Rare Shocks, Great Recessions

Vasco Cúrdia, Marco Del Negro, Daniel Greenwald*

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

May 18, 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 recession 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. This cannot be done using the Gaussian distribution, which is the standard assumption in the DSGE literature. We estimate a linearized DSGE model assuming that shocks are generated from a Student-t distribution, which is designed to capture fat tails. The number of degrees of freedom in the Student-t distribution, which determines the likelihood of observing rare large 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 (as described, for instance, in An and Schorfheide (2007)). In fact, the Gibbs sampler a simple extension of Geweke (1993)'s Gibbs sampler for a linear model to the DSGE framework.[1]

The focus of this paper is twofold. First, we ask whether there is evidence of fat tails, and whether tail events are important for the macroeconomy. Next, we ask whether allowing for Student-t distributed shocks affects the inference about time-variation in volatility. As to the first question, we provide strong evidence that the normality assumption in DSGE models is counterfactual. We show that allowing for Student-t distributed shocks substantially improves the model's fit – as measured by Bayesian marginal likelihood. In fact, the improvement in fit is greater than that obtained by allowing for (random walk) time variation in volatilities. To some extent this is not too surprising given that we had evidence, even before the recent recession, that macro variables are not normally distributed (see Christiano (2007)). Our finding may also not be too surprising to anyone who has computed smoothed shocks obtained from DSGE models estimated under normality. For instance, 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.

top panel of Figure 1 shows the time series of the smoothed "discount rate" shocks (in absolute value) from the Smets-Wouters model estimated under Gaussianity on post-war data. The shocks are normalized, that is, 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 chances of observing such large shocks under Gaussianity is slim.

We furthermore provide an estimate of the fatness of tails – that is, of the degrees of freedom in the Student-t distribution – and to do so shock by shock. We find that shocks that are associated with financial market and labor market frictions exhibit quite low degrees of freedom – that is, very fat tails. Finally, 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. In other words, without these rare events history would have been quite different.

There are of course ways to introduce departures from normality in the shocks distribution other than assuming that shocks are Student-t distributed. Two important papers in the DSGE literature, Justiniano and Primiceri (2008) and Fernández-Villaverde and Rubio-Ramírez (2007), have introduced stochastic volatility in the estimation of DSGE models. In both papers stochastic volatility captures low frequency movements in volatility however, as they focused on the sources of the Great Moderation. In particular, in Justiniano and Primiceri (2008) this is by assumption, as they postulate a random walk as the law of motion of the volatilities. We argue that it is important to allow for both fat tails and low frequency movements in volatility. Ignoring low frequency movements in volatility may bias the results toward finding evidence in favor of fat tails. For example, the bottom panel of Figure 1 shows the evolution of the smoothed monetary policy shocks (again, normalized, and in absolute value) estimated under Gaussianity. Imagine estimating a model with only fat tails but no slow moving time variation in volatility on shocks that fit the pattern shown in the bottom panel. Here the model would interpret all the large shocks in the late 70s and 80s as rare shocks, ignoring the evident clustering of these shocks – that is, the fact that the volatility of monetary policy shocks was low in the sixties, high during the Volcker period, and then low again. Vice versa, we argue that the presence of large shocks may potentially distort the assessment of these low frequency movements. 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. We show that indeed allowing for fat tails will change the inference about slow moving stochastic volatility.

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. We plan to address this issue in future drafts. A recent paper by Müller (2011) describes some of the dangers associated with departures from Gaussianity when the alternative shock distribution is also misspecified. Next, our model is linear, as in most of the estimated DSGE literature – the purpose of the paper is indeed to show that *in linear models* shocks exhibit fat tails. But our results provide evidence that the literature may need to move beyond linear models, following the lead of Fernández-Villaverde and Rubio-Ramírez (2007) and others. 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 are 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.

The next section discusses Bayesian inference. The section first describes the procedure used to estimate a DSGE model with Student-t distributed shocks, and then combines Student-t distributed shocks with time-variation in volatilities. 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:

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

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:

$$p(y_{1:T}|s_{1:T},\theta)p(s_{1:T}|\varepsilon_{1:T},\theta)p(\varepsilon_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta)p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})$$
$$p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}},\omega^2_{1:\bar{q}})p(\lambda_{1:\bar{q}})p(\rho_{1:\bar{q}}|\omega^2_{1:\bar{q}})p(\omega^2_{1:\bar{q}})p(\theta), \quad (12)$$

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):

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

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

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

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

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

where

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

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

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

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

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

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

where $\tilde{\sigma}^{*^2}_{q,1} = \tilde{\sigma}^2_{q,1}$ in the SV-UR case, and $\tilde{\sigma}^{*^2}_{q,1} = \tilde{\sigma}^2_{q,1}/(1-\rho^2_q)$ in the SV-S case. This implies that

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

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

$$p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}}, \omega^2_{1:\bar{q}})p(\rho_{1:\bar{q}}|\omega^2_{1:\bar{q}}) \propto \prod_{q=1}^{\bar{q}} \mathcal{N}(\bar{\rho}_q, \bar{V}_q)\mathcal{I}(\rho_q)p(\tilde{\sigma}_1|\rho_q, \omega^2_q),$$

where $\bar{V}_q = (\bar{v}_\rho^{-1} + \omega_q^{-2}\sum_{t=2}^{T}\tilde{\sigma}^2_{q,t-1})^{-1}$ and $\bar{\rho}_q = \bar{V}_q\left(\bar{v}_\rho^{-1}\bar{\rho} + \omega_q^{-2}\sum_{t=2}^{T}\tilde{\sigma}_{q,t}\tilde{\sigma}_{q,t-1}\right)$. This probability is non-standard because of the likelihood of the first observation $p(\tilde{\sigma}_1|\rho_q, \omega^2_q)$. We therefore use the Metropolis-Hastings step proposed by Chib and Greenberg (1994), where $\mathcal{N}(\bar{\rho}_q, \bar{V}_q)\mathcal{I}(\rho_q)$ is the proposal density and the acceptance ratio simplifies to $\frac{p(\tilde{\sigma}_1|\rho^*_q, \omega^{2\ j}_q)}{p(\tilde{\sigma}_1|\rho^{j-1}_q, \omega^{2\ j}_q)}$, with ρ^{j-1} and ρ^* 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. We further consider an extension in which we add credit frictions to this framework, following the "financial accelerator" model described in Bernanke et al. (1999). The actual implementation of the credit frictions follows closely that of Christiano et al. (2009).

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

$$\tilde{z}_t = \rho_z \tilde{z}_{t-1} + \sigma_z \varepsilon_{z,t}. \tag{19}$$

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:

$$z_t = \ln(Z_t/Z_{t-1}) - \gamma = \frac{1}{1-\alpha}(\rho_z - 1)\tilde{z}_{t-1} + \frac{1}{1-\alpha}\sigma_z \epsilon_{z,t}. \tag{20}$$

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:

$$\begin{aligned} c_t = &-\frac{(1-he^{-\gamma})}{\sigma_c(1+he^{-\gamma})}\left(R_t - \mathbb{E}_t[\pi_{t+1}] + b_t\right) + \frac{he^{-\gamma}}{(1+he^{-\gamma})}\left(c_{t-1} - z_t\right) \\ &+ \frac{1}{(1+he^{-\gamma})}\mathbb{E}_t\left[c_{t+1} + z_{t+1}\right] + \frac{(\sigma_c - 1)}{\sigma_c(1+he^{-\gamma})}\frac{w_* L_*}{c_*}\left(L_t - \mathbb{E}_t[L_{t+1}]\right), \end{aligned} \tag{21}$$

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:

$$q_t^k = S'' e^{2\gamma}(1+\beta e^{(1-\sigma_c)\gamma})\Big(i_t - \frac{1}{1+\beta e^{(1-\sigma_c)\gamma}}(i_{t-1} - z_t) - \frac{\beta e^{(1-\sigma_c)\gamma}}{1+\beta e^{(1-\sigma_c)\gamma}} I\!E_t\left[i_{t+1} + z_{t+1}\right] - \mu_t\Big), \quad (22)$$

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

$$\bar{k}_t = \left(1 - \frac{i_*}{\bar{k}_*}\right)\left(\bar{k}_{t-1} - z_t\right) + \frac{i_*}{\bar{k}_*} i_t + \frac{i_*}{\bar{k}_*} S'' e^{2\gamma}(1+\beta e^{(1-\sigma_c)\gamma})\mu_t, \quad (23)$$

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:

$$\frac{r_*^k}{r_*^k + (1-\delta)} I\!E_t[r_{t+1}^k] + \frac{1-\delta}{r_*^k + (1-\delta)} I\!E_t[q_{t+1}^k] - q_t^k = R_t + b_t - I\!E_t[\pi_{t+1}], \quad (24)$$

where r_t^k 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

$$k_t \;=\; u_t - z_t + \bar{k}_{t-1}. \quad (25)$$

The optimality condition determining the rate of utilization is given by

$$\frac{1-\psi}{\psi} r_t^k = u_t, \tag{26}$$

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:

$$k_t \;=\; w_t - r_t^k + L_t. \tag{27}$$

Real marginal costs for firms are given by

$$mc_t \;=\; w_t + \alpha L_t - \alpha k_t, \tag{28}$$

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:

$$y_t \;=\; \Phi_p \left(\alpha k_t + (1-\alpha) L_t\right) + \mathcal{I}\{\rho_z < 1\}(\Phi_p - 1)\frac{1}{1-\alpha}\tilde{z}_t, \tag{29}$$

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:

$$y_t = g_t + \frac{c_*}{y_*}c_t + \frac{i_*}{y_*}i_t + \frac{r_*^k k_*}{y_*}u_t - \mathcal{I}\{\rho_z < 1\}\frac{1}{1-\alpha}\tilde{z}_t, . \tag{30}$$

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:

$$\begin{aligned}\pi_t = {} & \frac{(1-\zeta_p \beta e^{(1-\sigma_c)\gamma})(1-\zeta_p)}{(1+\iota_p \beta e^{(1-\sigma_c)\gamma})\zeta_p((\Phi_p - 1)\epsilon_p + 1)} mc_t \\ & + \frac{\iota_p}{1+\iota_p\beta e^{(1-\sigma_c)\gamma}}\pi_{t-1} + \frac{\beta e^{(1-\sigma_c)\gamma}}{1+\iota_p\beta e^{(1-\sigma_c)\gamma}} I\!\!E_t[\pi_{t+1}] + \lambda_{f,t},\end{aligned} \tag{31}$$

and

$$
\begin{aligned}
w_t = \frac{(1-\zeta_w \beta e^{(1-\sigma_c)\gamma})(1-\zeta_w)}{(1+\beta e^{(1-\sigma_c)\gamma})\zeta_w((\lambda_w-1)\epsilon_w+1)}\left(w_t^h - w_t\right) \\
- \frac{1+\iota_w \beta e^{(1-\sigma_c)\gamma}}{1+\beta e^{(1-\sigma_c)\gamma}}\pi_t + \frac{1}{1+\beta e^{(1-\sigma_c)\gamma}}\left(w_{t-1} - z_t - \iota_w \pi_{t-1}\right) \\
+ \frac{\beta e^{(1-\sigma_c)\gamma}}{1+\beta e^{(1-\sigma_c)\gamma}}\mathbb{E}_t\left[w_{t+1} + z_{t+1} + \pi_{t+1}\right] + \lambda_{w,t},
\end{aligned} \tag{32}
$$

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:

$$
w_t^h = \frac{1}{1-he^{-\gamma}}\left(c_t - he^{-\gamma}c_{t-1} + he^{-\gamma}z_t\right) + \nu_l L_t, \tag{33}
$$

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::[2]

$$
\begin{aligned}
R_t &= \rho_R R_{t-1} + (1-\rho_R)\left(\psi_1(\pi_t - \pi_t^*) + \psi_2(y_t - y_t^f)\right) \\
&\quad + \psi_3\left((y_t - y_t^f) - (y_{t-1} - y_{t-1}^f)\right) + r_t^m,
\end{aligned} \tag{34}
$$

where the flexible price/wage output y_t^f obtains from solving the version of the model without nominal rigidities (that is, Equations (21) through (30) and (33)), and the residual r_t^m follows an AR(1) process with parameters ρ_{r^m} and σ_{r^m}. This rule differs from the one in SW in that it has a time-varying inflation target, which evolves according to:

$$
\pi_t^* = \rho_{\pi^*}\pi_{t-1}^* + \sigma_{\pi^*}\epsilon_{\pi^*,t}, \tag{35}
$$

[2] We follow the specification in Del Negro and Eusepi (2011), while Aruoba and Schorfheide (2010) assume that the inflation target also affects the intercept in the feedback rule.

where $0 < \rho_{\pi^*} < 1$ and $\epsilon_{\pi^*,t}$ is an iid shock. We follow Erceg and Levin (2003) and model π_t^* as following a stationary process, although our prior for ρ_{π^*} will force this process to be highly persistent. We make this change as we include long-run inflation expectations to the set of observables.

3.2 Including Financial Frictions

We add to this model credit frictions as in Christiano et al. (2009). This amounts to replacing (24) with the following conditions:

$$E_t\left[\tilde{R}^k_{t+1} - R_t\right] = -b_t + \zeta_{sp,b}\left(q^k_t + \bar{k}_t - n_t\right) + \tilde{\sigma}_{\omega,t} \tag{36}$$

and

$$\tilde{R}^k_t - \pi_t = \frac{r^k_*}{r^k_* + (1-\delta)} r^k_t + \frac{(1-\delta)}{r^k_* + (1-\delta)} q^k_t - q^k_{t-1}, \tag{37}$$

where $\tilde{R}^k_t$ is the gross nominal return on capital for entrepreneurs, n_t is entrepreneurial equity, and $\tilde{\sigma}_{\omega,t}$ captures mean-preserving changes in the cross-sectional dispersion of ability across entrepreneurs (see Christiano et al. (2009)) and follows an AR(1) process with parameters ρ_{σ_ω} and σ_{σ_ω}. The second condition defines the return on capital, while the first one determines the spread between the expected return on capital and the riskless rate.[3] The following condition describes the evolution of entrepreneurial net worth:

$$\begin{array}{rcl} \hat{n}_t & = & \zeta_{n,\tilde{R}^k}\left(\tilde{R}^k_t - \pi_t\right) - \zeta_{n,R}\left(R_{t-1} - \pi_t\right) + \zeta_{n,qK}\left(q^k_{t-1} + \bar{k}_{t-1}\right) + \zeta_{n,n} n_{t-1} \\ & & -\frac{\zeta_{n,\sigma_\omega}}{\zeta_{sp,\sigma_\omega}} \tilde{\sigma}_{\omega,t-1} \end{array} . \tag{38}$$

3.3 State-Space Representation of the DSGE Model

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

[3] Note that if $\zeta_{sp,b} = 0$ and the financial friction shocks are zero, (24) coincides with (36) plus (37).

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:

$$s_t = \mathcal{T}(\theta)s_{t-1} + \mathcal{R}(\theta)\epsilon_t, \tag{39}$$

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

$$y_t = \mathcal{Z}(\theta)s_t + \mathcal{D}(\theta), \tag{40}$$

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.

The implied equilibrium law of motion in log-linear approximation form can be written as

$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t. \tag{41}$$

The system matrices Φ_1 and Φ_ϵ are functions of the DSGE model parameters θ, and s_t spans the state variables of the model economy.

The measurement equation linking the observable variables to the state variables is:

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)s_t. \tag{42}$$

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:

$$\begin{array}{rcl} \textit{Output growth} & = & \gamma + 100\left(y_t - y_{t-1} + z_t\right) \\ \textit{Consumption growth} & = & \gamma + 100\left(c_t - c_{t-1} + z_t\right) \\ \textit{Investment growth} & = & \gamma + 100\left(i_t - i_{t-1} + z_t\right) \\ \textit{Real Wage growth} & = & \gamma + 100\left(w_t - w_{t-1} + z_t\right)\ , \\ \textit{Hours} & = & \bar{l} + 100 l_t \\ \textit{Inflation} & = & \pi_* + 100\pi_t \\ \textit{FFR} & = & R_* + 100R_t \end{array} \tag{43}$$

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).

We further augment the set of measurement equations (43) by

$$\begin{aligned} \pi_t^{O,40} &= \pi_* + 100 I\!E_t \left[\frac{1}{40} \sum_{k=1}^{40} \pi_{t+k} \right] \\ &= \pi_* + \frac{100}{40} \Psi_2(\theta)_{(\pi,.)} (I - \Phi_1(\theta))^{-1} \left(\Phi_1(\theta) - \Phi_1(\theta)^{41} \right) s_t, \end{aligned} \tag{44}$$

where $\pi_t^{O,40}$ represents observed long run inflation expectations obtained from surveys (in percent per quarter), and the right-hand-side of (44) corresponds to expectations obtained from the DSGE model (in deviation from the mean π_*). The second line shows how to compute these expectations using the transition equation (41), where $\Psi_2(\theta)_{(\pi,.)}$ is the row of the matrix $\Psi_2(\theta)$ entering the measurement equation (42) corresponding to inflation.

Appendix A.1 provides further details on the data. In our benchmark specification we use data from 1964Q4 to 2011Q1. In Section 4.5 we consider a shorter sample in which end the sample in 2004Q4, so that we exclude the great recession.

Whenever we consider the extended model with credit frictions, the set of equations (43) is augmented to include:

$$\text{Spread} = SP_* + 100 I\!E_t \left[\tilde{R}^k_{t+1} - R_t \right], \tag{45}$$

where the parameter SP_* measures the steady state spread. We specify priors for the parameters SP_*, $\zeta_{sp,b}$, in addition to ρ_{σ_ω} and σ_{σ_ω}, and fix the parameters $\bar{F}_*$ and γ_* (steady state default probability and survival rate of entrepreneurs, respectively).

4 Results

4.1 Prior and Posterior for the DSGE Parameters

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.

4.2 Evidence against Gaussianity

In the introduction we showed evidence of rare large shocks and time varying volatility based on historical shocks extracted from standard gaussian estimation. In this section we assess the fit of models estimated with Student-t distributed shocks (St-t henceforth) and/or stochastic volatility (SV henceforth). Table 3 shows the log marginal likelihood for the different shock assumptions. We consider specifications with and without stochastic volatility (time-varying $\sigma_{q,t}$) and with different prior means for the degrees of freedom of the Student-t distributed component of the shocks (with $\lambda_q = \infty$ being the gaussian case). We consider three different priors for the degrees of freedom λ, which differ in terms of their mean $\bar{\lambda}$. 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. 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

The three priors capture three different views of the world on the importance of fat tails. The first prior, with $\bar{\lambda} = 15$, captures the view that the world is not quite Gaussian, but not too far from Gaussianity either. The second prior ($\bar{\lambda} = 9$) embodies the idea that the world is quite far from Gaussian, yet not too extreme. The last prior ($\bar{\lambda} = 6$) stands for a model with quite heavy tails. In the middle panel we consider the three different priors with 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 the assumption of Student-t distributed shocks vis-a-vis the Gaussian assumption. The fit also increases when we consider stochastic volatility instead of constant volatility. Furthermore, the fit is best when we consider both of these features, rather than only one. Importantly, the fit increases the lower the prior mean for the degrees of freedom for the Student-t distributed shocks, and this result is independent of how loose the prior is, although the difference across priors for λ naturally widens the tighter the prior. 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 197 log points in this case, compared to an increase of 161 log points for the latter. But considering both features is ideal, leading to an increase in fit of 222 log points.

Marginal likelihoods are difficult to compute, especially for these models, and for some readers are difficult to interpret. Therefore we also show the posterior distribution of λ obtained under the (almost) flat prior. Specifically, Table 4 shows the posterior mean and the posterior 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 6, and the posterior 90% bands around the median are all fairly tight. 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 ($\bar{\lambda}$ equal to 6 or 9) matters little for the posterior.

It is interesting to notice that 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). While the shocks to the price markup (λ_f) have a higher posterior mean in all cases. The two shocks related to monetary policy (inflation target, π^*, and deviations from the systematic response to the economy, r^m) exhibit posterior distributions 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 means 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. On the other hand, all of the shocks related to real frictions (as opposed to nominal frictions) are better explained by the model if we allow for fat tails, with a more limited role for the stochastic volatility component.

Figure 2 shows the smoothed shocks and the "tamed" version of these 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.

4.3 Do Fat Tails Matter for the Macroeconomy?

We have shown that the tails are fat – the estimated degrees of freedom are low. But what does this mean in terms of business cycle fluctuations? This section simply 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.[4] 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 the fluctuation variables. 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. Another example is that in the sharp contraction in output growth in 1980, about 1.5 percentage points was accounted the Student-t

[4] 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.

component of the shocks. In general, without rare shocks all recession would be of roughly the same magnitude. Since the fat tails accounted for a significant part of the large fluctuations 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) then it is even more clear that a substantial part of peaks and troughs (especially the latter ones) have a strong contribution from rare but large shocks. In the "Great Recession" the contraction in consumption growth would be a more modest 1% contraction, rather than the nearly 2.5% contraction observed. And the same can be said about the contractions in 1975 and 1980. Given this it is not surprising that the Student-t component accounts for about 26% of the peak volatility in the late 1970s.

Finally, in the lower panel we have the same experiments for hours worked (in logs). In the "Great Recession" the Student-t component accounted for about 4 percentage points decline in hours worked.

4.4 Student-t shocks and inference about time-variation in volatility

From the marginal likelihood analysis it is clear that the stochastic volatility has a role in explaining the data. Relative to the simple constant volatility and gaussian shocks it improves fit by as much as 161 log points, consistent with the work of Justiniano and Primiceri (2008). Once we had fat tails in the model the additional improvements is much smaller, but is nevertheless in there. We now discuss the extent to which accounting for fat tails makes us reevaluate the role of stochastic volatility in explaining the data and in particular the volatility in the data.

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 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.

For the discount rate shock, once we account for the fat tails, the stochastic volatility component captures only the slow moving trend in the volatility of the shock path. For the policy shock, in the lower panels, there is hardly any difference in the stochastic volatility component. In this case it seems to capture most of the movement in the shock volatility — consistent with the fact that the estimated degrees of freedom of the Student-t component for this shock is much higher once we allow for stochastic volatility (shown in Table 4).

With stochastic volatility, the model-implied volatility can change 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. The black lines show this volatility for the estimation with both stochastic volatility and Student-t components, while the red line shows this measure for the estimation with stochastic volatility but gaussian shocks. Solid line is the posterior median and the dashed lines correspond to the 90% bands around the median.

For all three variables shown the model-implied volatility is lower when we do not account for fat tails, but it is not simply a parallel shift, affecting all periods in the same way. For output and consumption growth 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. For hours worked we find a smaller difference across estimations, but there is nevertheless a non-negligible change in the model-implied volatility.

We now ask whether the evidence the "great moderation" is influenced by the presence of Student-t shocks. shows that the magnitude of the Great Moderation is still substantial even allowing for Student-t shocks, but that introducing the latter has a non-negligible impact. The right 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.

It is clear that most of the probability mass is above one, confirming the high probability of a fall in volatility between 1981 and 1994. However, the posterior distribution for this ratio in the estimation with Student-t shocks for output growth is shifted to the left relative to the case with gaussian shocks (indeed the median is 1.87 in the former, compared to 2.18 in the latter). This same pattern is also evident for the consumption growth, shown in the middle panel.

4.5 Sub-Sample Excluding the Great Recession

See Table 5: similar results to those in Table 3.

5 Conclusions

To be written

References

An, Sungbae and Frank Schorfheide, "Bayesian Analysis of DSGE Models–Rejoinder," *Econometric Reviews*, 2007, *26* (2-4), 211–219.

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

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, "The Financial Accelerator in a Quantitative Business Cycle Framework," in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, Vol. 1C, Amsterdam: North-Holland, 1999, chapter 21, pp. 1341–93.

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

_ **and Srikanth Ramamurthy**, "DSGE Models with Student-t errors," *mimeo, Washington University in St. Louis*, 2011.

Christiano, Lawrence J., "Comment on 'On the Fit of New Keynesian Models' by Del Negro ,Schorfheide, Smets ,Wouters," *Journal of Business and Economic Statistics*, 2007, *25* (2), 143–151.

Christiano, Lawrence, Martin Eichenbaum, and Charles Evans, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 2005, *113*, 1–45.

_ **, Roberto Motto, and Massimo Rostagno**, "Financial Factors in Economic Fluctuations," *Unpublished*, May 2009.

Durbin, James and Siem Jan Koopman, "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 2002, *89* (3), 603–616.

Erceg, Christopher J. and Andrew T. Levin, "Imperfect Credibility and Inflation Persistence," *Journal of Monetary Economics*, 2003, *50*, 915–944.

Fernández-Villaverde, Jesús and Juan F. Rubio-Ramírez, "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 2007, *74* (4), 1059–1087.

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

_ , "Priors for Macroeconomic Time Series and Their Application," *Econometric Theory*, 1994, *10*, 609–632.

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

_ , *Contemporary Bayesian Econometrics and Statistics*, Wiley, 2005.

Greenwood, Jeremy, Zvi Hercovitz, and Per Krusell, "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review*, 1998, *87* (3), 342–36.

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

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 1998, *65* (3), 361–393.

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

Negro, Marco Del and Stefano Eusepi, "Fitting Observed Ination Expectations," *Journal of Economic Dynamics and Control*, 2011, *35*, 2105–2131.

Sims, Christopher A., "Solving Rational Expectations Models," *Computational Economics*, 2002, *20*, 1–20.

Smets, Frank and Raf Wouters, "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 2003, *1* (5), 1123 – 1175.

_ **and** _ , "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 2007, *97* (3), 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 (43). 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.

The long-run inflation forecasts are obtained from the Blue Chip Economic Indicators survey and the Survey of Professional Forecasters (SPF) available from the FRB Philadelphia. Long-run inflation expectations (average CPI inflation over the next 10 years) are available from 1991:Q4 onwards. Prior to 1991:Q4, we use the 10-year expectations data from the Blue Chip survey to construct a long time series that begins in 1979:Q4. Since the Blue Chip survey reports long-run inflation expectations only twice a year, we treat these expectations in the remaining quarters as missing observations and adjust the measurement equation of the Kalman filter accordingly.

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:

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

Long-run inflation expectations $\pi_t^{O,40}$ are therefore measured as

$$\pi_t^{O,40} = (\text{10-YEAR AVERAGE CPI INFLATION FORECAST} - 0.50)/4.$$

where .50 is the average difference between CPI and GDP annualized inflation, and where we divide by 4 since the data are expressed in quarterly terms.

In the estimation of the DSGE model with financial frictions we measure *Spread* as the annualized Moody's Seasoned Baa Corporate Bond Yield spread over the 10-Year Treasury Note Yield at Constant Maturity. Both series are available from the Federal Reserve Board's H.15 release, and averaged over each quarter. Spread data is also not revised.

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:

$$\begin{aligned} p(y_{1:T}) &= \int p(y_{1:T}|s_{1:T},\theta)p(s_{1:T}|\varepsilon_{1:T},\theta)p(\varepsilon_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta) \\ &\qquad p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\tilde{\sigma}_{1:T}|\omega^2_{1:\bar{q}})p(\lambda_{1:\bar{q}})p(\omega^2_{1:\bar{q}})p(\theta) \\ &\qquad\qquad d(s_{1:T},\varepsilon_{1:T},\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}},\theta), \\ &= \int p(y_{1:T}|\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\theta)p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\tilde{\sigma}_{1:T}|\omega^2_{1:\bar{q}}) \\ &\qquad p(\lambda_{1:\bar{q}})p(\omega^2_{1:\bar{q}})p(\theta)d(\tilde{h}_{1:T},\tilde{\sigma}_{1:T},\lambda_{1:\bar{q}},\rho_{1:\bar{q}},\omega^2_{1:\bar{q}},\theta) \end{aligned} \tag{46}$$

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

$$f(\theta,\tilde{h}_{1:T}) = f(\theta)\cdot p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\tilde{\sigma}_{1:T}|\omega^2_{1:\bar{q}})p(\lambda_{1:\bar{q}})p(\omega^2_{1:\bar{q}}), \tag{47}$$

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

$$\hat{p}(y_{1:T}) = \left[\frac{1}{n_{sim}}\sum_{j=1}^{n_{sim}}\frac{f(\theta^j)}{p(y_{1:T}|\tilde{h}^j_{1:T},\tilde{\sigma}^j_{1:T},\theta^j)p(\theta^j)}\right]^{-1} \tag{48}$$

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 (47), 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:

$$\varepsilon^*_{q,t} = 2\tilde{\sigma}_{q,t} + \eta^*_{q,t} \tag{49}$$

where

$$\varepsilon^*_{q,t} = \log(\sigma_q^{-2}\tilde{h}_{q,t}\varepsilon^2_{q,t} + c), \tag{50}$$

$c = .001$ 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:

$$p(\eta^*_{q,t}) = \sum_{k=1}^{K} \pi^*_k \mathcal{N}(m^*_k - 1.2704, {\nu^*_k}^2) \tag{51}$$

The parameters that optimize this approximation, namely $\{\pi^*_k, m^*_k, \nu^*_k\}_{k=1}^K$ 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:

$$\eta^*_{q,t}|\varsigma_{q,t} = k \sim \mathcal{N}(m^*_k - 1.2704, {\nu^*_k}^2), \ Pr(s_{i,t} = k) = \pi^*_k. \tag{52}$$

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 (51) for each q. Specifically:

$$Pr\{\varsigma_{q,t} = k|\tilde{\sigma}_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}\dots\} \propto \pi^*_k \nu^{-1}_k \exp\left[-\frac{1}{2{\nu^*_k}^2}(\eta^*_{q,t} - m^*_k + 1.2704)^2\right]. \tag{53}$$

where from (49) $\eta^*_{q,t} = \varepsilon^*_{q,t} - 2\tilde{\sigma}_{q,t}$.

(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 (49) 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	0.62	0.10
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
ρ_{π^*}	Beta	0.50	0.20	σ_{π^*}	InvG	0.03	6.00
η_{λ_f}	Beta	0.50	0.20	η_{λ_w}	Beta	0.50	0.20
η_{gz}	Beta	0.50	0.20				
Financial Frictions ($SW-FF$)							
SP_*	Gamma	2.00	0.10	$\zeta_{sp,b}$	Beta	0.05	0.005
ρ_{σ_w}	Beta	0.75	0.15	σ_{σ_w}	InvG	0.05	4.00

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 $\varepsilon_p = 10$. In addition, for the model with financial frictions we fix $\bar{F}_* = .03$ and $\gamma_* = .99$. 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	0.146	0.144	0.153	0.143
ζ_p	0.672	0.703	0.671	0.702
ι_p	0.184	0.200	0.229	0.221
Φ	1.604	1.622	1.643	1.659
S''	5.410	6.301	6.826	6.856
h	0.687	0.655	0.740	0.701
psi	0.704	0.686	0.681	0.618
ν_l	1.771	2.019	2.032	1.834
ζ_w	0.726	0.771	0.718	0.717
ι_w	0.563	0.488	0.544	0.435
β	0.203	0.181	0.215	0.201
ψ_1	1.944	2.177	2.180	2.228
ψ_2	-0.006	-0.015	-0.017	-0.016
ψ_3	0.219	0.150	0.159	0.125
π^*	0.805	0.556	0.577	0.604
σ_c	1.200	1.491	1.257	1.561
ρ	0.785	0.844	0.841	0.847
γ	0.367	0.310	0.362	0.320
ρ_g	0.978	0.987	0.981	0.990
ρ_b	0.476	0.483	0.455	0.394
ρ_μ	0.745	0.721	0.742	0.721
ρ_z	0.981	0.994	0.984	0.994
ρ_{λ_f}	0.919	0.837	0.911	0.812
ρ_{λ_w}	0.986	0.985	0.978	0.981
ρ_{rm}	0.238	0.257	0.248	0.259
ρ^*_{pi}	0.990	0.990	0.990	0.990
σ_g	2.920	0.615	2.390	0.355
σ_b	0.187	0.093	0.144	0.065
σ_μ	0.417	0.166	0.324	0.092
σ_z	0.486	0.371	0.345	0.167
σ_{λ_f}	0.162	0.056	0.133	0.048
σ_{λ_w}	0.298	0.095	0.219	0.048
σ_{rm}	0.242	0.055	0.132	0.041
σ^*_{pi}	0.033	0.021	0.017	0.019
η_{gz}	0.779	0.778	0.769	0.773
η_{λ_f}	0.771	0.703	0.771	0.685
η_{λ_w}	0.912	0.939	0.871	0.886

Notes: We consider a prior mean of 6 degrees of freedom for the Student-t distributed component. 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		
	-973.3	-812.1
Student-t distributed shocks, prior with 4 degrees of freedom		
$\lambda = 15$	-810.5	-762.0
$\lambda = 9$	-796.7	-765.1
$\lambda = 6$	-780.3	-753.0
Student-t distributed shocks, prior with 1 degree of freedom		
$\lambda = 15$	-779.7	-759.9
$\lambda = 9$	-783.2	-754.5
$\lambda = 6$	-776.9	-751.7

Notes: The parameter λ 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*	
	$\lambda = 9$	$\lambda = 6$	$\lambda = 9$	$\lambda = 6$
g	7.1	6.1	6.3	5.6
	(2.5,11.9)	(2.5,9.9)	(2.2,10.7)	(2.1,9.2)
b	4.8	4.7	5.5	5.4
	(2.4,7.1)	(2.5,7.0)	(2.4,8.5)	(2.4,8.4)
μ	5.8	5.4	7.3	6.5
	(2.4,9.1)	(2.5,8.5)	(2.4,12.4)	(2.4,10.7)
z	3.9	3.7	5.2	4.3
	(1.8,5.9)	(1.8,5.6)	(1.7,9.0)	(1.7,7.2)
λ_f	7.3	6.8	16.1	12.7
	(2.5,12.4)	(2.5,11.3)	(4.6,28.0)	(4.2,21.2)
λ_w	5.9	5.5	4.9	4.8
	(2.4,9.4)	(2.5,8.5)	(2.4,7.4)	(2.4,7.1)
r^m	2.8	2.7	13.3	10.9
	(1.7,3.8)	(1.7,3.7)	(3.9,23.2)	(3.7,18.0)
π^*	1.7	1.7	11.2	9.0
	(1.2,2.2)	(1.2,2.2)	(3.1,20.1)	(2.9,15.6)

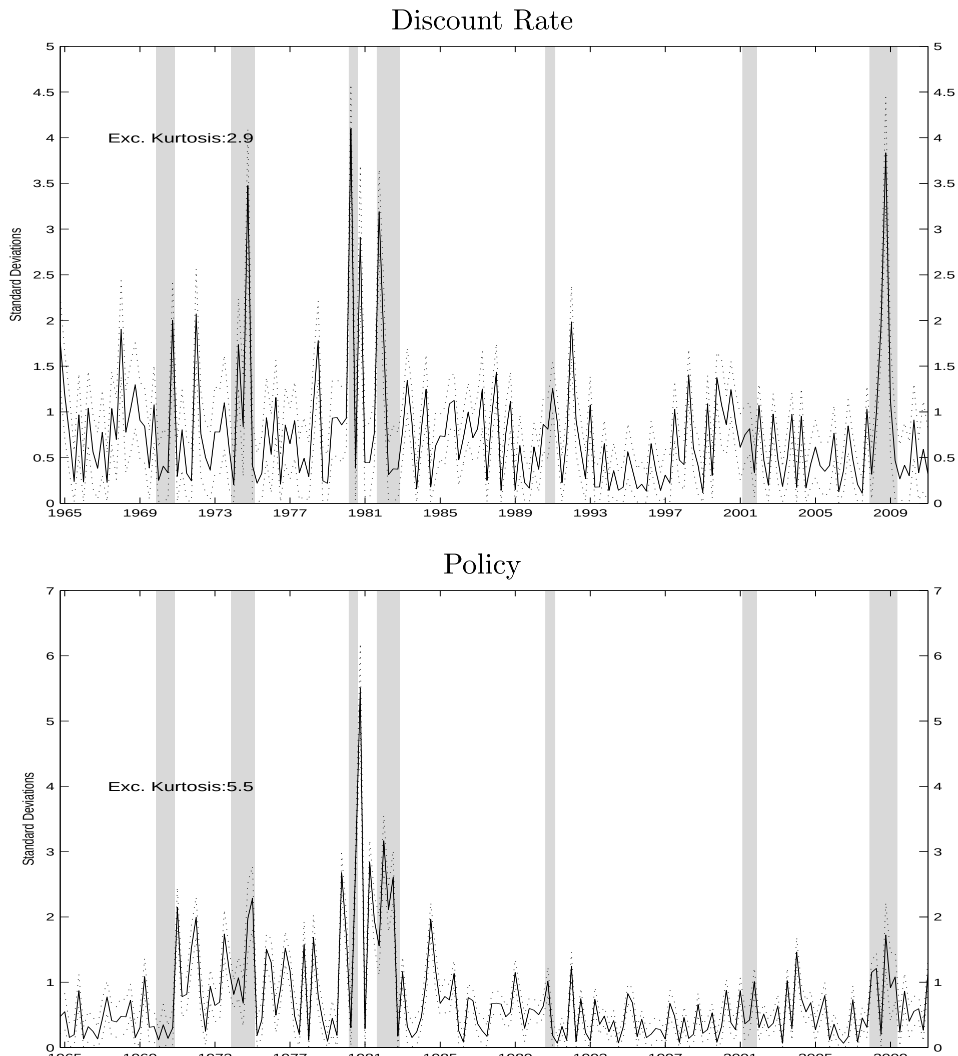
Notes: Numbers shown for the posterior mean and the 90% intervals of the degrees of freedom parameter.

Table 5: Marginal Likelihoods, Sample Ending in 2004Q4

	Constant Volatility	Stochastic Volatility
Gaussian shocks		
	-864.7	-742.1
Student-t distributed shocks, prior with 1 degree of freedom		
$\lambda = 15$	-731.9	-709.9
$\lambda = 9$	-731.8	-705.3
$\lambda = 6$	-729.58	-700.9
Student-t distributed shocks, prior with 4 degrees of freedom		
$\lambda = 15$	-746.8	-712.5
$\lambda = 9$	-735.7	-710.3
$\lambda = 6$	-719.2	-694.9

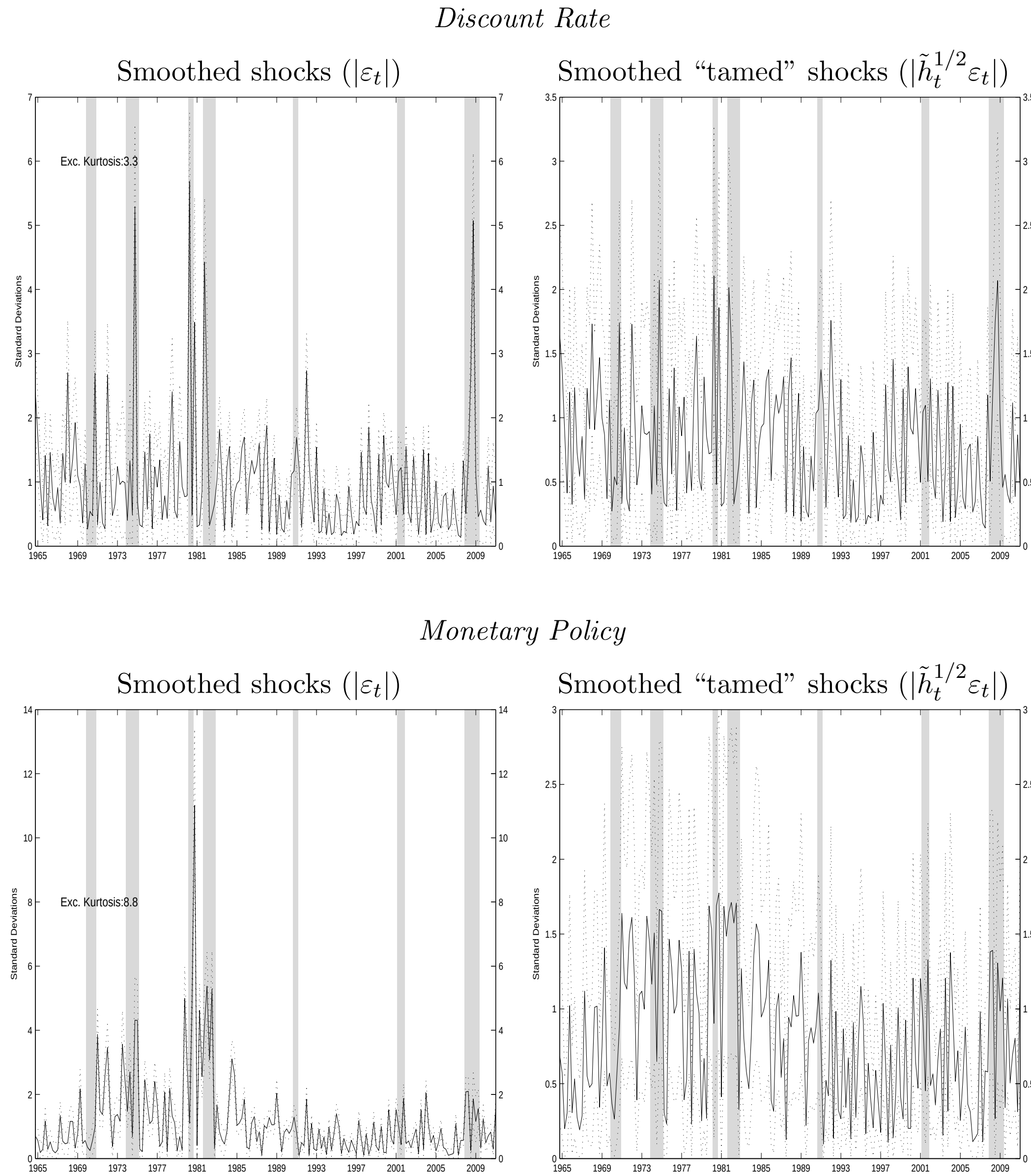
Notes: The parameter λ represents the prior mean for the degrees of freedom in the Student-t distribution.

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



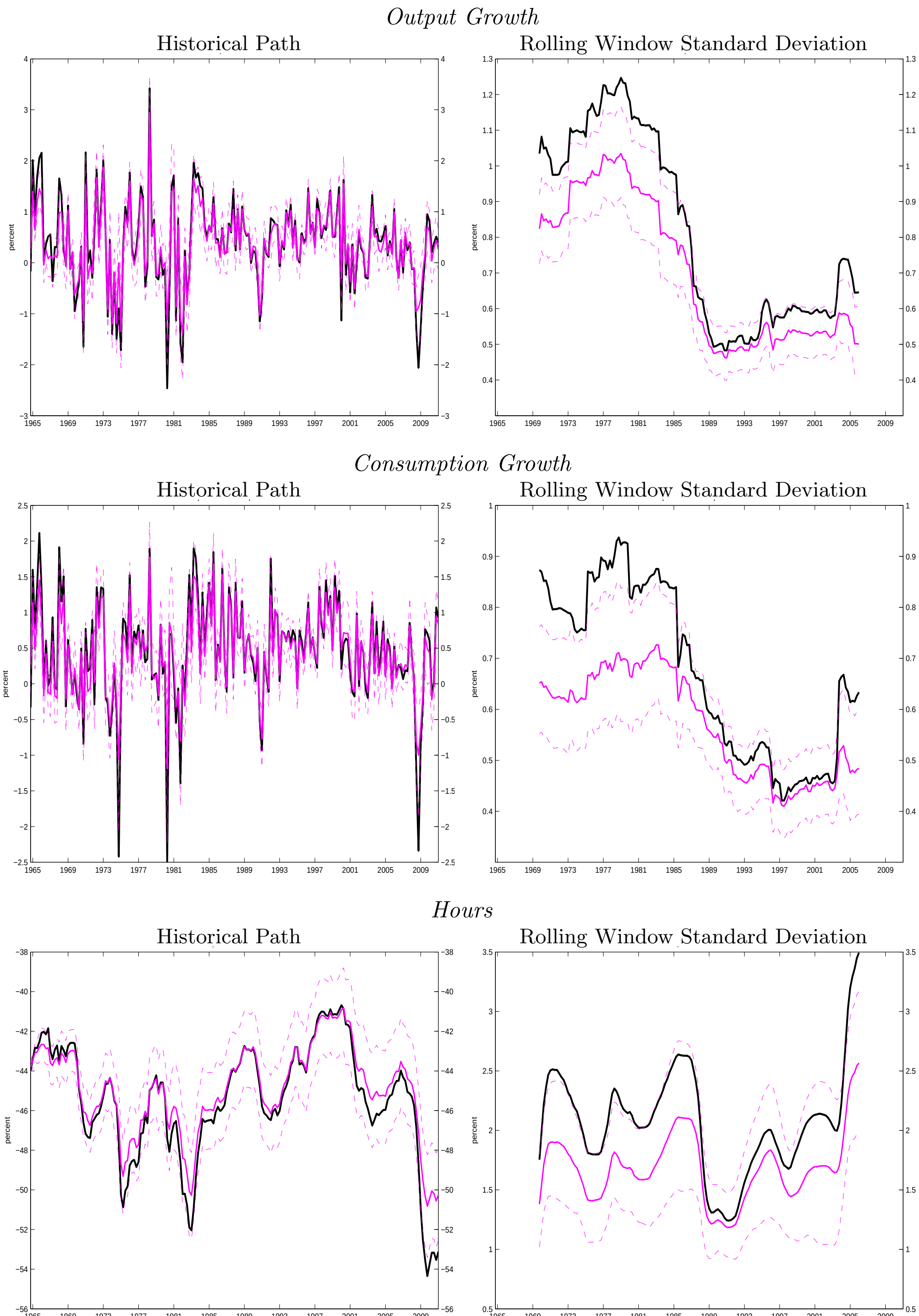
Notes: The solid line is the median, and the dashed lines are the posterior 90% bands. The vertical shaded regions identify NBER recession dates.

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.

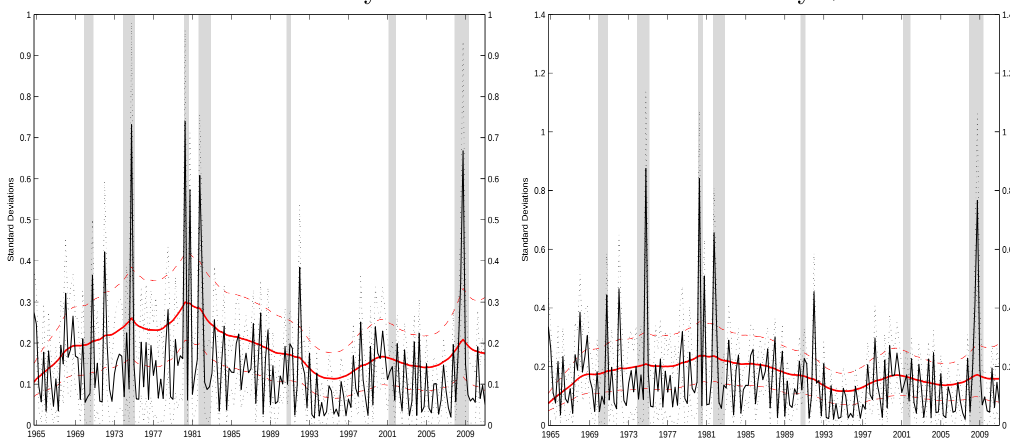


Notes: Black lines are the historical evolution of the variable, and pink lines are the median counterfactual evolution of the same variable if we shut down the Student-t distributed component of all shocks. The rolling window standard deviation uses 20 quarters before and 20 quarters after a given quarter.

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

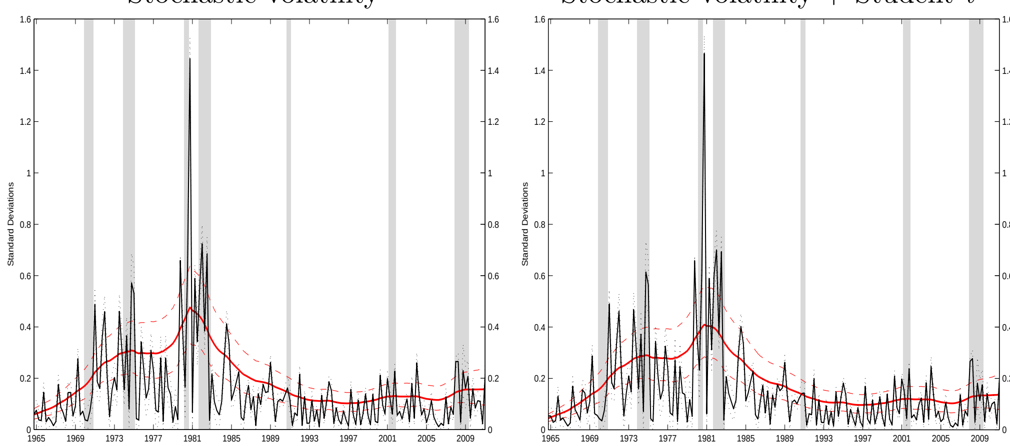
Discount Rate

Stochastic volatility | Stochastic volatility + Student-t



Monetary Policy

Stochastic volatility | Stochastic volatility + Student-t

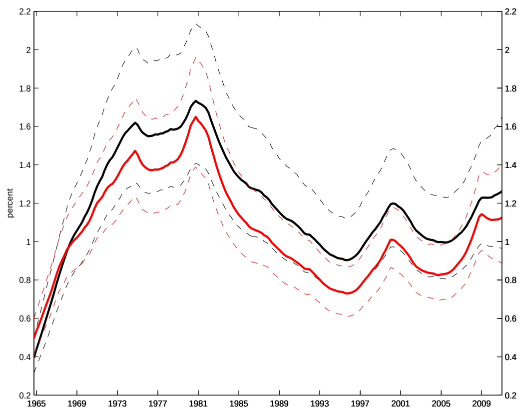


Notes: Estimation with Student-t distribution with $\underline{\lambda} = 6$. The solid line is the median, and the dashed lines are the posterior 90% bands. Black line is the absolute value of the shock, and the red line is the stochastic volatility component.

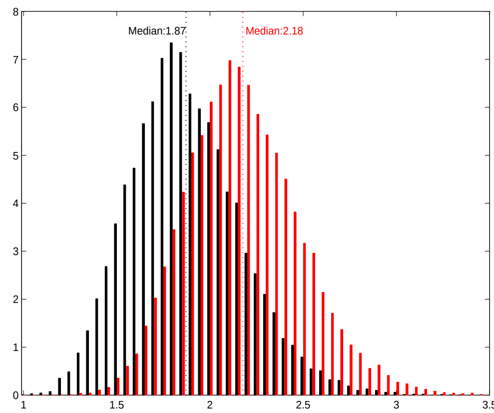
Figure 5: Time-Variation in the unconditional variance of output and consumption; models estimated with and without the Student-t distributed component.

Output Growth

Unconditional variance over time

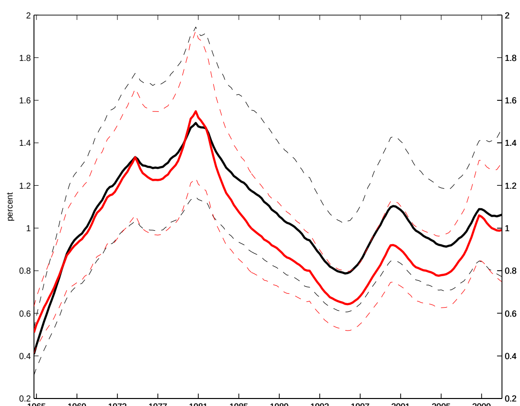


Ratio of variance in 1981 relative to 1994

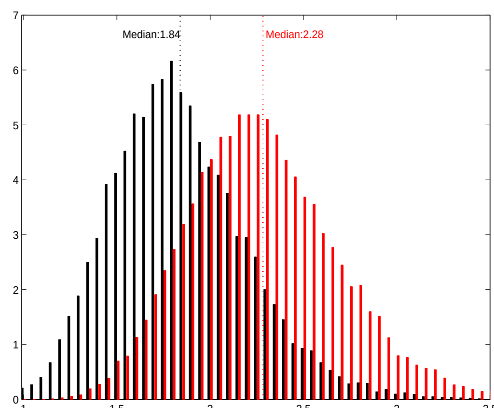


Consumption Growth

Unconditional variance over time



Ratio of variance in 1981 relative to 1994



Notes: Black line in the left 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 right 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.