reports the minimum observed strike year by year, in standard deviations below the spot price. In particular, for each day we computed the minimum available strike price, and the figure plots the average of these minimum strike price across all days in each year; this ensures that the number reported does not simply reflect outlier strikes that only appear for small parts of each year.

Figure A.1 shows that in the early part of our sample, we can typically observe options with strikes around 2 standard deviations below the spot price; this number increases to around 2.5 towards the end of the sample.

[3] In general this range of strikes is sufficient to calculate V_n. However, the model-free implied volatility technically involves an integral over the entire positive real line. Our calculation is thus literally a calculation of Andersen and Bondarenko's (2007) corridor implied volatility. We use this fact also when calculating realized volatility.

These figures show that while the number of options was significantly smaller at the beginning of the sample (1983), the maturities observed and the strikes observed did not change dramatically over time.

Figure A.3 shows an example of the SVI fitting procedure for a specific day in the early part of our sample (November 7th 1985). Each panel in the figure corresponds to a different maturity. On that day, we observe options at three different maturities, of approximately 1, 4, and 8 months. In each panel, the x's represent observed IVs at different values of log moneyness k. The line is the fitted SVI curve, that shows both the interpolation and the extrapolation obtained from the model.

Figure A.4 repeats the exercise in the later part of our sample (Nov. 1st 2006), where many more maturities and strikes are available.

Both figures show that the SVI model fits the observed variances extremely well. The bottom panel of Figure A.2 shows the average relative pricing error for the SVI model in absolute value. The graph shows that the typical pricing error for most of the sample is around 0.02, meaning that the SVI deviates from the observed IV by around 2% on average. Only in the very first years (up to 1985) pricing errors are larger, but still only around 10% of the observed IV.

Overall, the evidence in this section shows that our observed option sample since 1983 has been relatively stable along the main dimensions that matter for our analysis – maturity structure, strikes observed, and goodness of fit of the SVI model.

A.2 Lasso

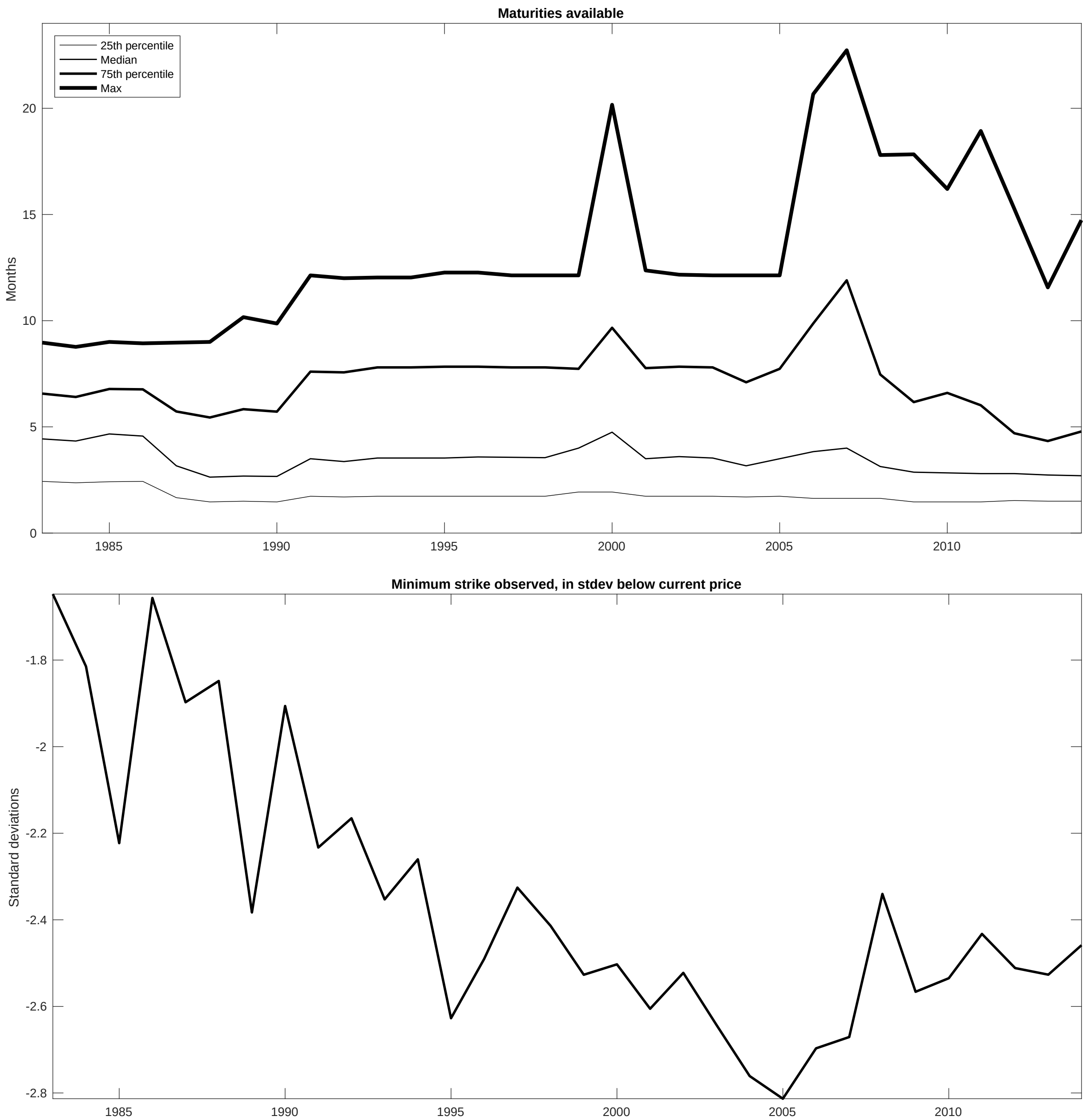
Lasso is a regularization method for regressions that penalizes coefficients based on their absolute values. Specifically, the objective that is minimized under lasso is the sum of squared residuals plus a tuning parameter, which we denote λ, multiplied by the sum of the absolute values of the coefficients.

Lasso is not invariant to the scaling of the variables in the regression. We therefore rescale the variables as follows. rv, v_6, and $slope$ are all translated into z-scores. The three macro variables (FFR, emp, and ip) are multiplied by constants so that their first differences have unit variances. We use that transformation because those three variables have approximate unit roots in our sample.

We examine two methods to select λ. The first is to use leave-one-out cross validation. We choose λ separately for the three volatility and three macro series. The cross-validation criterion implies setting $\lambda = 0.013$ for the volatility series and $\lambda = 0$ (i.e. no lasso) for the macro series. the results reported in the text use this choice of λ.

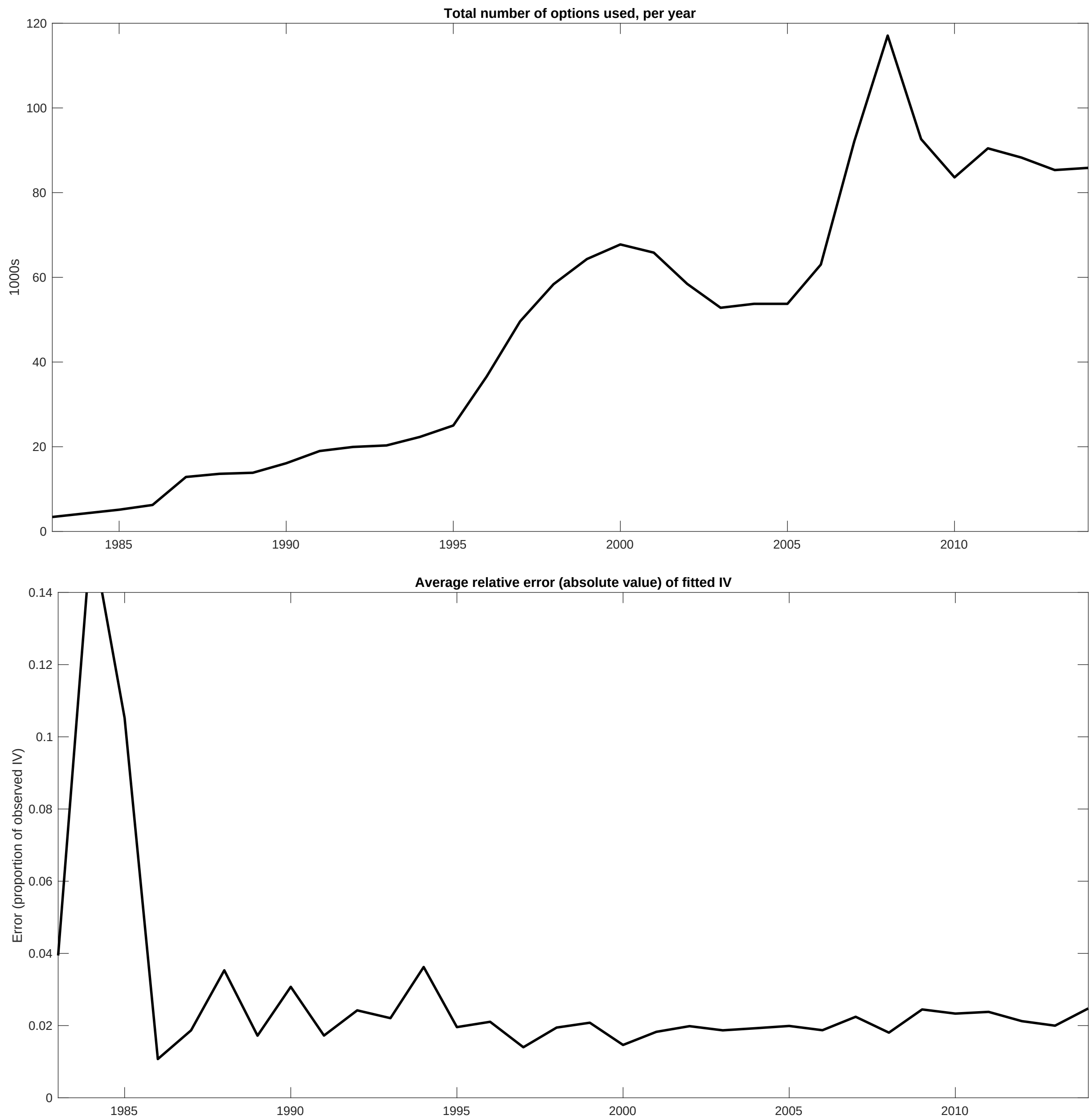
The second method is to choose the smallest (i.e. least restrictive) value of λ that causes the coefficients on all the lags of the macro variables in the rv equation to be zero, consistent with the benchmark specification. The motivation for this method is that it takes the restrictions that we impose on economic grounds and then essentially tries to impose similar restrictions on the other equations, for the sake of parity. In this case we find a value of λ of 0.055. The results with this value are not reported here but are consistent with out main findings. In this case we find slightly negative effects for uncertainty shocks, but they are still statistically significantly less negative than those for rv shocks, and the forecast error variance decompositions put an upper bound on the fraction driven by uncertainty shocks of 15 percent.

Figure A.1: Maturities and strikes in the CME dataset



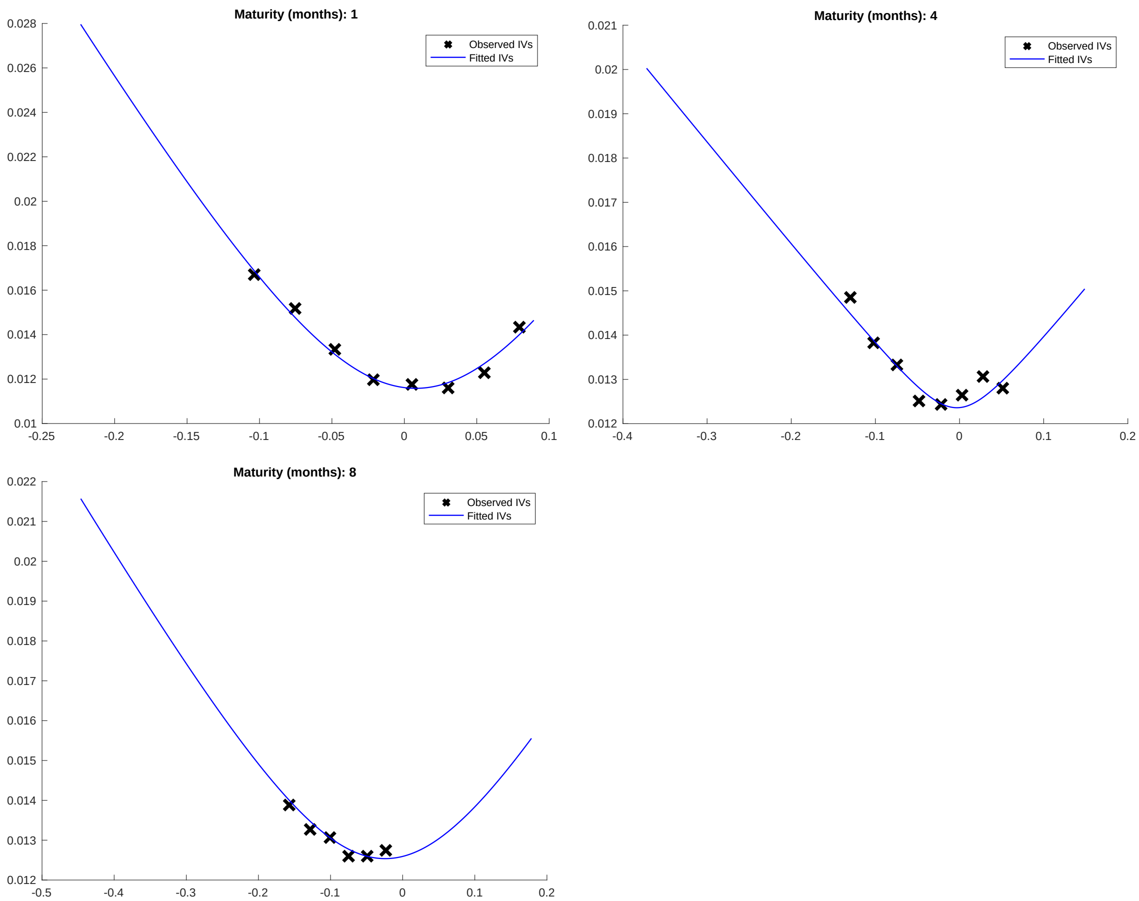
Note: The top panel reports the distribution of maturities of options used to compute implied volatility in each year, in months. The bottom panel reports the average minimum strike in each year, in standard deviations below the forward price. The number is obtained by computing the minimum observed strike in each date and at each maturity (in standard deviations below the forward price), and then averaging it within each year to minimize the effect of outliers.

Figure A.2: Number of options to construct implied volatility and pricing errors



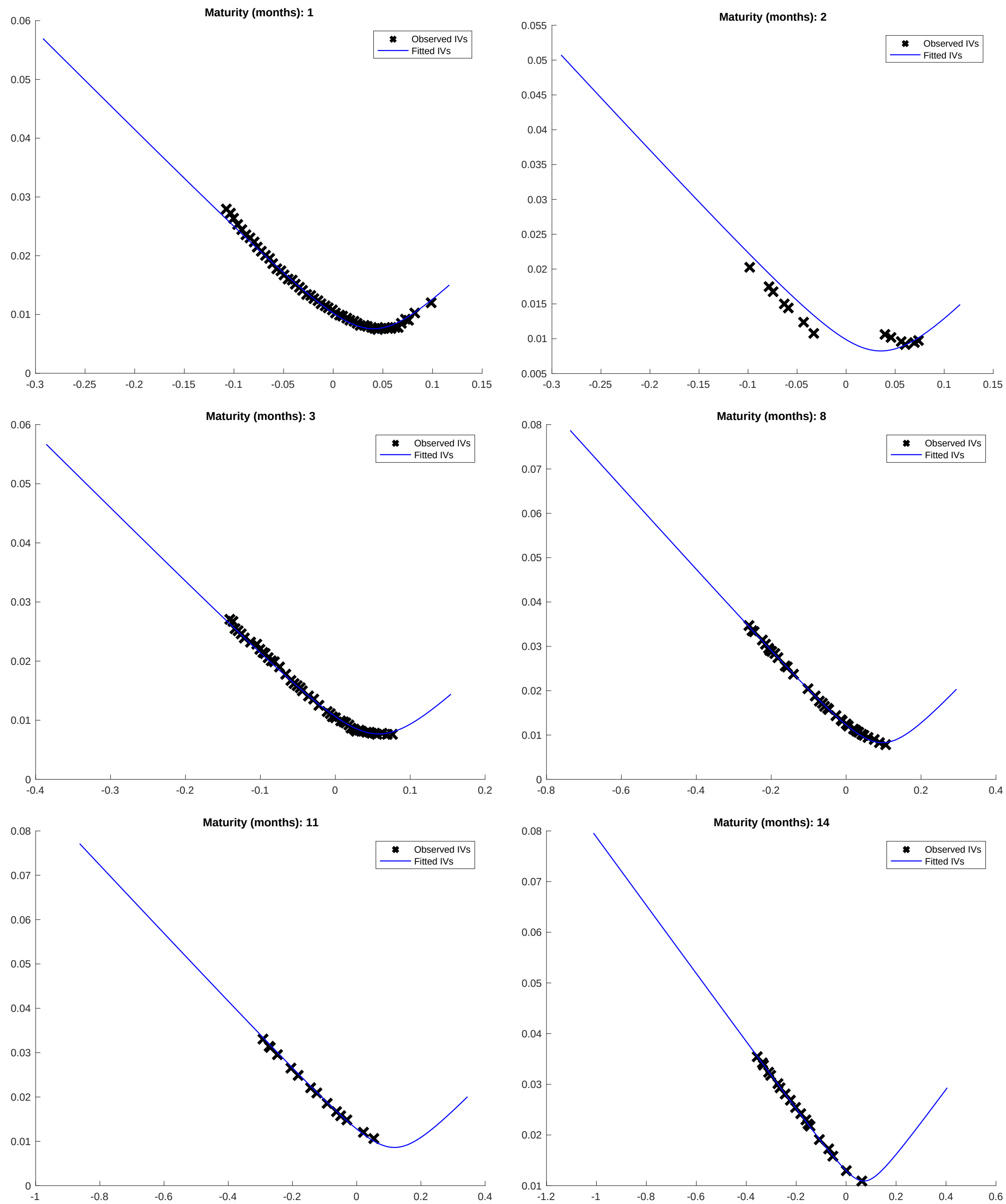
Note: The top panel reports the number of options used to compute implied volatility in each year, in thousands. The bottom panel reports the average absolute value of the pricing error of the SVI fitted line relative to the observed implied variances, in proportional terms (i.e. 0.02 means absolute value of the pricing error is 2% of the observed implied variance).

Figure A.3: SVI fit: 11/7/1985



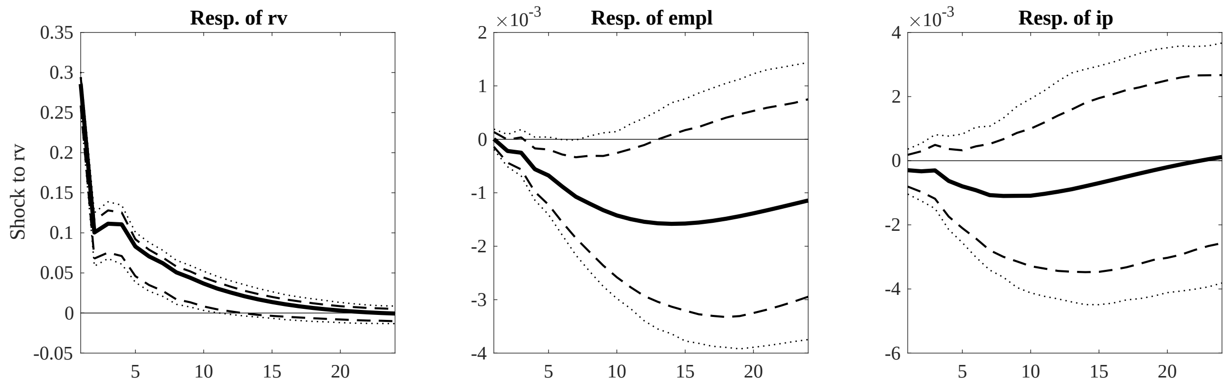
Note: Fitted implied variance curve on 11/7/1987, for the three available maturities. X axis is the difference in log strike and log forward price. x's correspond to the observed implied variances, and the line is the fitted SVI curve.

Figure A.4: SVI fit: 11/1/2006



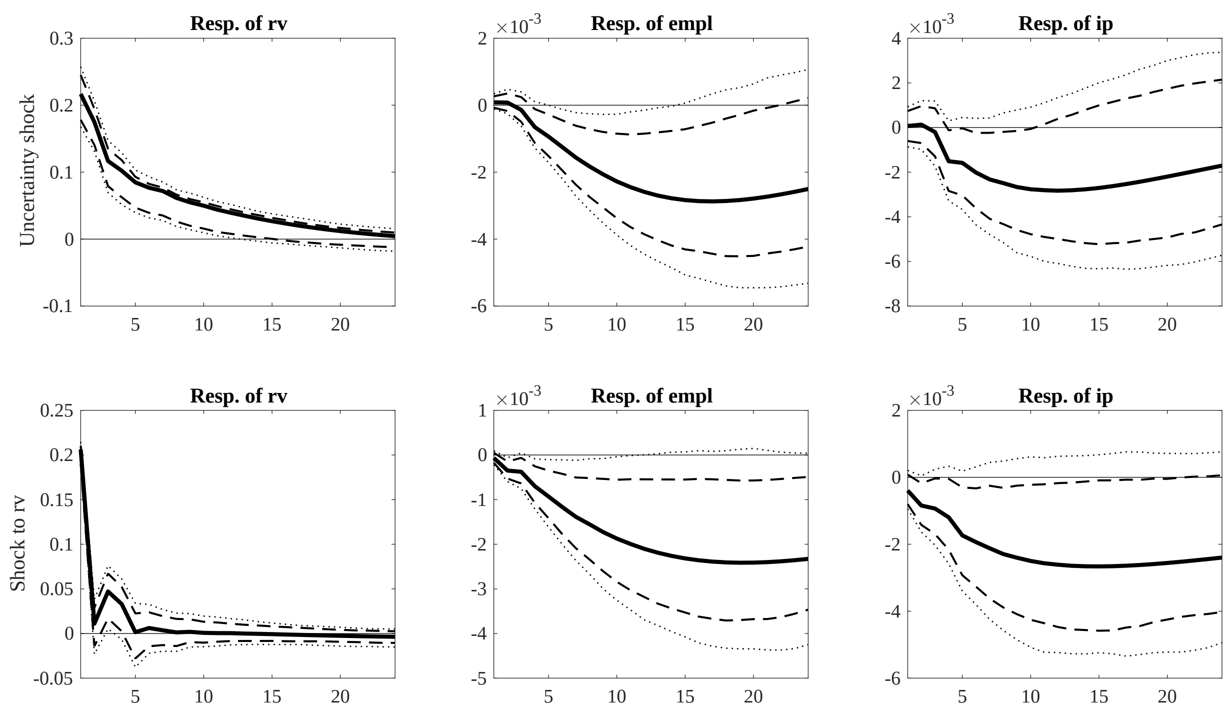
Note: Fitted implied variance curve on 11/1/2006, for the three available maturities. X axis is the difference in log strike and log forward price. x's correspond to the observed implied variances, and the line is the fitted SVI curve. On 11/1/2006 also a maturity of 5 months was available (not plotted for reasons of space).

Figure A.5: Impulse response functions from VAR with v_1 but not rv



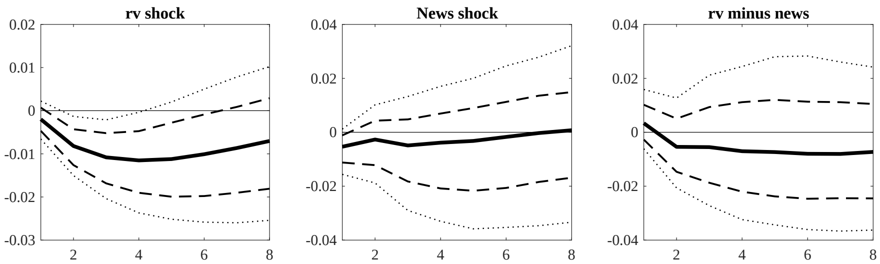
Note: The figure shows responses of volatility (measured by v_1), log employment, and log industrial production to a reduced-form shock to v_1 in a VAR with v_1, the Fed funds rate, log employment, and log industrial production with 68% and 90% confidence intervals. Sample period 1986-2014.

Figure A.6: Impulse response functions from VAR ordering uncertainty first and rv second

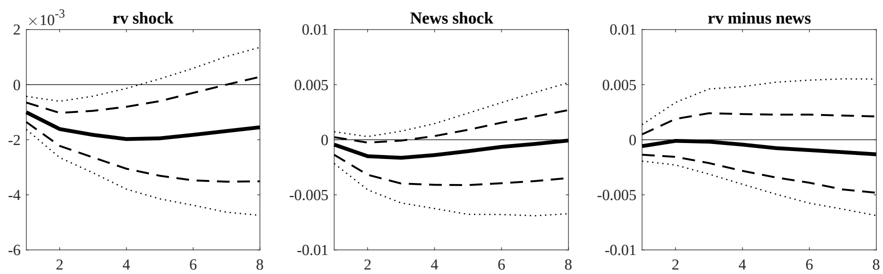


Note: See figure 2. Unlike in the baseline identification, the identified uncertainty shock is not orthogonalized with respect to rv. The rv shock in this case is the remaining part of reduced-form innovation to rv that is not spanned by the uncertainty shock.

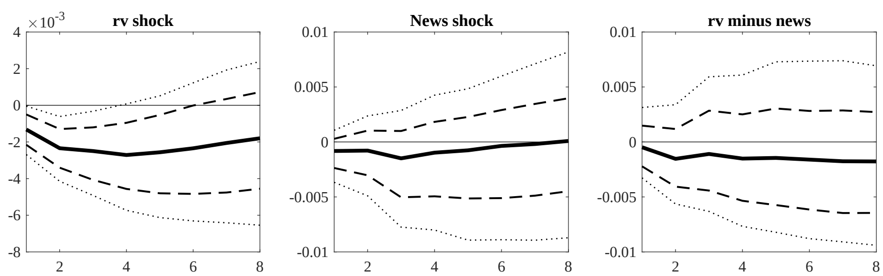
Figure A.7: Robustness: data from Basu and Bundick (2016), using v_1



(a) Investment



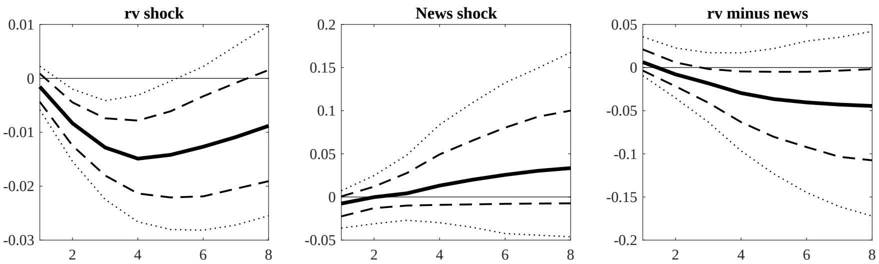
(b) Consumption



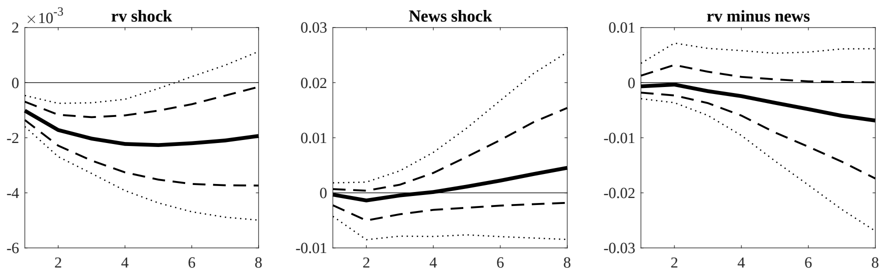
(c) GDP

Note: See figure 2. Here we use the quarterly data from Basu and Bundick (2016) as the macro time series.

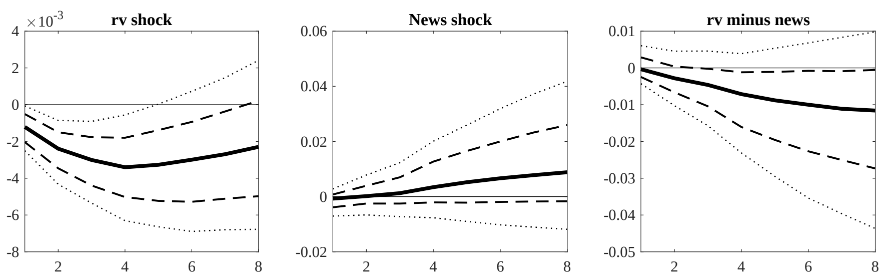
Figure A.8: Robustness: data from Basu and Bundick (2016), using v_6



(a) Investment



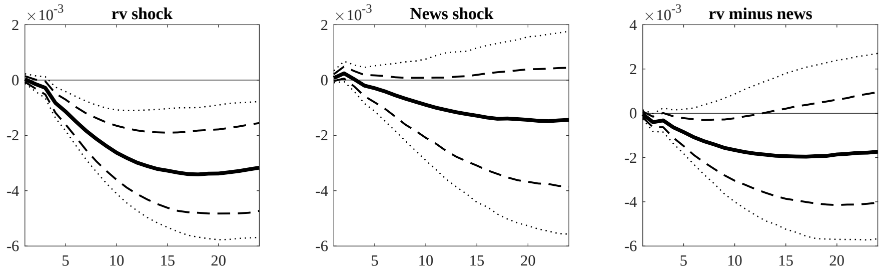
(b) Consumption



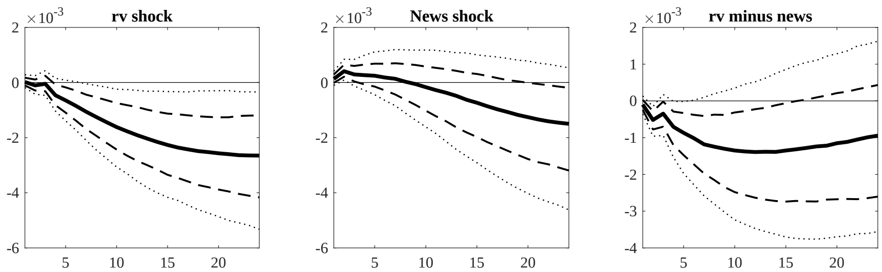
(c) GDP

Note: See figure 2. Here we use the quarterly data from Basu and Bundick (2016) as the macro time series. We use v_6 instead of v_1.

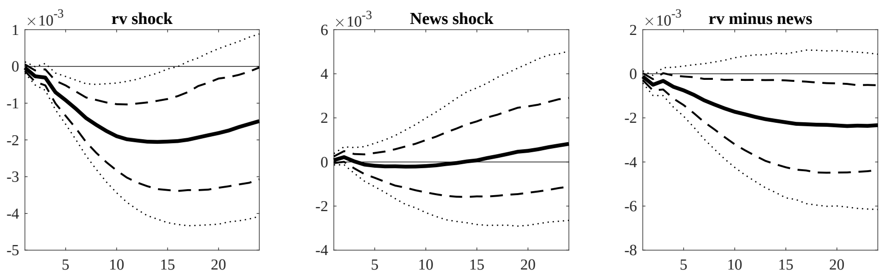
Figure A.9: Robustness: response of Employment to rv and uncertainty shocks across specifications



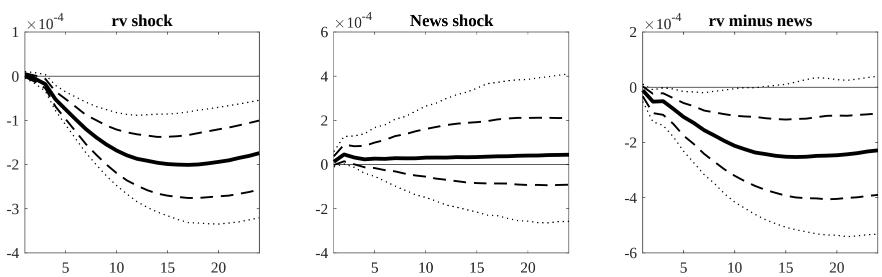
(a) Detrending the macroeconomic time series via HP filter



(b) Subperiod 1988-2006 (excluding 1987 crash and financial crisis)



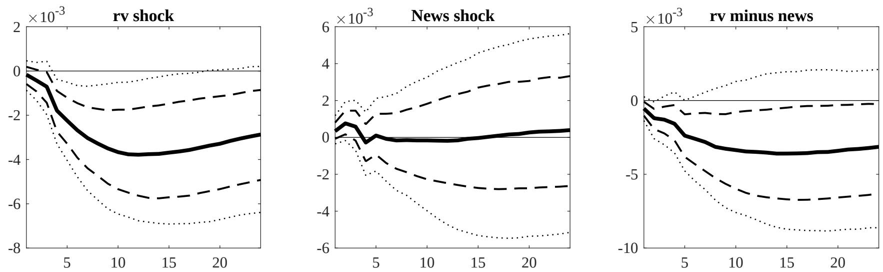
(c) Adding the S&P 500 level as first shock



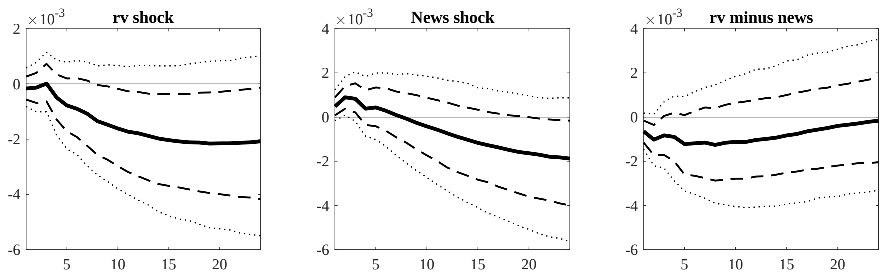
(d) Using RV and V_1 in levels, not logs

Note: Response of employment to RV shocks (left panels) and uncertainty (middle panels) with the difference in the right panel and different model specifications in each row. Row (a) detrends the macroeconomic time series via HP filter. Row (b) estimates the VAR in the subsample 1988-2006, which excludes both RV peaks (1987 crash and financial crisis). Row (c) orthogonalizes both the rv and the uncertainty shocks with respect to the reduced-form innovation in the S&P 500, as in Bloom (2009). Row (d) uses RV and V_1 in levels, not logs.

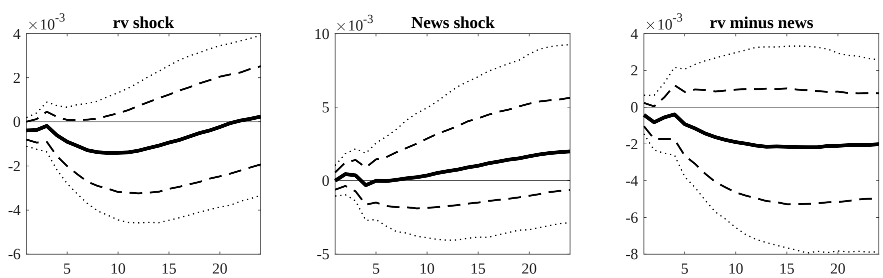
Figure A.10: Robustness: response of IP to rv and uncertainty shocks across specifications



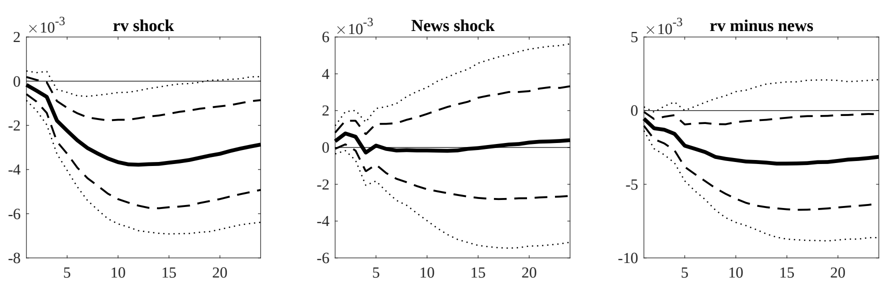
(a) Detrending the macroeconomic time series via HP filter



(b) Subperiod 1988-2006 (excluding 1987 crash and financial crisis)



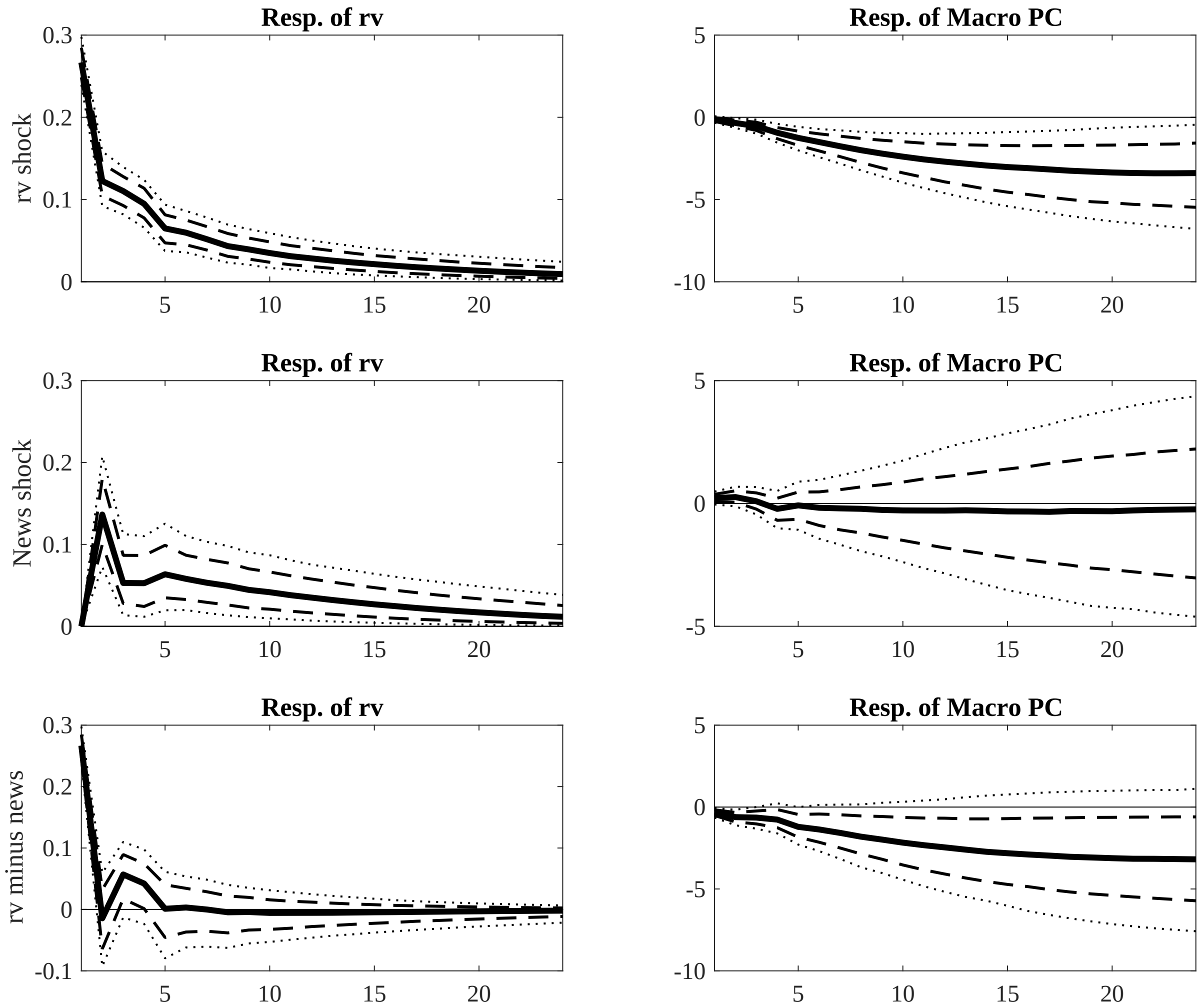
(c) Adding the S&P 500 level as first shock



(d) Using RV and V_1 in levels, not logs

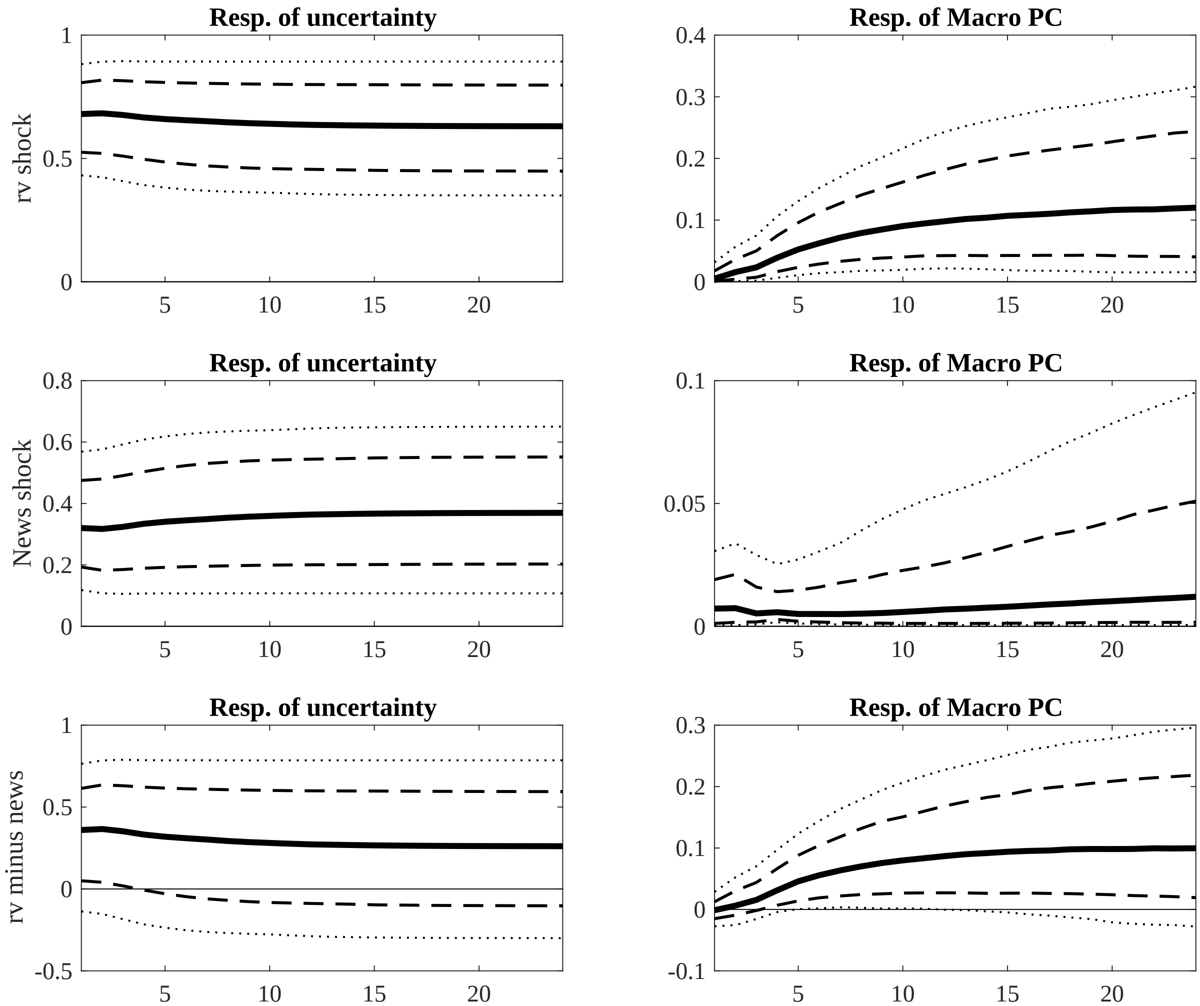
Note: See figure A.9. In this case the responses of IP are reported instead of employment.

Figure A.11: Impulse response functions from FAVAR



Note: Responses of rv and the first principal component from Ludvigson and Ng (2007) to shocks to rv and the identified uncertainty shock, in a VAR with rv, v_1, federal funds rate, and the PC. The IRFs are scaled so that the two shocks have equal cumulative effects on rv over months 2–24 following the shock. The sample period is 1983–2014. The dotted lines are 68% and 90% confidence intervals.

Figure A.12: Forecast error variance decomposition in FAVAR



Note: Fraction of the forecast error variance of uncertainty (the expected sum of rv over the next 24 months) and the first principal component from Ludvigson and Ng (2007) to shocks to rv and uncertainty in the VAR of figure A.11.

Table A.1: Predictability of rv with and without v_1 as predictor

Predictors	(1)	(2)	(3)	(4)
rv		0.14		0.09
		(0.09)		(0.09)
v_1		0.56***		0.60***
		(0.14)		(0.15)
FFR	0.01	0.01		
	(0.03)	(0.02)		
Δemp	-31.03*	-2.93		
	(17.67)	(10.92)		
Δip	-5.27	-5.74		
	(5.37)	(3.52)		
PC 1			-0.029**	-0.010
			(0.013)	(0.009)
PC 2			-0.014	-0.017
			(0.018)	(0.012)
PC 3			-0.014	-0.011
			(0.010)	(0.007)
$R_{S\&P}$			-1.56***	0.12
			(0.43)	(0.41)
Def			0.21	0.07
			(0.14)	(0.09)
N	377	377	377	377
Adj. R^2	0.10	0.45	0.20	0.46

Note: Regressions of 6-month realized volatility on lagged rv, option-implied volatility, and various macroeconomic variables, with Hansen-Hodrick standard errors using a 6-month lag window. PC 1 – PC3 are the first three principal components from a large set of macroeconomic time series. $R_{S\&P}$ is the return on the S&P 500. Def is the default spread, the gap between yields on Aaa and Baa bonds.

Table A.2: Predictability of 6-month-ahead v_1

Predictors	(1)	(2)	(3)	(5)	(6)
rv_t	0.02	0.01	0.03	0.01	0.01
	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)
$v_{1,t}$	0.57***	0.49***	0.51**	0.54***	0.60***
	(0.13)	(0.15)	(0.26)	(0.13)	(0.13)
$v_{6,t}$			0.06		
			(0.27)		
rv_{t-1}		0.10*			
		(0.06)			
FFR				0.02	
				(0.015)	
$\Delta empl$				-11.94	
				(9.75)	
Δip				-0.15	
				(3.22)	
PC1					-0.010
					(0.008)
PC2					-0.010
					(0.010)
PC3					-0.008
					(0.006)
$R_{S\&P}$					0.78**
					(0.38)
Default spread					-0.05
					(0.08)
Adj. R^2	0.35	0.36	0.35	0.37	0.37

Note: Regressions of 6-month-ahead v_1 on lagged rv, option-implied volatility, and various macroeconomic variables, with Hansen–Hodrick (1980) standard errors using a 6-month lag window. PC1–3 are principal components from the data set used in Ludvigson and Ng (2007). The default spread is the difference in yields on Baa and Aaa bonds. The sample is 1983–2014.