Rare Shocks, Great Recessions

Vasco Cúrdia, Marco Del Negro, Daniel Greenwald*

Federal Reserve Bank of New York and *New York University*

August 8, 2012; First Draft: March 2011

Work in progress

Abstract

We estimate a DSGE model where rare large shocks can occur, by replacing the commonly used Gaussian assumption with a Student-t distribution. We show that the latter is strongly favored by the data in the context of the Smets and Wouters (2007) model, even when we allow for low frequency variation in the shocks' volatility. To assess the quantitative impact of rare shocks on the business cycle we perform a counterfactual experiment where we show that, absent "rare shocks", all recessions would have been of roughly the same magnitude. Further, we show that inference about low frequency changes in volatility – and in particular, inference about the magnitude of Great Moderation – is different once we allow for fat tails. Finally, we show that the evidence of fat tails is just as strong when we exclude the recent financial crisis from our sample.

JEL CLASSIFICATION: C32, E3

KEY WORDS: Bayesian Analysis, DSGE Models, Fat tails, Stochastic volatility, Great Recession.

---

*Vasco Cúrdia: vasco.curdia@ny.frb.org; Marco Del Negro: marco.delnegro@ny.frb.org; Research Department, Federal Reserve Bank of New York, 33 Liberty Street, New York NY 10045. Dan Greenwald: dlg340@nyu.edu; New York University. We thank Ulrich Mueller, Andriy Norets, as well as seminar participants at the 2011 EEA, FRB Chicago, FRB St Louis Econometrics Workshop, Seoul National University (Conference in Honor of Chris Sims), 2011 SCE, for useful comments and suggestions. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

Great Recessions do not happen every decade – which is why they are dubbed "Great" in the first place. To the extent that DSGE models rely on shocks in order to generate macroeconomic fluctuations, they may need to account for the occurrence of rare large shocks. We therefore estimate a linearized DSGE model assuming that shocks are generated from a Student-t distribution, which is designed to capture fat tails, as opposed to the Gaussian distribution, which is the standard assumption in the DSGE literature. The number of degrees of freedom in the Student-t distribution, which determines the likelihood of observing rare large shocks and which we allow to vary across shocks, is estimated from the data. We show that estimating DSGE models with Student-t distributed shocks is a fairly straightforward extension of current methods (described, for instance, in An and Schorfheide (2007)). In fact, the Gibbs sampler is a simple extension of Geweke (1993)'s Gibbs sampler for a linear model to the DSGE framework.[1]

In light of the evidence provided by several recent papers in the DSGE literature (Justiniano and Primiceri (2008), Fernández-Villaverde and Rubio-Ramírez (2007), Liu et al. (2011), among others), we also allow for low frequency changes in the volatility of the shocks. We do so because ignoring these low frequency movements in volatility may bias the results toward finding evidence in favor of fat tails, as we discuss below. Specifically, we follow the approach in Justiniano and Primiceri (2008), who postulate a random walk as the law of motion of the volatilities.

We apply our methodology to the Smets and Wouters (2007) model (henceforth, SW), estimated on the same seven macroeconomic time series used in SW. Our baseline data set starts in 1964Q4 and ends in 2011Q1, but we also consider a subsample ending in 2004Q4 to analyze the extent to which our findings depend on the

[1] The paper is closely related to Chib and Ramamurthy (2011) who in independent and contemporaneous work also propose a similar approach to the one developed here for estimating DSGE models with student-t distributed shocks. Differently from Chib and Ramamurthy (2011), we also introduce time-varying volatilities following the approach in Justiniano and Primiceri (2008). As discussed below, this is important to obtain a proper assessment of the importance of fat tails.

inclusion of the Great Recession in our sample. We use the SW model both because it is a prototypical medium-scale DSGE model, and because its empirical success has been widely documented.[2] Models that fit the data poorly will necessarily have large shocks. We therefore chose a DSGE model that is at the frontier in terms of empirical performance to assess the extent to which macro variables have fat tails.

The motivation for our work arises from evidence like the one displayed in Figure 1. The top panel of Figure 1 shows the time series of the smoothed "discount rate" shocks (in absolute value) from the SW model estimated under Gaussianity. The shocks are normalized, that is, they are expressed in standard deviations units. The solid line is the median, and the dashed lines are the posterior 90% bands. The Figure shows that the size of the shock is between 3.5 and 4 standard deviations in a few occasions, one of which is the recent recession. The probability of observing such large shocks under Gaussianity is very low. In addition to this DSGE model-based evidence, there is work showing that that the unconditional distribution of macro variables is not Gaussian (see Christiano (2007) for pre-Great Recession evidence, and Ascari et al. (2012) for more recent work).

Simply staring at standardized smoothed shocks may not necessarily be the best approach for determining the importance of fat tails, however. First, these shocks are obtained under the counterfactual assumption of Gaussianity. Second, this approach does not provide any quantitative estimate of the fatness of the tails. Third, it is important to disentangle the relative contribution of fat tails from that of (slow moving) time-varying volatility. The bottom panel of Figure 1, which shows the evolution of the smoothed monetary policy shocks estimated under Gaussianity (again, normalized, and in absolute value), provides a case in point: The clustering of large shocks in the late 70s and 80s is quite evident. In general, studying the importance of fat tails only from looking at the kurtosis in the unconditional distribution of either macro variables directly (as in Ascari et al. (2012)) or shocks can be mislead-

[2] The forecasting performance of the SW model was found to be competitive in terms of accuracy relative to private forecasters and reduced form models not only during the Great Moderation period (see Smets and Wouters (2007) and Edge and Gürkaynak (2010)), but also including data for the Great Recession (Del Negro and Schorfheide (2012)).

ing, because the evidence against Gaussianity can be due to low frequency changes in volatility.

Conversely, the presence of large shocks may potentially distort the assessment of low frequency movements in volatility. Imagine estimating a model with only slow moving time variation in volatility, but no fat tails, in presence of shocks that fit the pattern shown in the top panel of Figure 1. As the stochastic volatility will try to fit the squared residuals, such model will produce a time series of volatilities peaking around 1980, and then again during the Great Recession. Put it differently, very large shocks may be interpreted as permanent changes in volatility, when they may be simply rare realizations from a process with a time invariant distribution. For instance, the extent to which the Great Recession can be interpreted as a permanent rise in macroeconomic volatility depends on whether we allow for rare large shocks.

Finally, we expect that the evidence provided in this paper will be further motivation for the study of non linear models. This is for two reasons. First, since shocks have fat tails, linearity may be a poor approximation. Second, non-linearities may explain away the fat tails: what we capture as large rare shocks may be Gaussian shocks whose effect is amplified through a non-linear propagation mechanism. Assessing whether this is the case will be an important line of research. In fact, the extent to which non linearities can alleviate the need for fat tailed shocks in order to explain business cycles can possibly become an additional metric for evaluating their usefulness.

Our findings are the following. We provide strong evidence that the Gaussianity assumption in DSGE models is counterfactual, even after allowing for low frequency changes in the volatility of shocks. Such strong evidence may be surprising considering that our sample only consists of macro variables. Part of this evidence arises from comparing the fit of different specifications using Bayesian marginal likelihoods, and part from the posterior estimates of the degrees of freedom of the Student-t distribution. In terms of fit, we find that if we were to consider only fat tails or stochastic volatility, but not both, the fit of the model is highest, and by far, when we choose the former. Most importantly, if we allow for Student-t in addition to stochastic volatility, the fit improves considerably.

We consider different prior means for the degrees of freedom of the Student-t distribution, capturing different views on the extent to which shocks have fat tails, as well as different degrees of tightness for those priors. We find that whenever the priors on the degrees of freedom are informative, the marginal likelihood always favors the lowest prior mean (fatter tails). When the priors are less informative, the marginal likelihoods tend to be very similar across models, not surprisingly. Less informative priors tend to perform better than more informative ones, highlighting the heterogeneity across shocks in terms of the properties of the distribution.

The posterior estimates of the Student-t distribution's degrees of freedom for some shocks change quite dramatically when we allow for time varying volatility, as previewed earlier, while for other shocks these estimates barely change. The posterior estimates are generally not very sensitive to the prior mean, however, indicating that the likelihood is in many cases quite informative. We can cluster the shocks in the model into three broad categories. Shocks to productivity, the households discount rate, the marginal efficiency of investment, and the wage markup all have fat tails, even in the case with stochastic volatility. Conversely, shocks to government expenditures and to price markups have posterior means for the degrees of freedom that are somewhat high, indicating that their distribution is not far from Gaussian, whether or not we allow for stochastic volatility. Finally, the degrees of freedom for monetary policy shocks are estimated to be extremely low degrees of freedom in the case with constant volatility. When we allow for stochastic volatility then the posterior distribution shifts dramatically toward higher values. These results suggest that for some shocks fat tails is indeed the correct way to model the underlying stochastic process and stochastic volatility does not play a significant role, while for other types of shock the reverse is true.

Whenever we exclude the Great Recession from the sample the results are nearly identical, both in terms of marginal likelihood comparisons and posterior estimates of the degrees of freedom across shocks. This suggests that the evidence in favor of rare large shocks is not confined to the Great Recession. In order to evaluate the importance of accounting for fat tails in the study of the business cycle, we consider a counterfactual experiment in which we shut down the fat tails, so that we recreate the counterfactual path of the economy in the sample absent "rare shocks". We show that in this case all recessions in the sample would then have roughly the same magnitudes, as the Great Recession would have been "just" a run-of-the-mill recession.

Finally, we show that allowing for fat tails will change the inference about slow moving stochastic volatility. We reevaluate the evidence in favor of the Great Moderation hypothesis discussed for example in Justiniano and Primiceri (2008). We find that when we consider Student-t shocks the reduction in the volatility of output and other variables is still substantial, but the magnitude is quite a bit smaller. Likewise, we show that the evidence in favor of a permanent increase in volatility following the Great Recession is much weaker when we consider the possibility that shocks have a Student-t distribution.

The next section discusses Bayesian inference. Section 3 describes the model, as well as our set of observables. Section 4 describes the results.

2 Bayesian Inference

The first part of the section describes the estimation of a DSGE model with both Student-t distributed shocks and time-varying volatilities. The Gibbs sampler combines the algorithm proposed by Geweke (1993)'s for a linear model with Student-t distributed shocks (see also Geweke (1994), and Geweke (2005) for a textbook exposition) with the approach for sampling the parameters of DSGE models with time-varying volatilities discussed in Justiniano and Primiceri (2008). Section A.2 discusses the computation of the marginal likelihood.

The model consists of the standard measurement and transition equations:

$$y_t = Z(\theta)s_t, \tag{1}$$

$$s_{t+1} = T(\theta)s_t + R(\theta)\varepsilon_t, \tag{2}$$

for $t = 1, .., T$, where y_t, s_t, and ε_t are $n \times 1$, $k \times 1$, and $\bar{q} \times 1$ vector of observables, states, and shocks, respectively. Call $p(\theta)$ the prior on the vector of DSGE model parameters θ. We assume that:

$$\varepsilon_{q,t} = \sigma_{q,t} \tilde{h}_{q,t}^{-1/2} \eta_{q,t}, \text{ all } q, t, \tag{3}$$

where

$$\eta_{q,t} \sim \mathcal{N}(0,1), \text{ i.i.d. across } q, \ t, \tag{4}$$

$$\lambda_q \tilde{h}_{q,t} \sim \chi^2(\lambda_q), \text{ i.i.d. across } q, \ t. \tag{5}$$

For the prior on the parameters λ_q we assume a gamma distributions with parameters $\underline{\lambda}/\underline{\nu}$ and $\underline{\nu}$:

$$p(\lambda_q|\underline{\lambda}, \underline{\nu}) = \frac{(\underline{\lambda}/\underline{\nu})^{-\underline{\nu}}}{\Gamma(\underline{\nu})} \lambda_q^{\underline{\nu}-1} exp(-\underline{\nu}\frac{\lambda_q}{\underline{\lambda}}), \text{ i.i.d. across } q. \tag{6}$$

where $\underline{\lambda}$ is the mean and $\underline{\nu}$ is the number of degrees of freedom (Geweke (2005) assumes a Gamma with one degree of freedom).

Define

$$\tilde{\sigma}_{q,t} = \log\left(\sigma_{q,t}/\sigma_q\right), \tag{7}$$

where the parameters $\sigma_{1:\bar{q}}$ (the non-time varying component of the shock variances) are included in the vector of DSGE parameters θ. We assume that the $\tilde{\sigma}_{q,t}$ follows an autoregressive process:

$$\tilde{\sigma}_{q,t} = \rho_q \tilde{\sigma}_{q,t-1} + \zeta_{q,t}, \ \zeta_{q,t} \sim \mathcal{N}(0, \omega_q^2), \text{ i.i.d. across } q, \ t. \tag{8}$$

The prior distribution for ω_q^2 is an inverse gamma $\mathcal{IG}(\nu_\omega/2, \nu_\omega \underline{\omega}^2/2)$, that is:

$$p(\omega_q^2|\nu_\omega, \underline{\omega}^2) = \frac{\left(\nu_\omega \underline{\omega}^2/2\right)^{\frac{\nu_\omega}{2}}}{\Gamma(\nu_\omega/2)} (\omega_q^2)^{-\frac{\nu_\omega}{2}-1} exp\left[-\frac{\nu_\omega \underline{\omega}^2}{2\omega_q^2}\right], \text{ i.i.d. across } q. \tag{9}$$

We consider two types of priors for ρ_q:

$$p(\rho_q|\omega_q^2) = \begin{cases} 1 & \text{SV-UR} \\ \mathcal{N}(\bar{\rho}, \omega_q^2 \bar{v}_\rho)\mathcal{I}(\rho_q), \text{ i.i.d. across } q, \ \mathcal{I}(\rho_q) = \begin{cases} 1 \text{ if } |\rho_q| < 1 \\ 0 \text{ otherwise,} \end{cases} & \text{SV-S} \end{cases} \tag{10}$$

In the SV-UR case $\tilde{\sigma}_{q,t}$ follows a random walk as in Justiniano and Primiceri (2008), while in the SV-S it follows a stationary process as in Fernández-Villaverde and Rubio-Ramírez (2007). In both cases the $\sigma_{q,t}$ process is very persistent: in the SV-UR case the persistence is wired into the assumed law of motion for $\tilde{\sigma}_{q,t}$, while in the SV-AR case it is enforced by choosing the hyperparameters $\bar{\rho}$ and $\bar{\sigma}_\rho$ in such a way that the prior for ρ_q puts most mass on high values of ρ_q. As a consequence, $\sigma_{q,t}$ and $\tilde{h}_{q,t}$ play very different roles in (3): $\sigma_{q,t}$ allows for slow-moving trends in volatility, while $\tilde{h}_{q,t}$ allows for large shocks. Finally, to close the model we make the following distributional assumptions on the initial conditions $\tilde{\sigma}_{q,0},\ q = 1, .., \bar{q}$:

$$p(\tilde{\sigma}_{q,0}|\rho_q, \omega_q^2) = \begin{cases} 0 & \text{SV-UR} \\ \mathcal{N}(0, \omega_q^2/(1-\rho_q^2)), \text{ i.i.d. across } q & \text{SV-S} \end{cases} \tag{11}$$

where the restriction under the SV-UR case is needed to obtain identification. In the stationary case we have assumed that $\tilde{\sigma}_{q,0}$ is drawn from the ergodic distribution.

2.1 The Gibbs-Sampler

The joint distribution of data and unobservables (parameters and latent variables) is given by:

$$\begin{aligned} &p(y_{1:T}|s_{1:T}, \theta)p(s_{1:T}|\varepsilon_{1:T}, \theta)p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta)p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) \\ &\qquad p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2)p(\lambda_{1:\bar{q}})p(\rho_{1:\bar{q}}|\omega_{1:\bar{q}}^2)p(\omega_{1:\bar{q}}^2)p(\theta), \end{aligned} \tag{12}$$

where $p(y_{1:T}|s_{1:T}, \theta)$ and $p(s_{1:T}|\varepsilon_{1:T}, \theta)$ come from the measurement and transition equation, respectively, $p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta)$ obtains from (3) and (4):

$$p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) \propto \prod_{q=1}^{\bar{q}} \left(\prod_{t=1}^{T} \tilde{h}_{q,t}^{-1/2} \sigma_{q,t} \right) exp \left[- \sum_{t=1}^{T} \tilde{h}_{q,t} \varepsilon_{q,t}^2 / 2\sigma_{q,t}^2 \right], \tag{13}$$

$p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})$ obtains from (5)

$$p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) = \prod_{q=1}^{\bar{q}} \prod_{t=1}^{T} \left(2^{\lambda_q/2} \Gamma(\lambda_q/2) \right)^{-1} \lambda_q^{\lambda_q/2} \tilde{h}_{q,t}^{(\lambda_q - 2)/2} \exp(-\lambda_q \tilde{h}_{q,t}/2), \tag{14}$$

$p(\tilde{\sigma}_{1:T}|\omega^2_{1:\bar{q}})$ obtains from expression (8) and (11):

$$p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}},\omega^2_{1:\bar{q}}) \propto \prod_{q=1}^{\bar{q}} (\omega_q^2)^{-(T-1)/2} exp\left[-\sum_{t=2}^{T}(\tilde{\sigma}_{q,t}-\rho_q\tilde{\sigma}_{q,t-1})^2/2\omega_q^2\right] p(\tilde{\sigma}_{q,1}|\rho_q,\omega_q^2), \tag{15}$$

where

$$p(\tilde{\sigma}_{q,1}|\rho_q,\omega_q^2) \propto \begin{cases} (\omega_q^2)^{-1/2} exp\left(-\frac{\tilde{\sigma}^2_{q,1}}{2\omega_q^2}\right), & \text{SV-UR} \\ (\omega_q^2(1-\rho_q^2))^{-1/2} exp\left(-\frac{\tilde{\sigma}^2_{q,1}}{2\omega_q^2(1-\rho_q^2)}\right). & \text{SV-S} \end{cases} \tag{16}$$

Finally, $p(\lambda_{1:\bar{q}}) = \prod_{q=1}^{\bar{q}} p(\lambda_q|\underline{\lambda})$, $p(\omega^2_{1:\bar{q}}) = \prod_{q=1}^{\bar{q}} p(\omega_q^2|\nu,\underline{\omega}^2)$.

The sampler consists of six blocks.

(1) Draw from $p(\theta, s_{1:T}, \varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}}, y_{1:T})$. This is accomplished in two steps:

(1.1) Draw from the marginal $p(\theta|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}}, y_{1:T})$, where

$$\begin{aligned} &p(\theta|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}}, y_{1:T}) \\ &\qquad \propto \left[\int p(y_{1:T}|s_{1:T},\theta)p(s_{1:T}|\varepsilon_{1:T},\theta)p(\varepsilon_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta)\cdot d(s_{1:T},\varepsilon_{1:T})\right] p(\theta) \\ &\qquad = p(y_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta)p(\theta) \end{aligned} \tag{17}$$

where

$$p(y_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta) = \int p(y_{1:T}|s_{1:T},\theta)p(s_{1:T}|\varepsilon_{1:T},\theta)p(\varepsilon_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta)\cdot d(s_{1:T},\varepsilon_{1:T})$$

is computed using the Kalman filter with (1) as the measurement equation and (2) as transition equation, with

$$\varepsilon_t|\tilde{h}_{1:T},\tilde{\sigma}_{1:T} \sim \mathcal{N}(0,\Delta_t), \tag{18}$$

where Δ_t is a $\bar{q}\times\bar{q}$ diagonal matrices with $\sigma^2_{q,t}\cdot\tilde{h}^{-1}_{q,t}$ on the diagonal. The draw is obtained from a Metropolis-Hastings step.

(1.2) Draw from the conditional $p(s_{1:T}, \varepsilon_{1:T}|\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}}, y_{1:T})$. This is accomplished using the simulation smoother of Durbin and Koopman (2002).

(2) Draw from $p(\tilde{h}_{1:T}|\theta, s_{1:T}, \varepsilon_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}}, y_{1:T})$. This is accomplished by drawing from

$$p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta)p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) \propto \prod_{q=1}^{\bar{q}} \prod_{t=1}^{T} \tilde{h}_{q,t}^{(\lambda_q-1)/2} \exp(-\left[\lambda_q + \varepsilon^2_{q,t}/\sigma^2_{q,t}\right] \tilde{h}_{q,t}/2),$$

which implies

$$\left[\lambda_q + \varepsilon^2_{q,t}/\sigma^2_{q,t}\right] \tilde{h}_{q,t}|\theta, \varepsilon_{1:T}, \tilde{\sigma}_{1:T}, \lambda_q \sim \chi^2(\lambda_q + 1).$$

(3) Draw from $p(\lambda_{1:\bar{q}}|\tilde{h}_{1:T}, \theta, s_{1:T}, \varepsilon_{1:T}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}}, y_{1:T})$. This is accomplished by drawing from

$$\begin{aligned} p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\lambda_{1:\bar{q}}) \quad \propto \quad & \textstyle\prod_{q=1}^{\bar{q}} \left((\underline{\lambda}/\underline{\nu})^{\underline{\nu}}\,\Gamma(\underline{\nu})\right)^{-1} [2^{\lambda_q/2}\Gamma(\lambda_q/2)]^{-T} \lambda_q^{T\lambda_q/2+\underline{\nu}-1} \\ & \left(\textstyle\prod_{t=1}^{T} \tilde{h}_{q,t}^{(\lambda_q-2)/2}\right) \exp\left[-\left(\frac{\underline{\nu}}{\underline{\lambda}} + \frac{1}{2}\textstyle\sum_{t=1}^{T} \tilde{h}_{q,t}\right)\lambda_q\right]. \end{aligned}$$

This is a non-standard distribution, hence the draw is obtained from a Metropolis-Hastings step.

(4) Draw from $p(\tilde{\sigma}_{1:T}|\theta, s_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}}, y_{1:T})$. This is accomplished by drawing from

$$p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta)p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}}, \omega^2_{1:\bar{q}})$$

using the algorithm developed by Kim et al. (1998), which we briefly describe in appendix A.3.

(5) Draw from $p(\omega^2_{1:\bar{q}}, \rho_{1:\bar{q}}|\tilde{\sigma}_{1:T}, \theta, s_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}, \lambda_{1:\bar{q}}, y_{1:T})$ using

$$\begin{aligned} & p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}}, \omega^2_{1:\bar{q}})p(\omega^2_{1:\bar{q}})p(\rho_{1:\bar{q}}|\omega^2_{1:\bar{q}}) \propto \\ & \prod_{q=1}^{\bar{q}} (\omega^2_q)^{-\frac{\nu+T-1}{2}-1} exp\left[-\frac{\nu\underline{\omega}^2 + \sum_{t=2}^{T}(\tilde{\sigma}_{q,t} - \rho_q\tilde{\sigma}_{q,t-1})^2}{2\omega^2_q}\right] p(\tilde{\sigma}_{q,1}|\rho_q, \omega^2_q)p(\rho_q|\omega^2_q), \end{aligned} \tag{19}$$

where $p(\tilde{\sigma}_{q,1}|\rho_q, \omega^2_q)$ is given by equation (16). In the SV-UR case ρ_q is fixed to 1, and we can draw ω^2_q from:

$$\omega^2_q|\tilde{\sigma}_{1:T}, \dots \sim \mathcal{IG}\left(\frac{\nu+T}{2}, \frac{1}{2}\left(\nu\underline{\omega}^2 + \sum_{t=2}^{T}(\tilde{\sigma}_{q,t} - \tilde{\sigma}_{q,t-1})^2 + \tilde{\sigma}^2_{q,1}\right)\right), \text{i.i.d. across } q.$$

In the SV-S case the joint posterior of ρ_q, ω_q^2 is non-standard because of the likelihood of the first observation $p(\tilde{\sigma}_1|\rho_q, \omega_q^2)$. We therefore use the Metropolis-Hastings step proposed by Chib and Greenberg (1994). Specifically, we use as proposal density the standard Normal-Inverted Gamma distribution, that is,

$$\begin{aligned} \omega_q^2|\tilde{\sigma}_{1:T}, \ldots \quad & \sim \mathcal{IG}\left(\tfrac{\nu+T-1}{2}, \tfrac{1}{2}\left(\nu\underline{\omega}^2 + \textstyle\sum_{t=2}^T \tilde{\sigma}_{q,t}^2 + \bar{v}_\rho^{-1}\bar{\rho}^2 - \hat{V}_q^{-1}\hat{\rho}_q^2\right)\right), \\ \rho_q|\omega_q^2, \tilde{\sigma}_{1:T}, \ldots \quad & \sim \mathcal{N}\left(\hat{\rho}_q, \omega_q^2\hat{V}_q\right), \text{i.i.d. across } q, \end{aligned}$$

where $\hat{\rho}_q = \hat{V}_q\left(\bar{v}_\rho^{-1}\bar{\rho} + \sum_{t=2}^T \tilde{\sigma}_{q,t}\tilde{\sigma}_{q,t-1}\right)$, $\hat{V}_q = (\bar{v}_\rho^{-1} + \sum_{t=2}^T \tilde{\sigma}_{q,t-1}^2)^{-1}$. We then accept/reject this draw using the proposal density and the acceptance ratio $\frac{p(\tilde{\sigma}_1, \rho_q^{(*)}, \omega_q^{2\ (*)})\mathcal{I}(\rho_q^{(*)})}{p(\tilde{\sigma}_1, \rho_q^{(j-1)}, \omega_q^{2\ (j-1)})\mathcal{I}(\rho_q^{(j-1)})}$, with $(\rho^{(j-1)}, \omega_q^{2\ (j-1)})$ and $(\rho^{(*)}, \omega_q^{2\ (*)})$ being the draw at the $(j-1)^{\text{th}}$ iteration and the proposed draw, respectively.

3 The DSGE Model

The model considered is the one used in Smets and Wouters (2007), which is based on earlier work by Christiano et al. (2005) and Smets and Wouters (2003). It is a medium-scale DSGE model, which augments the standard neoclassical stochastic growth model by nominal price and wage rigidities as well as habit formation in consumption and investment adjustment costs.

3.1 The Smets-Wouters Model

We begin by briefly describing the log-linearized equilibrium conditions of the Smets and Wouters (2007) model. We deviate from Smets and Wouters (2007) in that we detrend the non-stationary model variables by a stochastic rather than a deterministic trend. This approach makes it possible to express almost all equilibrium conditions in a way that encompasses both the trend-stationary total factor productivity process in Smets and Wouters (2007), as well as the case where technology follows a unit root process. We refer to the model presented in this section as SW model. Let $\tilde{z}_t$ be the linearly detrended log productivity process which follows the autoregressive law of motion

$$\tilde{z}_t = \rho_z \tilde{z}_{t-1} + \sigma_z \varepsilon_{z,t}. \tag{20}$$

We detrend all non stationary variables by $Z_t = e^{\gamma t + \frac{1}{1-\alpha}\tilde{z}_t}$, where γ is the steady state growth rate of the economy. The growth rate of Z_t in deviations from γ, denoted by z_t, follows the process:

$$z_t = \ln(Z_t/Z_{t-1}) - \gamma = \frac{1}{1-\alpha}(\rho_z - 1)\tilde{z}_{t-1} + \frac{1}{1-\alpha}\sigma_z \epsilon_{z,t}. \tag{21}$$

All variables in the subsequent equations are expressed in log deviations from their non-stochastic steady state. Steady state values are denoted by $*$-subscripts and steady state formulas are provided in a Technical Appendix (available upon request). The consumption Euler equation takes the form:

$$\begin{aligned} c_t = & -\frac{(1-he^{-\gamma})}{\sigma_c(1+he^{-\gamma})}\left(R_t - I\!E_t[\pi_{t+1}] + b_t\right) + \frac{he^{-\gamma}}{(1+he^{-\gamma})}\left(c_{t-1} - z_t\right) \\ & + \frac{1}{(1+he^{-\gamma})} I\!E_t\left[c_{t+1} + z_{t+1}\right] + \frac{(\sigma_c - 1)}{\sigma_c(1+he^{-\gamma})}\frac{w_* L_*}{c_*}\left(L_t - I\!E_t[L_{t+1}]\right), \end{aligned} \tag{22}$$

where c_t is consumption, L_t is labor supply, R_t is the nominal interest rate, and π_t is inflation. The exogenous process b_t drives a wedge between the intertemporal ratio of the marginal utility of consumption and the riskless real return $R_t - I\!E_t[\pi_{t+1}]$, and follows an AR(1) process with parameters ρ_b and σ_b. The parameters σ_c and h capture the relative degree of risk aversion and the degree of habit persistence in the utility function, respectively. The next condition follows from the optimality condition for the capital producers, and expresses the relationship between the value of capital in terms of consumption q_t^k and the level of investment i_t measured in terms of consumption goods:

$$\begin{aligned} q_t^k = S'' e^{2\gamma}(1+\beta e^{(1-\sigma_c)\gamma})\Big(& i_t - \frac{1}{1+\beta e^{(1-\sigma_c)\gamma}}\left(i_{t-1} - z_t\right) \\ & - \frac{\beta e^{(1-\sigma_c)\gamma}}{1+\beta e^{(1-\sigma_c)\gamma}} I\!E_t\left[i_{t+1} + z_{t+1}\right] - \mu_t\Big), \end{aligned} \tag{23}$$

which is affected by both investment adjustment cost (S'' is the second derivative of the adjustment cost function) and by μ_t, an exogenous process called "marginal efficiency of investment" that affects the rate of transformation between consumption and installed capital (see Greenwood et al. (1998)). The latter, called $\bar{k}_t$, indeed evolves as

$$\bar{k}_t = \left(1 - \frac{i_*}{\bar{k}_*}\right)\left(\bar{k}_{t-1} - z_t\right) + \frac{i_*}{\bar{k}_*} i_t + \frac{i_*}{\bar{k}_*} S'' e^{2\gamma}(1 + \beta e^{(1-\sigma_c)\gamma})\mu_t, \tag{24}$$

where $i_*/\bar{k}_*$ is the steady state ratio of investment to capital. μ_t follows an AR(1) process with parameters ρ_μ and σ_μ. The parameter β captures the intertemporal discount rate in the utility function of the households. The arbitrage condition between the return to capital and the riskless rate is:

$$\frac{r^k_*}{r^k_* + (1-\delta)} I\!E_t[r^k_{t+1}] + \frac{1-\delta}{r^k_* + (1-\delta)} I\!E_t[q^k_{t+1}] - q^k_t = R_t + b_t - I\!E_t[\pi_{t+1}], \tag{25}$$

where r^k_t is the rental rate of capital, r^k_* its steady state value, and δ the depreciation rate. Capital is subject to variable capacity utilization u_t. The relationship between $\bar{k}_t$ and the amount of capital effectively rented out to firms k_t is

$$k_t \;=\; u_t - z_t + \bar{k}_{t-1}. \tag{26}$$

The optimality condition determining the rate of utilization is given by

$$\frac{1-\psi}{\psi} r^k_t = u_t, \tag{27}$$

where ψ captures the utilization costs in terms of foregone consumption. From the optimality conditions of goods producers it follows that all firms have the same capital-labor ratio:

$$k_t \;=\; w_t - r^k_t + L_t. \tag{28}$$

Real marginal costs for firms are given by

$$mc_t \;=\; w_t + \alpha L_t - \alpha k_t, \tag{29}$$

where α is the income share of capital (after paying markups and fixed costs) in the production function.

All of the equations so far maintain the same form whether technology has a unit root or is trend stationary. A few small differences arise for the following two equilibrium conditions. The production function is:

$$y_t \;=\; \Phi_p\left(\alpha k_t + (1-\alpha) L_t\right) + \mathcal{I}\{\rho_z < 1\}(\Phi_p - 1)\frac{1}{1-\alpha}\tilde{z}_t, \tag{30}$$

under trend stationarity. The last term $(\Phi_p - 1)\frac{1}{1-\alpha}\tilde{z}_t$ drops out if technology has a stochastic trend, because in this case one has to assume that the fixed costs are proportional to the trend. Similarly, the resource constraint is:

$$y_t = g_t + \frac{c_*}{y_*}c_t + \frac{i_*}{y_*}i_t + \frac{r_*^k k_*}{y_*}u_t - \mathcal{I}\{\rho_z < 1\}\frac{1}{1-\alpha}\tilde{z}_t, . \tag{31}$$

The term $-\frac{1}{1-\alpha}\tilde{z}_t$ disappears if technology follows a unit root process. Government spending g_t is assumed to follow the exogenous process:

$$g_t = \rho_g g_{t-1} + \sigma_g \varepsilon_{g,t} + \eta_{gz}\sigma_z\varepsilon_{z,t}.$$

Finally, the price and wage Phillips curves are, respectively:

$$\pi_t = \frac{(1-\zeta_p\beta e^{(1-\sigma_c)\gamma})(1-\zeta_p)}{(1+\iota_p\beta e^{(1-\sigma_c)\gamma})\zeta_p((\Phi_p-1)\epsilon_p+1)}mc_t + \frac{\iota_p}{1+\iota_p\beta e^{(1-\sigma_c)\gamma}}\pi_{t-1} + \frac{\beta e^{(1-\sigma_c)\gamma}}{1+\iota_p\beta e^{(1-\sigma_c)\gamma}}\mathbb{E}_t[\pi_{t+1}] + \lambda_{f,t}, \tag{32}$$

and

$$w_t = \frac{(1-\zeta_w\beta e^{(1-\sigma_c)\gamma})(1-\zeta_w)}{(1+\beta e^{(1-\sigma_c)\gamma})\zeta_w((\lambda_w-1)\epsilon_w+1)}\left(w_t^h - w_t\right) - \frac{1+\iota_w\beta e^{(1-\sigma_c)\gamma}}{1+\beta e^{(1-\sigma_c)\gamma}}\pi_t + \frac{1}{1+\beta e^{(1-\sigma_c)\gamma}}\left(w_{t-1} - z_t - \iota_w\pi_{t-1}\right) + \frac{\beta e^{(1-\sigma_c)\gamma}}{1+\beta e^{(1-\sigma_c)\gamma}}\mathbb{E}_t\left[w_{t+1} + z_{t+1} + \pi_{t+1}\right] + \lambda_{w,t}, \tag{33}$$

where ζ_p, ι_p, and ϵ_p are the Calvo parameter, the degree of indexation, and the curvature parameters in the Kimball aggregator for prices, and ζ_w, ι_w, and ϵ_w are the corresponding parameters for wages. The variable w_t^h corresponds to the household's marginal rate of substitution between consumption and labor, and is given by:

$$w_t^h = \frac{1}{1-he^{-\gamma}}\left(c_t - he^{-\gamma}c_{t-1} + he^{-\gamma}z_t\right) + \nu_l L_t, \tag{34}$$

where ν_l characterizes the curvature of the disutility of labor (and would equal the inverse of the Frisch elasticity in absence of wage rigidities). The mark-ups $\lambda_{f,t}$ and $\lambda_{w,t}$ follow exogenous ARMA(1,1) processes

$$\lambda_{f,t} = \rho_{\lambda_f}\lambda_{f,t-1} + \sigma_{\lambda_f}\varepsilon_{\lambda_f,t} + \eta_{\lambda_f}\sigma_{\lambda_f}\varepsilon_{\lambda_f,t-1}, \text{ and}$$

$$\lambda_{w,t} = \rho_{\lambda_w} \lambda_{w,t-1} + \sigma_{\lambda_w} \varepsilon_{\lambda_w,t} + \eta_{\lambda_w} \sigma_{\lambda_w} \varepsilon_{\lambda_w,t-1},$$

respectively. Last, the monetary authority follows a generalized feedback rule:

$$\begin{aligned} R_t &= \rho_R R_{t-1} + (1-\rho_R)\left(\psi_1 \pi_t + \psi_2 (y_t - y_t^f)\right) \\ &+ \psi_3 \left((y_t - y_t^f) - (y_{t-1} - y_{t-1}^f)\right) + r_t^m, \end{aligned} \tag{35}$$

where the flexible price/wage output y_t^f obtains from solving the version of the model without nominal rigidities (that is, Equations (22) through (31) and (34)), and the residual r_t^m follows an AR(1) process with parameters ρ_{r^m} and σ_{r^m}.

3.2 Observation Equation, Data, and Priors

We use the method in Sims (2002) to solve the log-linear approximation of the DSGE model. We collect all the DSGE model parameters in the vector θ, stack the structural shocks in the vector ϵ_t, and derive a state-space representation for our vector of observables y_t, which is composed of the transition equation:

$$s_t = \mathcal{T}(\theta) s_{t-1} + \mathcal{R}(\theta)\epsilon_t, \tag{36}$$

which summarizes the evolution of the states s_t, and of the measurement equations:

$$y_t = \mathcal{Z}(\theta) s_t + \mathcal{D}(\theta), \tag{37}$$

which maps the states onto the vector of observables y_t, where $\mathcal{D}(\theta)$ represents the vector of steady state values for these observables. Specifically, the SW model is estimated based on seven quarterly macroeconomic time series. The measurement equations for real output, consumption, investment, and real wage growth, hours, inflation, interest rates and long-run inflation expectations are given by:

$$\begin{aligned} \textit{Output growth} &= \gamma + 100\,(y_t - y_{t-1} + z_t) \\ \textit{Consumption growth} &= \gamma + 100\,(c_t - c_{t-1} + z_t) \\ \textit{Investment growth} &= \gamma + 100\,(i_t - i_{t-1} + z_t) \\ \textit{Real Wage growth} &= \gamma + 100\,(w_t - w_{t-1} + z_t)\ , \\ \textit{Hours} &= \bar{l} + 100 l_t \\ \textit{Inflation} &= \pi_* + 100\pi_t \\ \textit{FFR} &= R_* + 100 R_t \end{aligned} \tag{38}$$

where all variables are measured in percent, π_* and R_* measure the steady state level of net inflation and short term nominal interest rates, respectively and $\bar{l}$ captures the mean of hours (this variable is measured as an index).

Appendix A.1 provides further details on the data. In our benchmark specification we use data from 1964Q4 to 2011Q1. In Section 4.4 we consider a shorter sample in which end the sample in 2004Q4, so that we exclude the Great Recession.

Table 1 shows the priors for the DSGE model parameters. These are based on the priors used in Smets and Wouters (2007). Table 2 presents the posterior mean for the standard parameters of the DSGE model for the different specifications with Gaussian shocks, Student-t distributed shocks, stochastic volatility, and both.

4 Results

4.1 Evidence against Gaussianity

In the introduction we showed some qualitative evidence of the presence of rare large shocks based on smoothed shocks extracted from standard Gaussian estimation. In this section we provide quantitative evidence in favor of fat-tailed shocks. First, we assess the improvement in fit obtained by allowing for Student-t distributed shocks. Table 3 shows the log marginal likelihood for models with different assumptions on the shocks distribution. We consider four different combinations: i) Gaussian shocks with constant volatility (baseline), ii) time-varying volatility, iii) Student-t distributed shocks but constant volatility, and iv) both Student-t shocks and time-variation in volatility. We consider specifications with different prior means for the degrees of freedom λ of the Student-t distribution. The three priors capture three different views of the world on the importance of fat tails. The first prior, $\underline{\lambda} = 15$, captures the view that the world is not quite Gaussian, but not too far from Gaussianity either. The second prior, $\underline{\lambda} = 9$, embodies the idea that the world is quite far from Gaussian, yet not too extreme. The last prior, $\underline{\lambda} = 6$, stands for a model with quite heavy tails. The following table provides a quantitative feel for what these different means imply in terms of the model's ability to generate fat tailed shocks (with $\lambda = \infty$ being the Gaussian case). Specifically, the table shows the number of shocks larger (in abs. value) than x standard deviations per 200 periods, which is the size of our sample.

λ, x:	3	4	5
∞	.54	.012	$1e^{-4}$
15	1.79	.23	.03
9	2.99	.62	.15
6	4.80	1.42	.49

We consider different degrees of tightness in the prior on λ (the tightness is proportional to the degrees of freedom $\underline{\nu}$ of the Gamma distribution characterizing the prior for λ, see equation 6). In the middle panel of Table 3 we use four degree of freedom in the Gamma distribution. These priors are relatively informative and the difference in the views of the world hence quite stark. In the lower panel we consider the same three prior means but now with only one degrees of freedom. These priors are quite flat, and hence the difference among them in not as stark.

The table shows that the data strongly favor Student-t distributed shocks. The fit also increases when we consider stochastic volatility instead of constant volatility. If we were to consider only Student-t distributed shocks or stochastic volatility, but not both, the data seems to prefer the former, as the marginal likelihood increases by 150 log points in this case, compared to an increase of 67 log points for the latter. But the fit is best when we consider both of these features, indicating that both low frequency changes in volatility and fat tails are features of the data. Whenever the prior on the degrees of freedom of the Student-t distribution is (relatively) tight, the fit increases the lower the prior mean for the degrees of freedom. Whenever the prior is loose, the difference in fit across priors naturally shrinks, and all three priors perform approximately the same, although the $\underline{\lambda} = 15$ prior performs slightly better than the others. Importantly, as we show below, with a loose prior the posterior estimates for the degrees does not differ much across the different priors (and is very low for some shocks).

Marginal likelihoods are difficult to compute, especially for these models (see the appendix). Therefore we also show the posterior distribution of λ obtained under less informative prior. Specifically, Table 4 shows the posterior mean and 90% bands for the degrees of freedom for each shock in the specifications with and without stochastic volatility. In the case without stochastic volatility component the posterior means are mostly below or relatively close to 6, with the exception of the price markup. Furthermore, the posterior 90% bands around the median are all fairly tight, with the 95th percentile well below 15 (again with the exception of the price markup, and only in the case with prior mean of 9 or 15). Once we introduce stochastic volatility, the posterior means for some of the shocks increase considerably. This finding points to the fact that for a proper assessment of the importance of fat tails we need to allow for low-frequency time variation in volatility. In fact, for some of the shocks, the stochastic volatility explains the data fairly well without the need for fat tails. However, for several of the shocks the posterior means are still fairly low and with tight bands, confirming that adding stochastic volatility is not enough to fully explain the data, and fat tails still play an important role. Note that the location of the prior ($\underline{\lambda}$ equal to 6, 9, or 15) matters little for the shocks with a low posterior mean.

The shocks with the lowest posterior degrees of freedom are those affecting the discount rate (b), TFP productivity (z), the marginal efficiency of investment (μ) and wage markup (λ_w). The shock related to monetary policy (deviations from the systematic response to the economy, r^m) exhibit a posterior distribution for the degrees of freedom concentrated in very low levels when we do not allow for stochastic volatility, but once we allow for stochastic volatility the posterior shifts to values well above the prior mean. This suggests that the shocks related to monetary policy exhibit a pattern that is better explained by stochastic volatility, consistent with a period in the late 70s and early 80s in which shocks related to monetary policy were especially volatile.

Figure 2 shows the smoothed shocks and the "tamed" version of these shocks (that is, the counterfactual shocks after shutting down the Student-t component –

see below) for both the discount rate and policy shocks for the estimation without stochastic volatility. On the left plots we consider the absolute value of the shock histories, much like in Figure 1, while on the right side we consider the absolute value of the shocks once we shut down the Student-t component. Consistent with the above analysis, the right side plots look much more consistent with a Gaussian distribution than those on the left side, confirming the role of fat tails. However, notice that for the policy shock there is still a cluster of higher variance innovations in the late 70s and early 80s, suggestive that stochastic volatility has an important role in this type of shocks.

4.2 Do Fat Tails Matter for the Macroeconomy?

We have shown that quite a few important shocks in the SW model have fat tails – their estimated degrees of freedom are low. But what does this mean in terms of business cycle fluctuations? This section tries to provide a quantitative answer to this question. We do so by performing a counterfactual experiment. Recall that from equation (3)

$$\varepsilon_{q,t} = \sigma_{q,t}\tilde{h}_{q,t}^{-1/2}\eta_{q,t}.$$

We compute the posterior distribution of $\varepsilon_{q,t}$ (the smoothed shocks) and $\tilde{h}_{q,t}$. Next, we purge $\varepsilon_{q,t}$ from the Student-t component, that is, we compute

$$\tilde{\varepsilon}_{q,t} = \sigma_{q,t}\eta_{q,t},$$

and compute counterfactual histories had the shocks been $\tilde{\varepsilon}_{q,t}$ instead of $\varepsilon_{q,t}$. All these counterfactuals are computed for the best fitting model – that with stochastic volatility and, for the Student-t, a prior for λ centered at 6 with one degree of freedom.

The left panels of Figure 3 shows these counterfactual histories for output, consumption growth, and hours (solid black lines are the actual data). The right panel uses actual and counterfactual histories to compute a rolling window standard deviation, where each window contains the prior 20 quarters as well as the following 20 quarters, for a total of 41 quarters. These rolling window standard deviations are commonly used measured of time-variation in the volatility of the series. The difference between actual and counterfactual standard deviations measures the extent to which the change in volatility is accounted for by rare shocks.[3] For all plots the pink solid lines are the median counterfactual paths while the pink dashed lines represent the 90% bands.

The left panels show that rare shocks seem to account for a non negligible part of fluctuations in two of the variables of interest, output and consumption growth. For output growth, the Student-t component accounted for about half of the contraction in output growth in the Great Recession. If the fat tail component were absent the Great Recession would be no worse than the average recession in the sample. In general, without rare shocks all recessions would be of roughly the same magnitude. Since the fat tails accounted for a significant part of the large movements in output, the rolling window standard deviation shown in the top right panel shows that the Student-t component explains a non-negligible part of changes in the realized volatility in the data. One can interpret this evidence as saying that the 70s and early 80s were more volatile than the Great Moderation period partly because rare shocks took place. For example at the peak of the volatility in 1978, the data standard deviation is about 1.25, but once we shut down the Student-t component it drops to 1.05, which is a reduction of about 16% in volatility. If we turn to the evolution of consumption growth (middle panels) the results are somewhat stronger, while this is not the case for hours (bottom panel), although in the Great Recession the Student-t component accounted for about 2 percentage points decline in hours worked.

4.3 Student-t Shocks and Inference about Time-Variation in Volatility

It is clear from the marginal likelihood analysis that the stochastic volatility has a role in explaining the data. Relative to the simple constant volatility and Gaussian

[3] The distribution of $\tilde{h}_{q,t}$ is non-time varying. However, since large shocks occur rarely, they may account for changes in the rolling window volatility.

shocks it improves fit by as much as 67 log points, consistent with the work of Justiniano and Primiceri (2008). Once we add fat tails to the model the additional improvement due to stochastic volatility is smaller, but is nevertheless there. We now discuss the extent to which accounting for fat tails makes us reevaluate the magnitude of low frequency changes in volatility.

Figure 4 shows the stochastic volatility component for the discount rate and policy shocks. On the left panel we show these for the estimation with stochastic volatility and otherwise Gaussian shocks, while on the right panels we consider both stochastic volatility and Student-t shocks. The black lines correspond to the absolute value of the shocks, as in Figure 1, and the red lines correspond to the evolution of the stochastic volatility component, $\sigma_q \sigma_{q,t}$. Solid lines correspond to the median and dot/dashed lines to the 90% bands around the median. The Figure shows that when we ignore fat tails fluctuations in volatility for the discount rate shock are sizable: volatility drops by half between the early 1980s and the mid 1990s, and the decline is significant. Once we account for the fat tails movements in the volatility are more subdued, and less significant. Recall that the degrees of freedom for the discount rate shock are still quite low even after allowing for stochastic volatility. For the policy shock, in the lower panels, there is hardly any difference in the stochastic volatility component whether or not we allow for Student-t shocks – consistent with the fact that the estimated degrees of freedom of the Student-t component for this shock is high once we allow for stochastic volatility (shown in Table 4).

An implication of stochastic volatility is that the model-implied variance of the endogenous variables changes over time (see Figure 5 of Justiniano and Primiceri (2008)). Figure 5 shows the model-implied volatility of output and consumption growth, as measured by the unconditional standard deviation of the series computed keeping the estimated $\sigma_{q,t}$ constant for each t. In the top panel, the red line shows this measure for the estimation with stochastic volatility but Gaussian shocks, while the black lines show this volatility for the estimation with both stochastic volatility and Student-t components. Solid line is the posterior median and the dashed lines correspond to the 90% bands around the median. For both variables the model-implied volatility is mostly lower when we do not account for fat tails, but it is not simply a parallel shift. The difference seems to be more substantial in the periods with low volatility, than in the periods with high volatility. Indeed, in the case of consumption growth the unconditional volatility path is very similar for the two estimations in the first part of the sample up to 1981.

We now ask whether the evidence concerning the Great Moderation is influenced by the presence of Student-t shocks. The middle panel of Figure 5 shows the posterior histogram of the ratio of the volatility in 1981 relative to the volatility in 1994 for the three variables. Whether or not we allow for Student-t shocks most of the probability mass is above one, confirming that output volatility fell between 1981 and 1994. The magnitude of the fall depends on whether or not we allow for fat-tailed shocks, however. The posterior distribution for the ratio of the output growth volatility in 1981 relative to the volatility in 1994 in the estimation with Student-t shocks is shifted to the left relative to the case with Gaussian shocks (indeed the median is 1.7 in the former, compared to 2.5 in the latter). This same pattern is also evident for the consumption growth, shown in the the right part of the middle panel. The bottom panel of Figure 5 shows that the evidence in favor of a permanent increase in volatility following the Great Recession is much weaker when we consider the possibility that shocks have a Student-t distribution.

As a result of the Great Recession there has been an increase in volatility in many macroeconomic variables since 2008, as measured by the rolling window standard deviations shown in Figure 3. To what extent does this increase reflect a permanent increase in the volatility of the underlying shocks, that is, the end of the Great Moderation? The bottom panel of Figure 5 shows that the evidence in favor of a permanent increase in volatility following the Great Recession is much weaker when we consider the possibility that shocks have a Student-t distribution. The panel shows the posterior histogram of the ratio of volatility in 2011 over the volatility in 2005. The probability of the ratio being below one shifts from 4.6% to 12% in the case of output, and from 8.2% to 21% in the case of consumption growth.

4.4 Sub-Sample Excluding the Great Recession

This paper was motivated in large part by the Great Recession. But do the results depend on the Great Recession being part of the estimation sample? In order to address this question we estimate the model for the sub-sample ending in the fourth quarter of 2004 (the same sample used in Justiniano and Primiceri (2008)). Table 5 shows the marginal likelihood for all the specifications considered above but estimated on the shorter sub-sample. The results displayed in Table 5 are aligned with the results for the full sample. Namely, adding a Student-t component improves the fit, whether we also consider stochastic volatility or not. If we were to have only Student-t or stochastic volatility, but not both, the data strongly favors the Student-t specification (the marginal likelihood is higher by 72 log-points). Interestingly, the lower the prior mean for the degrees of freedom the higher the marginal likelihood is. Importantly, this is true regardless of how tight the prior is or whether we also include a stochastic volatility component. The posterior means of the degrees of freedom of the Student-t distribution (not shown) are also in line with those presented in Table 4. These results are not surprising given our previous discussion of the counterfactual analysis shown in Figure 3. The contribution of the Student-t component to explain the path of the series shown or their volatility is visible throughout the sample. The Great Recession is just one instance of the evidence in favor rare large shocks, while there are several such instances in the 1970s, and so our inference about whether we should include the Student-t component is not much dependent on this last part of the sample.

5 Conclusions

We provide strong evidence that the Gaussianity assumption in DSGE models is counterfactual, even after allowing for low frequency changes in the volatility of shocks. It is important to point out a number of caveats regarding our analysis. For one, in the current draft we allow for excess Kurtosis but not for skewness. The shocks plots in in Figure 1 make it plain that most large shocks occur during recessions. A recent paper by Müller (2011) describes some of the dangers associated with departures from Gaussianity when the alternative shock distribution is also misspecified. Second, we allow for permanent (random walk) and i.i.d (Student-t distribution) changes in the variance of the shocks. These assumptions are convenient, but also extreme. We will consider relaxing these assumptions (and yet maintain identification) in future drafts.

References

An, S. and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models–Rejoinder," *Econometric Reviews*, 26, 211 – 219.

Aruoba, S. B. and F. Schorfheide (2010): "Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-offs," *American Economic Journal: Macroeconomics*, forthcoming.

Ascari, G., G. Fagiolo, and A. Roventini (2012): "Fat Tails Distributions and Business Cycle Models," *Manuscript, OFCE-Science-Po.*

Chib, S. and E. Greenberg (1994): "Bayes inferences in regression models with ARMA(p,q) errors," *Journal of Econometrics*, 64, 183–206.

Chib, S. and S. Ramamurthy (2011): "DSGE Models with Student-t errors," *mimeo, Washington University in St. Louis.*

Christiano, L., M. Eichenbaum, and C. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

Christiano, L. J. (2007): "Comment on 'On the Fit of New Keynesian Models' by Del Negro ,Schorfheide, Smets ,Wouters," *Journal of Business and Economic Statistics*, 25, 143–151.

Del Negro, M. and S. Eusepi (2011): "Fitting Observed Ination Expectations," *Journal of Economic Dynamics and Control*, 35, 2105–2131.

Del Negro, M. and F. Schorfheide (2012): "DSGE Model-Based Forecasting," *FRBNY Working Paper.*

Durbin, J. and S. J. Koopman (2002): "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89, 603–616.

Edge, R. and R. Gürkaynak (2010): "How Useful Are Estimated DSGE Model

Forecasts for Central Bankers," *Brookings Papers of Economic Activity*, forthcoming.

Erceg, C. J. and A. T. Levin (2003): "Imperfect Credibility and Inflation Persistence," *Journal of Monetary Economics*, 50, 915–944.

Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74, 1059–1087.

Geweke, J. (1993): "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics,*, 8, S19–S40.

——— (1994): "Priors for Macroeconomic Time Series and Their Application," *Econometric Theory*, 10, 609–632.

——— (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18, 1–126.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*, Wiley.

Greenwood, J., Z. Hercovitz, and P. Krusell (1998): "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review*, 87, 342–36.

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

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–393.

Liu, Z., D. F. Waggoner, and T. Zha (2011): "Sources of macroeconomic fluctuations: A regime-switching DSGE approach," *Quantitative Economics*, 2, 251–301.

Müller, U. K. (2011): "Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix," *Mimeo, Princeton University.*

Sims, C. A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20(1-2), 1–20.

Smets, F. and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123 – 1175.

——— (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586 – 606.

A Appendix

A.1 Data

The data set is obtained from Haver Analytics (Haver mnemonics are in italics). We compile observations for the variables that appear in the measurement equation (38). Real GDP (GDPC), the GDP price deflator (GDPDEF), nominal personal consumption expenditures (PCEC), and nominal fixed private investment (FPI) are constructed at a quarterly frequency by the Bureau of Economic Analysis (BEA), and are included in the National Income and Product Accounts (NIPA).

Average weekly hours of production and nonsupervisory employees for total private industries (PRS85006023), civilian employment (CE16OV), and civilian noninstitutional population (LNSINDEX) are produced by the Bureau of Labor Statistics (BLS) at the monthly frequency. The first of these series is obtained from the Establishment Survey, and the remaining from the Household Survey. Both surveys are released in the BLS Employment Situation Summary (ESS). Since our models are estimated on quarterly data, we take averages of the monthly data. Compensation per hour for the nonfarm business sector (PRS85006103) is obtained from the Labor Productivity and Costs (LPC) release, and produced by the BLS at the quarterly frequency. Last, the federal funds rate is obtained from the Federal Reserve Board's H.15 release at the business day frequency, and is not revised. We take quarterly averages of the annualized daily data.

All data are transformed following Smets and Wouters (2007). Specifically:

$$\begin{aligned}
\textit{Output growth} &= LN((GDPC)/LNSINDEX) * 100 \\
\textit{Consumption growth} &= LN((PCEC/GDPDEF)/LNSINDEX) * 100 \\
\textit{Investment growth} &= LN((FPI/GDPDEF)/LNSINDEX) * 100 \\
\textit{Real Wage growth} &= LN(PRS85006103/GDPDEF) * 100 \\
\textit{Hours} &= LN((PRS85006023 * CE16OV/100)/LNSINDEX) * 100 \\
\textit{Inflation} &= LN(GDPDEF/GDPDEF(-1)) * 100 \\
\textit{FFR} &= FEDERAL\ FUNDS\ RATE/4
\end{aligned}$$

A.2 Marginal likelihood

The marginal likelihood is the marginal probability of the observed data, and is computed as the integral of (12) with respect to the unobserved parameters and latent variables:

$$\begin{aligned} p(y_{1:T}) &= \int p(y_{1:T}|s_{1:T},\theta)p(s_{1:T}|\varepsilon_{1:T},\theta)p(\varepsilon_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta) \\ &\qquad p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\tilde{\sigma}_{1:T}|\omega^2_{1:\bar{q}})p(\lambda_{1:\bar{q}})p(\omega^2_{1:\bar{q}})p(\theta) \\ &\qquad d(s_{1:T},\varepsilon_{1:T},\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}},\theta), \\ &= \int p(y_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta)p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\tilde{\sigma}_{1:T}|\omega^2_{1:\bar{q}}) \\ &\qquad p(\lambda_{1:\bar{q}})p(\omega^2_{1:\bar{q}})p(\theta)d(\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}},\theta) \end{aligned} \tag{39}$$

where the quantity

$$\begin{aligned} p(y_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta) &= \int p(y_{1:T}|s_{1:T},\theta)p(s_{1:T}|\varepsilon_{1:T},\theta) \\ &\qquad p(\varepsilon_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta)\cdot d(s_{1:T},\varepsilon_{1:T}) \end{aligned}$$

is computed at step 1a of the Gibb-sampler described above.

We obtain the marginal likelihood using Geweke (1999)'s modified harmonic mean method. If $f(\theta,\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}})$ is any distribution with support contained in the support of the posterior density such that

$$\int f(\theta,\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}})\cdot d(\theta,\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}}) = 1,$$

it follows from the definition of the posterior density that:

$$\begin{aligned} \frac{1}{p(y_{1:T})} &= \int \frac{f(\theta,\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}})}{p(y_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta)p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\tilde{\sigma}_{1:T}|\omega^2_{1:\bar{q}})p(\lambda_{1:\bar{q}})p(\omega^2_{1:\bar{q}})p(\theta)} \\ &\qquad p(\theta,\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}}|y_{1:T})\cdot d(\theta,\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}}) \end{aligned}$$

We follow Justiniano and Primiceri (2008) in choosing

$$f(\theta,\tilde{h}_{1:T}) = f(\theta)\cdot p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\tilde{\sigma}_{1:T}|\omega^2_{1:\bar{q}})p(\lambda_{1:\bar{q}})p(\omega^2_{1:\bar{q}}), \tag{40}$$

where $f(\theta)$ is a truncate multivariate distribution as proposed by Geweke (1999). Hence we approximate the marginal likelihood as:

$$\hat{p}(y_{1:T}) = \left[\frac{1}{n_{sim}}\sum_{j=1}^{n_{sim}}\frac{f(\theta^j)}{p(y_{1:T}|\tilde{h}^j_{1:T},\tilde{\sigma}^j_{1:T},\theta^j)p(\theta^j)}\right]^{-1} \tag{41}$$

where θ^j, $\tilde{h}^j_{1:T}$, and $\tilde{\sigma}^j_{1:T}$ are draws from the posterior distribution, and n_{sim} is the total number of draws. We are aware of the problems with (40), namely that it does not ensure that the random variable

$$\frac{f(\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}})}{p(y_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta)p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\tilde{\sigma}_{1:T}|\omega^2_{1:\bar{q}})p(\lambda_{1:\bar{q}})p(\omega^2_{1:\bar{q}})p(\theta)}$$

has finite variance. Nonetheless, like Justiniano and Primiceri (2008) we found that this method delivers very similar results across different chains.

A.3 Drawing the stochastic volatilities

We draw the stochastic volatilities using the procedure in Kim et al. (1998), which we briefly describe. Taking squares and then logs of (3) one obtains:

$$\varepsilon^*_{q,t} = 2\tilde{\sigma}_{q,t} + \eta^*_{q,t} \tag{42}$$

where

$$\varepsilon^*_{q,t} = \log(\sigma_q^{-2}\tilde{h}_{q,t}\varepsilon^2_{q,t} + c), \tag{43}$$

$c = .001$ being an offset constant, and $\eta^*_{q,t} = \log(\eta^2_{q,t})$. If $\eta^*_{q,t}$ were normally distributed, $\sigma_{q,1:T}$ could be drawn using standard methods for state-space systems. In fact, $\eta^*_{q,t}$ is distributed as a $\log(\chi^2_1)$. Kim et al. (1998) address this problem by approximating the $\log(\chi^2_1)$ with a mixture of normals, that is, expressing the distribution of $\eta^*_{q,t}$ as:

$$p(\eta^*_{q,t}) = \sum_{k=1}^{K} \pi^*_k \mathcal{N}(m^*_k - 1.2704, \nu^{*\,2}_k) \tag{44}$$

The parameters that optimize this approximation, namely $\{\pi^*_k, m^*_k, \nu^*_k\}^K_{k=1}$ and K, are given in Kim et al. (1998). Note that these parameters are independent of the specific application. The mixture of normals can be equivalently expressed as:

$$\eta^*_{q,t}|\varsigma_{q,t} = k \sim \mathcal{N}(m^*_k - 1.2704, \nu^{*\,2}_k),\ Pr(s_{i,t} = k) = \pi^*_k. \tag{45}$$

Hence step (4) of the Gibbs sampler actually consists in two steps:

(4.1) Draw from $p(\varsigma_{1:T}|\tilde{\sigma}_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}, s_{1:T}\lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}}, y_{1:T})$ using (44) for each q. Specifically:

$$Pr\{\varsigma_{q,t} = k|\tilde{\sigma}_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T} \ldots\} \propto \pi^*_k \nu_k^{-1} \exp\left[-\frac{1}{2\nu^{*\,2}_k}(\eta^*_{q,t} - m^*_k + 1.2704)^2\right]. \tag{46}$$

where from (42) $\eta^*_{q,t} = \varepsilon^*_{q,t} - 2\tilde{\sigma}_{q,t}$.

(4.2) Draw from $p(\tilde{\sigma}_{1:T}|\varsigma_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}, s_{1:T}\lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega^2_{1:\bar{q}}, y_{1:T})$ using Durbin and Koopman (2002), where (42) is the measurement equation and (8) is the transition equation.

Note that in principle we should make it explicit that we condition on $\varsigma_{1:T}$ in the other steps of the Gibbs sampler as well. In practice, all other conditional distributions do not depend on $\varsigma_{1:T}$, hence we omit the term for simplicity.

Table 1: Priors for the Medium-Scale Model

	Density	Mean	St. Dev.		Density	Mean	St. Dev.
Policy Parameters							
ψ_1	Normal	1.50	0.25	ρ_R	Beta	0.75	0.10
ψ_2	Normal	0.12	0.05	ρ_{r^m}	Beta	0.50	0.20
ψ_3	Normal	0.12	0.05	σ_{r^m}	InvG	0.10	2.00
Nominal Rigidities Parameters							
ζ_p	Beta	0.50	0.10	ζ_w	Beta	0.50	0.10
Other "Endogenous Propagation and Steady State" Parameters							
α	Normal	0.30	0.05	π^*	Gamma	0.75	0.40
Φ	Normal	1.25	0.12	γ	Normal	0.40	0.10
h	Beta	0.70	0.10	S''	Normal	4.00	1.50
ν_l	Normal	2.00	0.75	σ_c	Normal	1.50	0.37
ι_p	Beta	0.50	0.15	ι_w	Beta	0.50	0.15
r_*	Gamma	0.25	0.10	ψ	Beta	0.50	0.15
ρs, σs, and ηs							
ρ_z	Beta	0.50	0.20	σ_z	InvG	0.10	2.00
ρ_b	Beta	0.50	0.20	σ_b	InvG	0.10	2.00
ρ_{λ_f}	Beta	0.50	0.20	σ_{λ_f}	InvG	0.10	2.00
ρ_{λ_w}	Beta	0.50	0.20	σ_{λ_w}	InvG	0.10	2.00
ρ_μ	Beta	0.50	0.20	σ_μ	InvG	0.10	2.00
ρ_g	Beta	0.50	0.20	σ_g	InvG	0.10	2.00
η_{λ_f}	Beta	0.50	0.20	η_{λ_w}	Beta	0.50	0.20
η_{gz}	Beta	0.50	0.20				

Notes: Note that $\beta = (1/(1 + r_*/100))$. The following parameters are fixed in Smets and Wouters (2007): $\delta = 0.025$, $g_* = 0.18$, $\lambda_w = 1.50$, $\varepsilon_w = 10.0$, and $\varepsilon_p = 10$. The columns "Mean" and "St. Dev." list the means and the standard deviations for Beta, Gamma, and Normal distributions, and the values s and ν for the Inverse Gamma (InvG) distribution, where $p_{\mathcal{IG}}(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region. The prior for $\bar{l}$ is $\mathcal{N}(-45, 5^2)$.

Table 2: Posterior Means of the DSGE Model Parameters

	Baseline	SV	St-t	St-t+SV
α	0.150	0.149	0.132	0.123
α	0.150	0.149	0.132	0.145
ζ_p	0.728	0.771	0.779	0.722
ι_p	0.311	0.366	0.318	0.354
Φ	1.582	1.587	1.546	1.551
S''	4.623	5.588	5.164	4.272
h	0.605	0.603	0.540	0.533
ψ	0.719	0.690	0.833	0.830
ν_l	2.070	2.435	2.366	2.413
ζ_w	0.800	0.824	0.830	0.762
ι_w	0.542	0.503	0.571	0.539
β	0.209	0.209	0.183	0.173
ψ_1	1.965	1.899	1.989	2.107
ψ_2	0.082	0.100	0.088	0.079
ψ_3	0.244	0.201	0.205	0.199
π^*	0.945	1.014	1.007	0.904
σ_c	1.272	1.294	1.293	1.226
ρ	0.834	0.868	0.859	0.846
γ	0.308	0.337	0.340	0.327
$\bar{l}$	-45.437	-44.488	-44.858	-43.753
ρ_g	0.978	0.983	0.991	0.982
ρ_b	0.752	0.800	0.828	0.780
ρ_μ	0.767	0.797	0.812	0.892
ρ_z	0.994	0.990	0.977	0.969
ρ_{λ_f}	0.804	0.796	0.831	0.824
ρ_{λ_w}	0.976	0.949	0.885	0.926
ρ_{rm}	0.152	0.216	0.210	0.219
σ_g	2.888	2.379	0.190	0.054
σ_b	0.125	0.080	0.074	0.052
σ_μ	0.422	0.317	0.226	0.058
σ_z	0.492	0.354	0.504	0.235
σ_{λ_f}	0.164	0.138	0.074	0.054
σ_{λ_w}	0.280	0.210	0.100	0.052
σ_{rm}	0.228	0.126	0.059	0.044
η_{gz}	0.793	0.783	0.765	0.796
η_{λ_f}	0.683	0.698	0.723	0.696
η_{λ_w}	0.939	0.900	0.813	0.825

Notes: We use a prior mean of 6 degrees of freedom for the Student-t distributed component when there is no stochastic volatility, and a prior mean of 15 degrees of freedom when we also include stochastic volatility. The stochastic volatility component assumes a prior mean for the size of the shocks to volatility of $(0.01)^2$.

Table 3: Marginal Likelihoods

	Without Stochastic Volatility	With Stochastic Volatility
Gaussian shocks		
	-1117.8	-1050.9
Student-t distributed shocks, prior with 4 degrees of freedom		
$\underline{\lambda} = 15$	-1004.1	-983.1
$\underline{\lambda} = 9$	-985.9	-969.0
$\underline{\lambda} = 6$	-974.6	-965.5
Student-t distributed shocks, prior with 1 degree of freedom		
$\underline{\lambda} = 15$	-982.9	-962.9
$\underline{\lambda} = 9$	-972.7	-964.3
$\underline{\lambda} = 6$	-968.0	-963.7

Notes: The parameter $\underline{\lambda}$ represents the prior mean for the degrees of freedom in the Student-t distribution.

Table 4: Posterior of the Student's t Degrees of Freedom

	Without Stochastic Volatility			*With Stochastic Volatility*		
	$\underline{\lambda} = 15$	$\underline{\lambda} = 9$	$\underline{\lambda} = 6$	$\underline{\lambda} = 15$	$\underline{\lambda} = 9$	$\underline{\lambda} = 6$
g	8.0	6.7	6.1	16.7	13.7	11.6
	(2.4,14.0)	(2.4,11.2)	(2.4,9.7)	(3.8,30.9)	(4.1,24.1)	(3.9,19.4)
b	4.1	4.0	3.9	6.4	6.2	5.7
	(2.1,6.0)	(2.1,5.8)	(2.2,5.6)	(2.3,10.7)	(2.3,10.0)	(2.4,9.0)
μ	7.4	6.6	6.1	8.3	7.3	6.5
	(2.4,12.6)	(2.4,10.8)	(2.4,10.0)	(2.5,14.7)	(2.5,12.5)	(2.5,10.6)
z	4.2	4.0	3.9	4.6	4.3	4.0
	(1.8,6.4)	(1.8,6.0)	(1.9,5.9)	(1.7,7.6)	(1.8,6.7)	(1.8,6.1)
λ_f	10.8	9.0	7.7	19.5	15.2	12.6
	(2.8,20.1)	(2.8,15.4)	(2.8,12.7)	(4.7,35.3)	(4.6,26.2)	(4.4,20.9)
λ_w	7.4	6.8	6.3	6.1	5.8	5.4
	(2.6,12.3)	(2.6,11.1)	(2.5,10.0)	(2.7,9.5)	(2.6,8.9)	(2.6,8.3)
r^m	2.8	2.7	2.7	17.1	13.5	11.1
	(1.6,3.9)	(1.6,3.8)	(1.6,3.7)	(4.0,31.7)	(3.9,23.7)	(3.6,18.6)

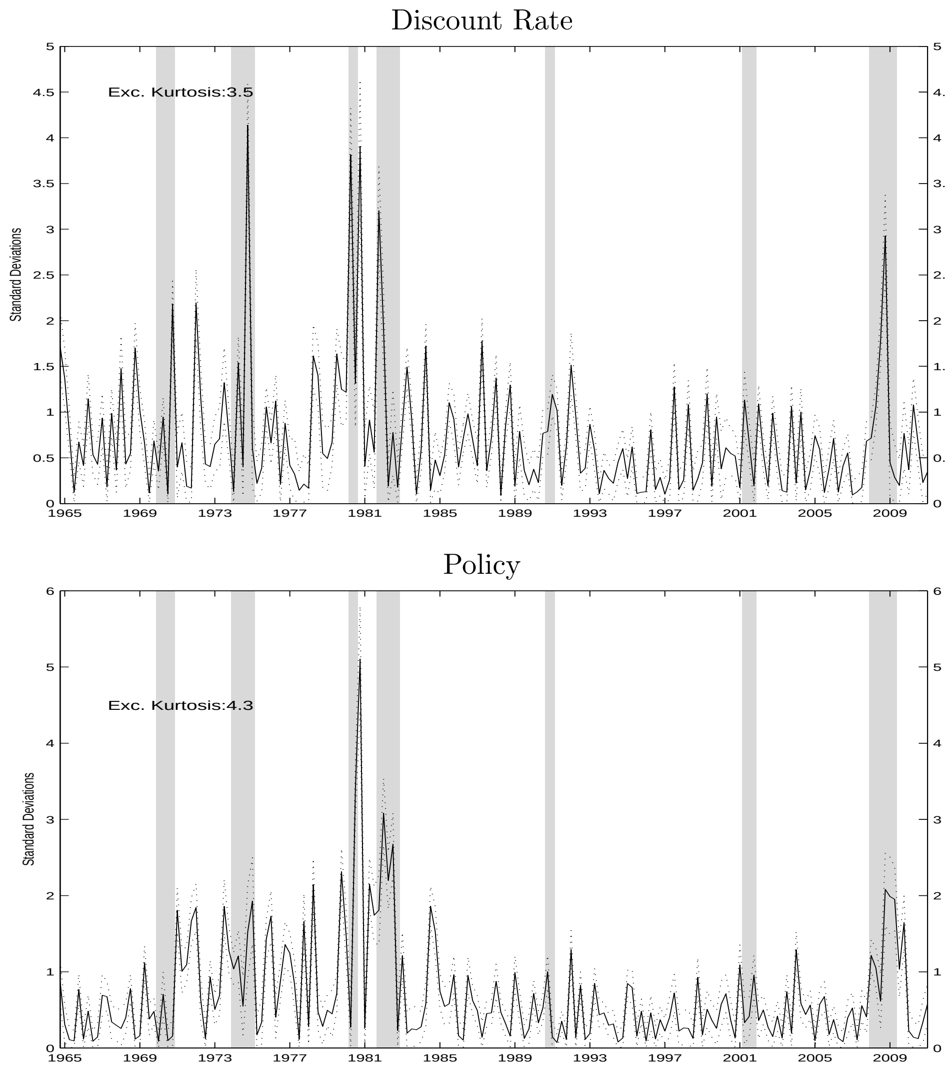
Notes: Numbers shown for the posterior mean and the 90% intervals of the degrees of freedom parameter.

Table 5: Marginal Likelihoods, Sample Ending in 2004Q4

	Constant Volatility	Stochastic Volatility
Gaussian shocks		
	-962.8	-926.7
Student-t distributed shocks, prior with 4 degrees of freedom		
$\underline{\lambda} = 15$	-878.4	-847.9
$\underline{\lambda} = 9$	-866.8	-842.2
$\underline{\lambda} = 6$	-853.9	-835.0
Student-t distributed shocks, prior with 1 degree of freedom		
$\underline{\lambda} = 15$	-860.3	-841.1
$\underline{\lambda} = 9$	-858.6	-837.8
$\underline{\lambda} = 6$	-854.6	-830.7

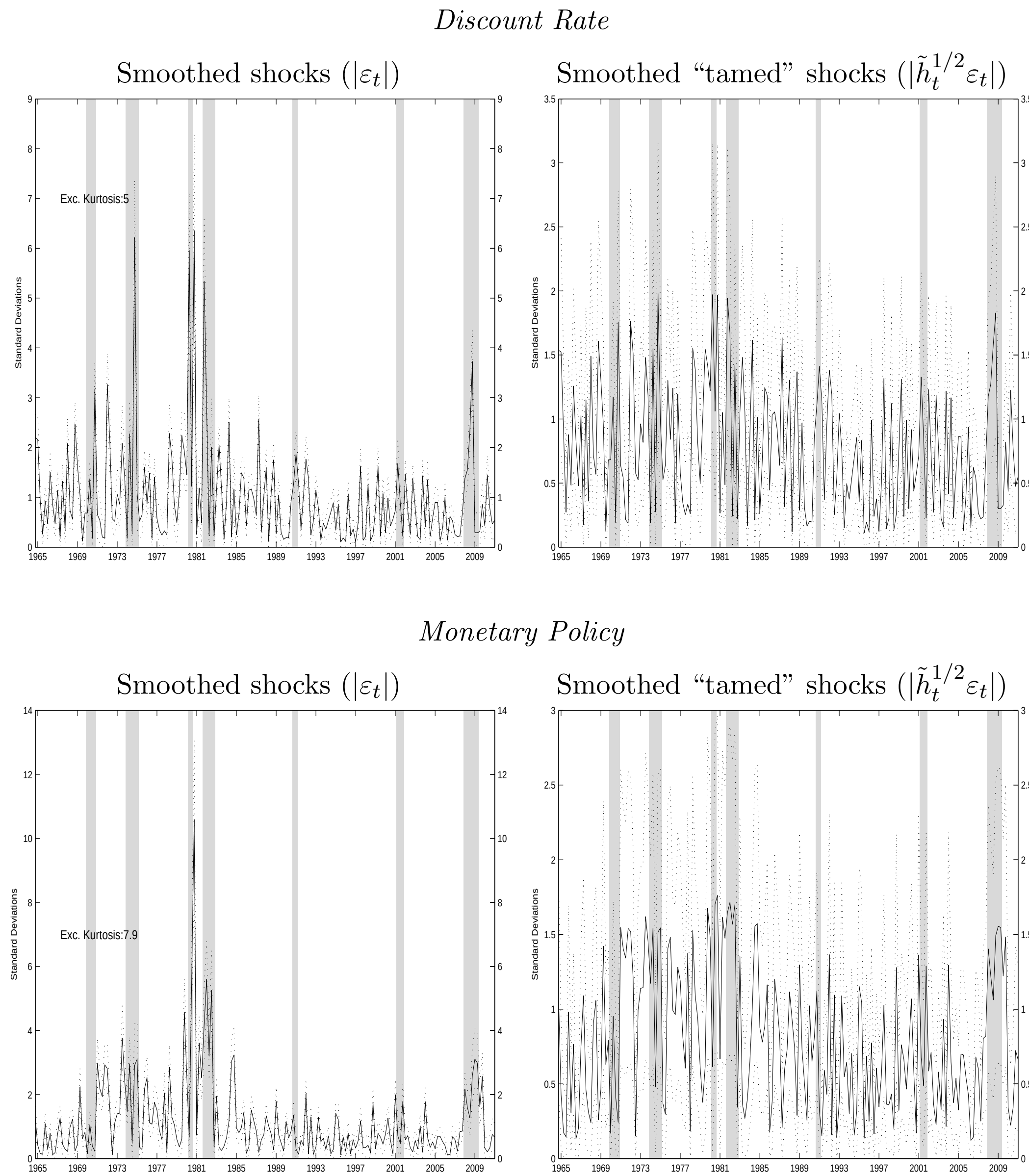
Notes: The parameter $\underline{\lambda}$ represents the prior mean for the degrees of freedom in the Student-t distribution.

Figure 1: Smoothed Shocks under Gaussianity (Absolute Value, Standardized)



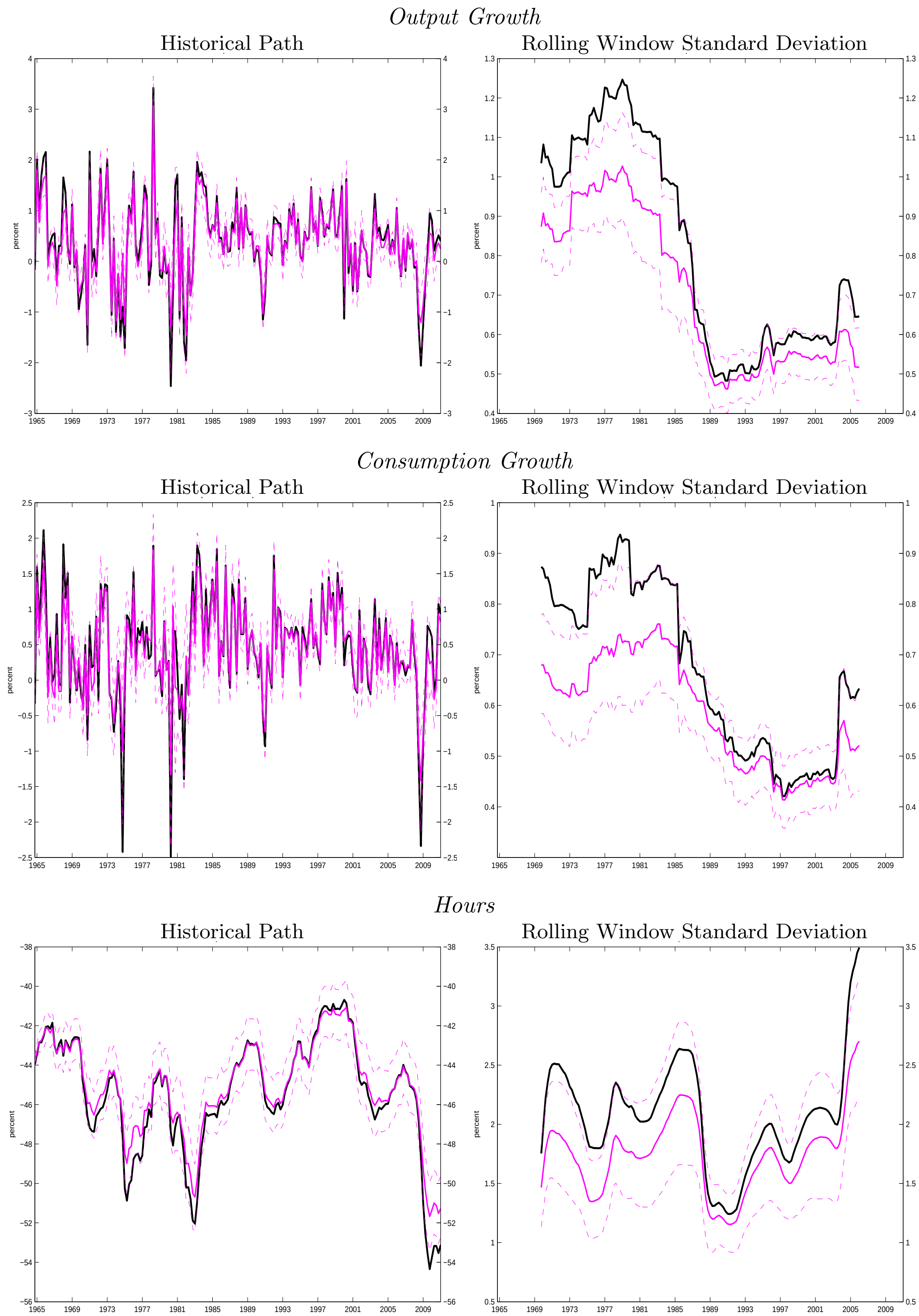
Notes: The solid line is the median, and the dashed lines are the posterior 90% bands. The vertical shaded regions identify NBER recession dates.

Figure 2: Shocks and "Tamed" Shocks (Absolute Value, Standardized)



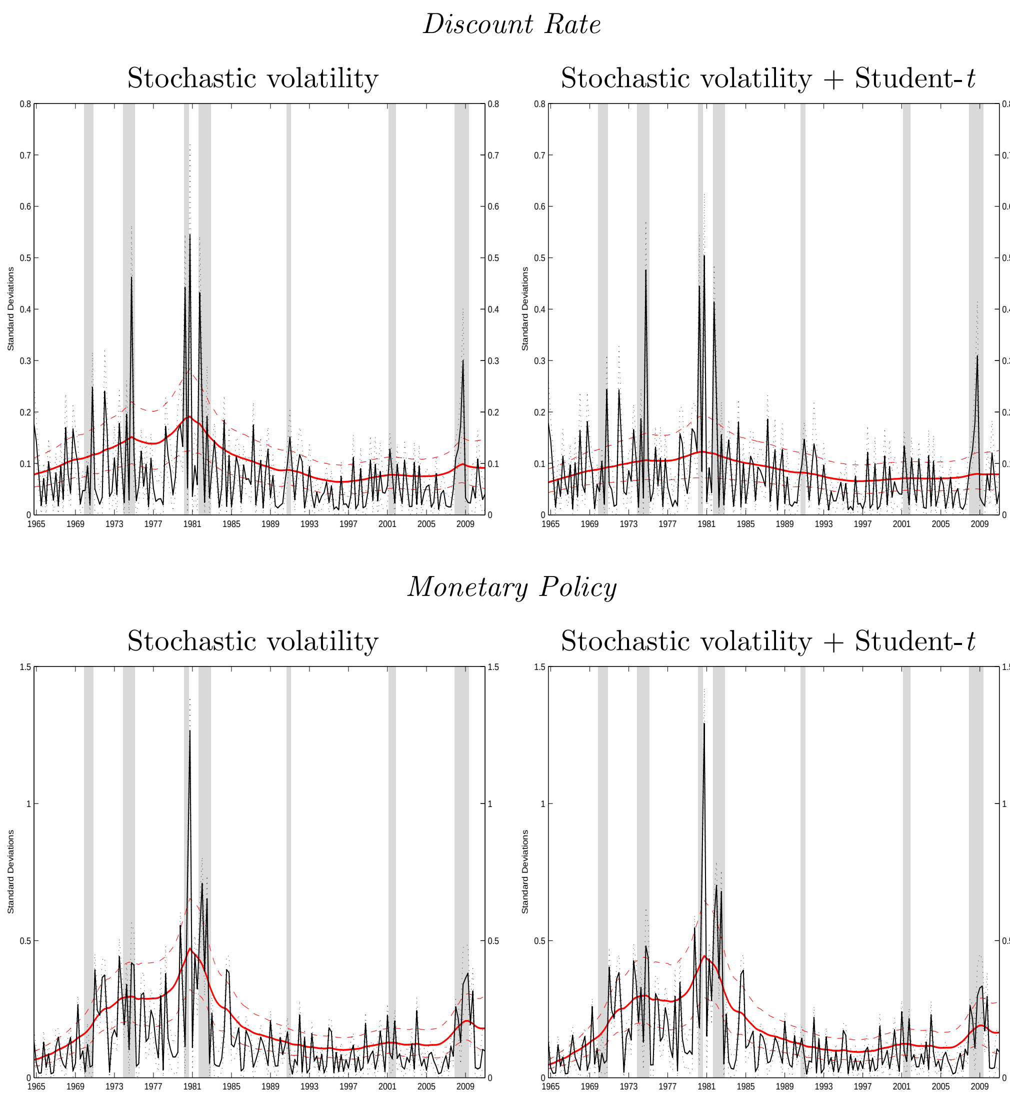
Notes: Estimation with Student-t distribution with $\underline{\lambda} = 6$. The solid line is the median, and the dashed lines are the posterior 90% bands. Shocks are expressed in units of the standard deviation σ_q. The vertical shaded regions identify NBER recession dates.

Figure 3: Counterfactual evolution of output, consumption and hours worked when the Student-t distributed component is turned off, estimation with Student-t distributed shocks and stochastic volatility.



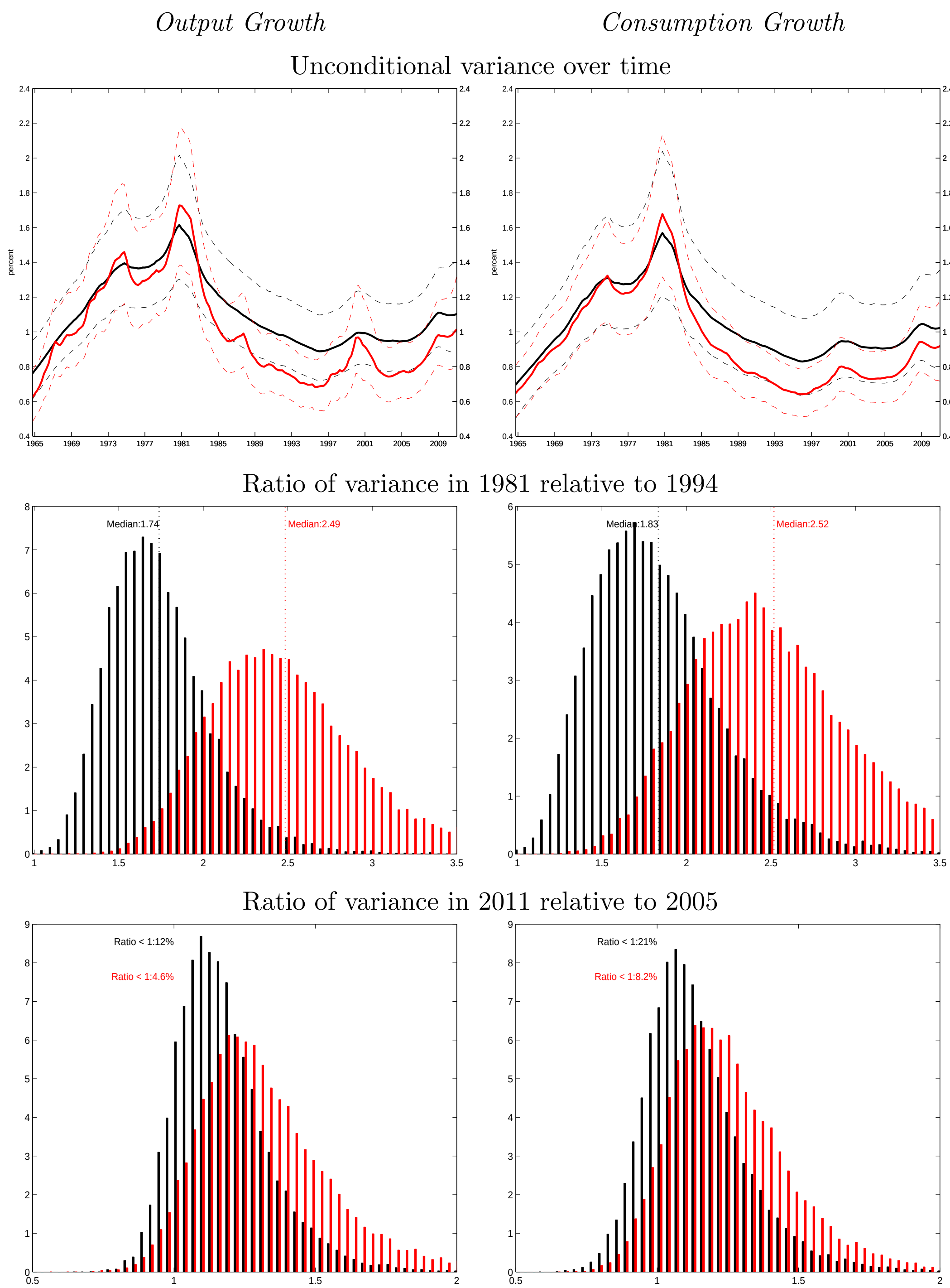
Notes: Black lines are the historical evolution of the variable, and pink lines are the median counterfactual evolution of the same variable if we shut down the Student-t distributed component of all shocks. The rolling window standard deviation uses 20 quarters before and 20 quarters after a given quarter.

Figure 4: Shocks (absolute values) and smoothed stochastic volatility component, $\sigma_q \sigma_{q,t}$



Notes: Estimation with Student-t distribution with $\underline{\lambda} = 15$. The solid line is the median, and the dashed lines are the posterior 90% bands. Black line is the absolute value of the shock, and the red line is the stochastic volatility component.

Figure 5: Time-Variation in the unconditional variance of output and consumption; models estimated with and without the Student-t distributed component.



Notes: Black line in the top panel is the unconditional variance in the estimation with both stochastic volatility and Student-t components, while the red line is the unconditional variance in the estimation with stochastic volatility component only. On the middle panel the black bars correspond to the posterior histogram of the ratio of volatility in 1981 over the variance in 1994 for the estimation with both stochastic volatility and Student-t components, while the red bars are for the estimation with with stochastic volatility component only. The lower panel replicates the same analysis as in the middle panel but for the ratio of volatility in 2011 over the variance in 2005.