COLLEGIUM OF ECONOMIC ANALYSIS
WORKING PAPER SERIES

Smets and Wouters model estimated with skewed shocks - empirical study of forecasting properties

Grzegorz Koloch

SGH KAE Working Papers Series Number: 2016/023 December 2016

Smets and Wouters model estimated with skewed shocks - empirical study of forecasting properties*

Grzegorz Koloch†

December 19, 2016

Abstract

In this paper we estimate a Smets and Wouters (2007) model with shocks following a closed skew normal distribution (csn), which was introduced by Gonzalez-Farias et al. (2004) and which nests a normal distribution as a special case. In the paper we present the identification procedure, discuss priors for model parameters, including skewness-related parameters of shocks, i.e. location, scale and skewness parameters. Using data ranging from $1990Q1$ to $2012Q2$ we estimate the model and recursively verify its out-of sample forecasting properties for time period $2007Q1$ - $2012Q2$, therefore including the recent financial crisis, within a forecasting horizon from 1 up to 8 quarters ahead. Using a RMSE measure we compare the forecasting performance of the model with skewed shocks with a model estimated using normally distributed shocks. We find that inclusion of skewness can help forecasting some variables (consumption, investment and hours worked), but, on the other hand, results in deterioration in the others (output, inflation, wages and the short rate).

Keywords: DSGE, Forecasting, Closed Skew-Normal Distribution.

JEL: C51, C13, E32

*This research was financed by the National Science Center grant No. 2012/07/E/HS4/01080.

†Warsaw School of Economics, Collegium of Economic Analysis, Institute of Econometrics, Decision Analysis and Support Unit, Al. Niepodleglosci 164, 00-950 Warsaw, Poland.

1 Introduction

In this paper we estimate a Smets and Wouters (2007) model with shocks following a closed skew normal distribution (csn) introduced in Gonzalez-Farias et al. (2004), which nests a normal distribution as a special case. In the paper we discuss the identification procedure, priors for model parameters, including skewness-related parameters of shocks, i.e. location, scale and skewness parameters. Using data ranging from 1990$Q1$ to 2012$Q2$ we estimate the model and recursively verify its out-of sample forecasting properties for time period 2007$Q1$ - 2012$Q2$, therefore including the recent financial crisis, within a forecasting horizon from 1 up to 8 quarters ahead. Using a RMSE measure we compare the forecasting performance of the model with skewed shocks wit a model estimated using normally distributed shocks. We find that inclusion of skewness can help forecasting some variables (consumption, investment and hours worked), but, on the other hand, results in deterioration in the others (output, inflation, wages and the short rate).

The paper is organized in the following way. In section two we outline the estimation procedure. Section three presents the Smets and Wouters (2007) model. Section four presents the dataset, *a priori* assumptions for model parameters and discusses estimation results for the entire data set, i.e. for time span 1990$Q1$ − 2012$Q2$. Section five discusses forecasting properties of the model with skewed shocks relative to the one estimated using a standard normal distribution based Kalman filter.

2 The skewed Kalman filter

In this section we briefly describe the CSN distribution, the state-space setting within which the model (see section 3) is formulated and estimated, as well as the filtering distributions and the likelihood function.

2.1 The CSN distribution

Let us denote a density function of a p-dimensional normal distribution with mean[1] μ and covariance matrix Σ by $\phi_p(z; \mu, \Sigma)$. Let us also denote a cumulative distribution

[1] All vectors are column vectors in this paper.

function of a q-dimensional normal distribution with mean μ and covariance matrix Σ by $\Phi_q(z;\mu,\Sigma)$. We will now define the closed skew-normal, possibly singular, distribution by means of the moment generating function (mgf) and then, under nonsingularity of the covariance matrix, by means of the probability density function (pdf). For explanation of this distribution we refer to Genton (2004).

Definition 2.1. (*csn distribution – mgf*) Let $\mu \in \mathrm{R}^p$ and $\vartheta \in \mathrm{R}^q$, $p,q \geq 1$. Let $\Sigma \in \mathrm{R}^{p\times p}$ and $\Delta \in \mathrm{R}^{q\times q}$, $\Sigma \geq 0$, $\Delta > 0$, and let $D \in \mathrm{R}^{q\times p}$. We say that random variable z has a (p,q)-dimensional closed skew-normal distribution with parameters μ, Σ, D, ϑ and Δ if moment generating function of z, $M_z(t)$, is given by:

$$M_z(t) = \frac{\Phi_q(D\Sigma t;\vartheta,\Delta + D\Sigma D^T)}{\Phi_q(0;\vartheta,\Delta + D\Sigma D^T)}\mathrm{e}^{t^T\mu+\frac{1}{2}t^T\Sigma t}$$

which henceforth will be denoted by:

$$z \sim csn_{p,q}(\mu,\Sigma,D,\vartheta,\Delta)$$

If $\Sigma > 0$, a *csn* random variable z obtains a probability density function.

Definition 2.2. (*csn distribution – pdf*) If a random variable z follows a (p,q)-dimensional, $p,q \geq 1$, closed skew-normal distribution with parameters μ, Σ, D, ϑ and Δ, where $\mu \in \mathrm{R}^p$, $\vartheta \in \mathrm{R}^q$, $\Sigma \in \mathrm{R}^{p\times p}$, $\Sigma > 0$, $\Delta \in \mathrm{R}^{q\times q}$, $\Delta > 0$ and $D \in \mathrm{R}^{q\times p}$, than probability density function of z is given by:

$$p(z) = \phi_p(z;\mu,\Sigma)\frac{\Phi_q(D(z-\mu);\vartheta,\Delta)}{\Phi_q(0;\vartheta,\Delta + D\Sigma D^T)}$$

Density function $p(z)$ defines a (p,q)-dimensional *nonsingular* closed skew-normal distribution in the sense that a random variable has (p,q)-dimensional nonsingular closed skew-normal distribution with parameters μ, Σ, D, ϑ and Δ if and only if its density function for every $z \in \mathrm{R}^p$ equals $p(z)$. Parameters μ, Σ and D have interpretation of location, scale and skewness parameters respectively. Parameters ϑ and Δ are artificial, but inclusion of them in the definition of $p(z)$ allows for closure of the *csn* distribution under conditioning and taking marginals respectively. The q-dimension in Φ_q is also artificial, but it allows for closure of sums of independent random variables and for taking the joint distribution of independent (not necessarily *iid*) random variables. For $D = 0$, the *csn* distribution reduces to a p-dimensional normal one. Dimension q can be interpreted as a skewness related degree of freedom.

2.2 State-space setting

Let us consider the following state-space model:

$$\begin{aligned} y_t &= F x_t + H u_t, \\ x_t &= A x_{t-1} + B \xi_t, \\ u_t &\sim N(0, \Sigma_u), \\ \xi_t &\sim p_\xi(\theta) \\ x_0 &\sim N(\mu_{x_0}, \Sigma_{x_0}) \end{aligned}$$

for $t \in \mathcal{T} = \{1, 2, ..., T\}$, where $x_t \in \mathrm{R}^p$ and $y_t \in \mathrm{R}^n$ denote states and observed variables[2], $\xi_t \in \mathrm{R}^{n_\xi}$ and $u_t \in \mathrm{R}^{n_u}$ denote shocks and measurement errors, whereas $p_\xi(\theta)$ denotes a probability distribution function of martingale difference shocks ξ_t, which depends on a vector of parameters θ.

A usual assumption is that p_ξ is a multivariate normal distribution, independent across its dimensions. In such a case Kalman filter constitutes an optimal filtration procedure[3], see Simon (2006). If normality assumption is relaxed, Kalman filter remains an optimal linear filter. In this paper, we relax normality assumption and assume that for some values of θ, probability density function p_ξ is skewed (asymmetric). In particular, we assume that shocks ξ_t follow a closed skew-normal distribution (*csn* henceforth), which nests the normal distribution as a special case, see Goznalez-Farias et. al. (2006) or Genton (2004). Shocks u_t in the measurement equation as well as initial states x_0 are assumed to be normally distributed for simplicity, they could also follow a *csn* distribution. The *csn* distribution is chosen because it nests a normal distribution and is analytically trackable. It constitutes a multivariate generalization of a class of one-dimensional skew-normal distributions which, among such generalizations, is closed under most forms of operations imposed on variables in the state-space setting (full rank linear transformations, conditioning, taking joints and marginals of independent variables)[4]. It also allows for analytical

[2]In general, the term states can refer both to observed and unobserved variables within the state-space setting, but in this paper we make a distinction, so that the term states is reserved to unobserved endogenous variables. Observed variables will also be called observables.

[3]In the sense that it minimizes the trace of one-step ahead in-sample forecast errors' covariance matrix.

[4]Details are provided in Section 2.

derivation of filtration and prediction densities as well as of the likelihood function. These features make it technically similar to the normal distribution.

To allow for prediction, filtration and estimation, we provide formulae for $p(y_t|Y_{t-1})$, $p(x_t|Y_{t-1})$, $p(x_t|Y_t)$ and for the likelihood function.

2.3 Conditional distributions

For $t \in \mathcal{T}$, let us define an information set $Y_t = \{y_0, y_1, y_2, ..., y_t\}$ which consists of observables up to time t. Assume, that the *a posteriori* distribution of state variables in time t, i.e. distribution of x_{t-1} conditional on Y_{t-1}, is:

$$x_{t-1}|Y_{t-1} \sim csn_{p,q_{t-1}}(\mu_{x,t-1|t-1}, \Sigma_{x,t-1|t-1}, D_{x,t-1|t-1}, \vartheta_{x,t-1|t-1}, \Delta_{x,t-1|t-1})$$

with parameters $\mu_{x,t-1|t-1} \in \mathrm{R}^p$, $\mathrm{R}^{p\times p} \ni \Sigma_{x,t-1|t-1} > 0$, $\mathrm{R}^{q\times p} \ni D_{x,t-1|t-1} > 0$, $\vartheta_{x,t-1|t-1} \in \mathrm{R}^{q\times 1}$ and $\Delta_{x,t-1|t-1} \in \mathrm{R}^{q\times q}$. If, so, then joint distribution of $s_t = (x_{t-1}^T, \xi_t^T)^T$, conditional on Y_{t-1} is given by:

$$s_t|Y_{t-1} \sim csn_{p+n_\xi, q_{t-1}+q_\xi}(\mu_{s,t|t-1}, \Sigma_{s,t|t-1}, D_{s,t|t-1}, \vartheta_{s,t|t-1}, \Delta_{s,t|t-1})$$

where:

$$\mu_{s,t|t-1} = (\mu_{x,t-1|t-1}^T, \mu_\xi^T)^T, \quad \Sigma_{s,t|t-1} = \Sigma_{x,t-1|t-1} \oplus \Sigma_\xi, \quad D_{s,t|t-1} = D_{x,t-1|t-1} \oplus D_\xi,$$

$$\vartheta_{s,t|t-1} = (\vartheta_{x,t-1|t-1}^T, \vartheta_\xi^T)^T, \quad \Delta_{s,t|t-1} = \Delta_{x,t-1|t-1} \oplus \Delta_\xi$$

and, assuming $G = [A, B]$ has a full row rank, the *a priori* distribution of $x_t = Ax_{t-1} + B\xi_t = Gs_t$, conditional on Y_{t-1} is given by:

$$x_t|Y_{t-1} \sim csn_{p,q_{t-1}+q_\xi}(\mu_{x,t|t-1}, \Sigma_{x,t|t-1}, D_{x,t|t-1}, \vartheta_{x,t|t-1}, \Delta_{x,t|t-1})$$

where:

$$\mu_{x,t|t-1} = G\mu_{s,t|t-1}, \quad \Sigma_{x,t|t-1} = G\Sigma_{s,t|t-1}G^T, \quad D_{x,t|t-1} = D_{s,t|t-1}\Sigma_{s,t|t-1}G^T\Sigma_{x,t|t-1}^{-1},$$

$$\vartheta_{x,t|t-1} = \vartheta_{s,t|t-1},$$

$$\Delta_{x,t|t-1} = \Delta_{s,t|t-1} + D_{s,t|t-1}\Sigma_{s,t|t-1}D_{s,t|t-1}^T - D_{s,t|t-1}\Sigma_{s,t|t-1}G^T\Sigma_{x,t|t-1}^{-1}G\Sigma_{s,t|t-1}D_{s,t|t-1}^T$$

Since distribution of measurement errors $u_t \sim N(0, \Sigma_u)$ can be represented as:

$$u_t \sim csn_{n_u,n_u}(0_{n_u,1}, \Sigma_u, 0_{n_u,n_u}, 1_{n_u,1}, I_{n_u,n_u})$$

and observation equation of the state-space setting, i.e. a likelihood equation, reads $y_t = Fx_t + u_t$, then *a posteriori* distribution of x_t, i.e. distribution of x_t conditional on Y_{t-1} and $y_t = \alpha$, is as follows:

$$x_t|Y_t = (x_t|y_t = \alpha)|Y_{t-1} \sim csn_{p,q_{t-1}+q_\xi}(\mu_{x,t|t}, \Sigma_{x,t|t}, D_{x,t|t}, \vartheta_{x,t|t}, \Delta_{x,t|t})$$

where:

$$\mu_{x,t|t} = \mu_{x,t|t-1} + K(\alpha - F\mu_{x,t|t-1})$$

$$\Sigma_{x,t|t} = \Sigma_{x,t|t-1} - KF\Sigma_{x,t|t-1}$$

$$D_{x,t|t} = \left(\begin{pmatrix} D_{x,t|t-1}\Sigma_{x,t|t-1} \\ 0 \end{pmatrix} - \begin{pmatrix} D_{x,t|t-1}\Sigma_{x,t|t-1}F^T \\ 0 \end{pmatrix} (F\Sigma_{x,t|t-1}F^T + \Sigma_u^T)^{-1} F\Sigma_{x,t|t-1} \right) \Sigma_{x,t|t}^{-1}$$

$$\vartheta_{x,t|t} = \begin{pmatrix} \vartheta_{x,t|t-1} \\ 0 \end{pmatrix} - \begin{pmatrix} D\Sigma_{x,t|t-1}F^T \\ 0 \end{pmatrix} (F\Sigma_{x,t|t-1}F^T + \Sigma_u)^{-1} F\Sigma_{x,t|t-1} \;(\alpha - F\mu_{x,t|t-1})$$

$$\Delta_{x,t|t} = \begin{pmatrix} \Delta_{x,t|t-1} + D_{x,t|t-1}\Sigma_{x,t|t-1}D_{x,t|t-1}^T & 0 \\ 0 & I_{n_u,n_u} \end{pmatrix} +$$

$$- \begin{pmatrix} D_{x,t|t-1}\Sigma_{x,t|t-1}F^T \\ 0 \end{pmatrix} (F\Sigma_{x,t|t-1}F^T + \Sigma_u)^{-1} \begin{pmatrix} D_{x,t|t-1}\Sigma_{x,t|t-1}F^T \\ 0 \end{pmatrix}^T - D_{x,t|t}\Sigma_{x,t|t}D_{x,t|t}^T$$

for $K = \Sigma_{x,t|t-1}F^T(F\Sigma_{x,t|t-1}F^T + \Sigma_u)^{-1}$.

Notice, that in the formula for $D_{x,t|t}$ one requires to perform inversion of $\Sigma_{x,t|t}$ and in the formula for $D_{x,t|t-1}$ one requires to perform inversion of $\Sigma_{x,t|t-1}$, therefore, for the *csn* distribution to propagate through the state-space setting, matrices $\Sigma_{x,t|t}$ and $\Sigma_{x,t|t-1}$ must be of full rank. However, $\Sigma_{x,t|t}$ is updated in every iteration of the filtering routine according to $\Sigma_{x,t|t} = \Sigma_{x,t|t-1} - KF\Sigma_{x,t|t-1}$, so that it can become singular for some $t > 1$. Since update of $\Sigma_{x,t|t}$ in the *csn* filter is, technically, the same as in case of the Kalman filter, this can also be the case under assumption of normality. The difference, however, is, that in the standard Kalman filter inversion not of $\Sigma_{x,t|t}$ is involved, but of $(F\Sigma_{x,t|t-1}F^T + \Sigma_u)^{-1}$, which is kept positive definite throughout the filter's iterations. Possible singularity of $\Sigma_{x,t|t}$ precludes, in general case, propagation of the *csn* distribution within the state-space. Moreover, since $\Sigma_{x,t|t-1} = G\Sigma_{s,t|t-1}G^T = G(\Sigma_{x,t-1|t-1} \oplus \Sigma_\xi)G^T$, singularity of $\Sigma_{x,t|t}$, even if G is full row rank, can translate into singularity of $\Sigma_{x,t+1|t}$.

Expression $D_{x,t|t-1} = D_{s,t|t-1}\Sigma_{s,t|t-1}G^T\Sigma_{x,t|t-1}^{-1}$ results as a solution, with respect to $D_{x,t|t-1}$, of a matrix equation $D_{x,t|t-1}\Sigma_{x,t|t-1} = D_{s,t|t-1}\Sigma_{s,t|t-1}G^T$. Therefore, in cases when $\Sigma_{x,t|t-1}$ is not positive definite, it can be subject to Tikhonov regularization, analogically to the way in which regularization is applied to the system of normal equations in the least squares method (ridge regression) or in which Levenberg-Marquardt correction is applied within a Rapshon-Newton algorithm, hence $D_{x,t|t-1}$ can be replaced by an optimal solution of the regularized system in the sense of Tikhonov. In particular, we use a L_2 regularization to matrix $\Sigma_{x,t|t-1}$ when calculating $D_{x,t|t-1}$ and to matrix $\Sigma_{x,t|t}$ when calculating $D_{x,t|t}$. Notice, that the regularization does not affect (directly) location and scale parameters of the distribution.

Note, that when we add two *csn* variables we have to add their q-dimensions, so that the q-dimension of x_t is the sum of q-dimensions of x_{t-1} and ξ_t, therefore contribution of ξ_t to the q-dimension of x_t in every period is equal to the size of ξ_t (i.e. n_ξ), hence the q-dimension of x_t quickly increases with t and so does the q-dimension of y_t. As a result, dimension of normal integrals involved within cumulative probability distribution functions in the formula of the likelihood function quickly becomes intracktable. To our best knowledge there is no efficient way of calculating such integrals with an arbitrary correlation structure, see (Genz and Bretz (2009)). Therefore, to make the estimation operational, we work with the following approximation: $\Phi_q(z, \mu, \Sigma) \approx \prod_{j=1}^{q} \Phi_1(z_j, \mu_j, \Sigma_{jj})$, which eliminates the curse of dimensionality. Accuracy of this approximation depends on the correlation structure implied by the covariance matrix Σ. In particular, if Σ is diagonal, the result is exact.

2.4 Likelihood function

To get the likelihood function we need to pin down distribution of y_t conditional on Y_{t-1}. Since

$$x_t|Y_{t-1} \sim csn_{p,q_{t-1}+q_\xi}(\mu_{x,t|t-1}, \Sigma_{x,t|t-1}, D_{x,t|t-1}, \vartheta_{x,t|t-1}, \Delta_{s,t|t-1})$$

and $u_t \sim N(0, \Sigma_u)$, assuming that F is full row rank, distribution of Fx_t conditional on Y_{t-1} is:

$$Fx_t|Y_{t-1} \sim csn_{n,q_{t-1}+q_\xi}(\mu_{Fx,t|t-1}, \Sigma_{Fx,t|t-1}, D_{Fx,t|t-1}, \vartheta_{Fx,t|t-1}, \Delta_{Fx,t|t-1})$$

where:

$$\mu_{Fx,t|t-1} = F\mu_{x,t|t-1}, \quad \Sigma_{Fx,t|t-1} = F\Sigma_{x,t|t-1}F^T, \quad D_{Fx,t|t-1} = D_{x,t|t-1}\Sigma_{x,t|t-1}F^T\Sigma^{-1}_{Fx,t|t-1},$$

$$\vartheta_{Fx,t|t-1} = \vartheta_{x,t|t-1},$$

$$\Delta_{Fx,t|t-1} = \Delta_{x,t|t-1} + D_{x,t|t-1}\Sigma_{x,t|t-1}D^T_{x,t|t-1} - D_{x,t|t-1}\Sigma_{x,t|t-1}F^T\Sigma^{-1}_{Fx,t|t-1}F\Sigma_{x,t|t-1}D^T_{x,t|t-1}$$

and distribution of $y_t|Y_{t-1} = Fx_t + u_t|Y_{t-1}$ is:

$$y_t|Y_{t-1} \sim csn_{n,q_{t-1}+q_\xi}(\mu_{y_t|Y_{t-1}}, \Sigma_{y_t|Y_{t-1}}, D_{y_t|Y_{t-1}}, \vartheta_{y_t|Y_{t-1}}, \Delta_{y_t|Y_{t-1}})$$

where:

$$\mu_{y_t|Y_{t-1}} = \mu_{Fx,t|t-1}$$

$$\Sigma_{y_t|Y_{t-1}} = \Sigma_{Fx,t|t-1} + \Sigma_u$$

$$D_{y_t|Y_{t-1}} = D_{Fx,t|t-1}\Sigma_{Fx,t|t-1}(\Sigma_{Fx,t|t-1} + \Sigma_u)^{-1}$$

$$\vartheta_{y_t|Y_{t-1}} = \vartheta_{Fx,t|t-1}$$

$$\Delta_{y_t|Y_{t-1}} = \Delta_{Fx,t|t-1} + (D_{Fx,t|t-1}(I - \Sigma_{Fx,t|t-1}(\Sigma_{Fx,t|t-1} + \Sigma_u)^{-1}))\Sigma_{Fx,t|t-1}D^T_{Fx,t|t-1}$$

therefore:

$$p(y_t|Y_{t-1}) = \phi_n(y_t; \mu_{y_t|Y_{t-1}}, \Sigma_{y_t|Y_{t-1}})\frac{\Phi_{q_{t-1}+p_\xi}(D_{y_t|Y_{t-1}}(y_t - \mu_{y_t|Y_{t-1}}); \vartheta_{y_t|Y_{t-1}}, \Delta_{y_t|Y_{t-1}})}{\Phi_{q_{t-1}+p_\xi}(0; \vartheta_{y_t|Y_{t-1}}, \Delta_{y_t|Y_{t-1}} + D_{y_t|Y_{t-1}}\Sigma_{y_t|Y_{t-1}}D^T_{y_t|Y_{t-1}})}$$

3 The model

In this section we present equations of the model, which is then used for forecasting experiment. Outlined model structure is in line with Smets and Wouters (2007). It presents a medium-scale DSGE economy with price and wage rigidities, partial indexation of not re-ooptimized prices and wages to lagged inflation, and with stochastic price and wage mark-ups. The model also involves habit formation in consumption, a distinction between installed capital and capital services, stochastic capital utilization as well as exogenous government spending. Monetary policy is assumed to be governed by a Taylor type of policy rule, in which apart from inflation and output gap - i.e. a difference between output and its potential level, also a first difference of output gap is involved. Potential output is defined as an hypothetical output level obtained without nominal rigidities and without mark-up wedges in prices and wages. The model involves 7 shocks, which are: a TFP shock (η_a), a risk premium (η_b) an exogenous spending shock (η_g), an investment-specific technology shock (η_I), a monetary policy shock (η_r) and two mark-up shocks - a shock in a price mark-up (η_p) and a shock in a wage mark-up (η_w). The model is estimated using 7 observed variables, which are consumption, investment, output, hours worked, inflation rate, real wages and short-term nominal interest rate.

3.1 Aggregate resource constraint

The aggregate resource constraint is give by:

$$y_t = c_y c_t + i_y i_t + z_y z_t + \epsilon_t^g$$

where y_t denotes output, c_t consumption, i_t - investment, z_t level of capital utilization, and ϵ_t^g exogenous spending, e.g. government spending, for:

$$c_y = 1 - g_y - i_y$$

denoting a steady-state share of consumption in output, where g_y and i_y are steady-state ratios of exogenous spending to output and of investment to output, wherein:

$$i_y = (\gamma - 1 + \delta)k_y$$

where γ is a steady-state growth rate, δ denotes depreciation rate, and k_y stands for a steady-state ratio of capital to output. Finally,

$$z_y = R_\star^k k_y$$

where $R_\star^k$ represents a steady-state value of capital rental rate. Exogenous spending is assumed to follow:

$$\epsilon_t^g = \rho_g \epsilon_{t-1}^g + \eta_t^g + \rho_{ga} \eta_t^a$$

3.2 Dynamics of consumption

Consumption dynamics is governed by:

$$c_t = c_1 c_{t-1} + (1 - c_1)\mathrm{E}_t(c_{t+1}) + c_2(l_t - \mathrm{E}_t(l_{t+1})) - c_3(r_t - \mathrm{E}_t(\pi_{t+1}) + \epsilon_t^b)$$

where:

$$c_1 = \frac{\frac{\lambda}{\gamma}}{1 + \frac{\lambda}{\gamma}}$$

$$c_2 = \frac{(\sigma_c - 1)\frac{W_\star^H L_\star}{C_\star}}{\sigma_c(1 + \frac{\lambda}{\gamma})}$$

$$c_3 = \frac{1 - \frac{\lambda}{\gamma}}{(1 + \frac{\lambda}{\gamma})\sigma_c}$$

where $\lambda = 0$ implies no external habit formation and $\sigma_c = 1$ implies logarithmic utility. The shock to required return on assets follows:

$$\epsilon_t^b = \rho_b \epsilon_{t-1}^g + \eta_t^b$$

3.3 Dynamics of investment

Investment dynamics is governed by:

$$i_t = i_1 i_t + (1 - i_1)\mathrm{E}_t(i_{t+1}) + i_2 q_t + \epsilon_t^i$$

where:

$$i_1 = \frac{1}{1 + \beta\gamma^{1-\sigma_c}}$$

$$i_2 = \frac{1}{(1 + \beta\gamma^{1-\sigma_c})\gamma^2\phi}$$

for ϕ denoting a steady-state elasticity of capital adjustment cost function and β is household's discount factor and q_t denoting a real value of capital stock. The shock to the investment specific technology is assumed to follow:

$$\epsilon_t^i = \rho_i \epsilon_{t-1}^i + \eta_t^i$$

Arbitrage equation for the real value of capital stock q_t is give by:

$$q_t = q_1 \mathrm{E}_t(q_{t+1}) + (1 - q_1)\mathrm{E}_t(r_{t+1}^k) - (r_t - \mathrm{E}_t(\pi_{t+1}) + \epsilon_t^b)$$

where:

$$q_1 = \beta\gamma^{-\sigma_c}(1 - \delta) = \frac{1 - \delta}{R_\star^k + 1 - \delta}$$

for r_t^k being a rental rate on capital.

3.4 Production technology

Production technology is give by:

$$y_t = \phi_p(\alpha k_t^s + (1-\alpha)l_t + \epsilon_t^a)$$

where y_t denotes output, which is produced using capital services k_t^s and labour supply (i.e. hours worked) l_t. Parameter α denotes a share of capital in production and parameter ϕ_p denotes one plus the share of fixed costs in production. Total factor productivity shock is assumed to follow:

$$\epsilon_t^a = \rho_a \epsilon_{t-1}^a + \eta_t^a$$

It is assumed that capital installed becomes effective with a lag of one quarter, so that capital services used in production equal capital installed in the previous period, denoted by k_{t-1}, plus the rate of capital utilization, denoted by z_t:

$$k_t^s = k_{t-1} + z_t$$

where:

$$z_t = z_1 r_t^k$$

with:

$$z_1 = \frac{1-\Psi}{\Psi}$$

for $\Psi \in [0,1]$ denoting a positive function of the elasticity of capital utilization adjustment cost. The law of motion of capital stock is given by:

$$k_t = k_1 k_{t-1} + (1-k_t)i_t + k_2 \epsilon_t^i$$

where:

$$k_1 = \frac{1-\delta}{\gamma}$$

$$k_2 = (1 - \frac{(1-\delta)}{\gamma})(1+\beta\gamma^{1-\sigma_c})\gamma^2\phi$$

3.5 Prices

A stochastic price mark-up (marginal product of labour less real wage) is given by:

$$\mu_t^p = \alpha(k_t^s - l_t) + \epsilon_t^a - w_t$$

where w_t denotes a real wage. Prices are assumed to be sticky and partial indexation of not re-optimized prices to the level of lagged inflation is assumed. Inflation is governed by a New-Keynesian Philips curve of the form:

$$\pi_t = \pi_1 \pi_{t-1} + \pi_2 \mathrm{E}_t(\pi^k_{t+1}) - \pi_3 \mu^p_t + \epsilon^p_t$$

where:

$$\pi_1 = \frac{\iota_p}{1 + \beta\gamma^{1-\sigma_c}\iota^p}$$

$$\pi_2 = \frac{\beta\gamma^{1-\sigma_c}}{1 + \beta\gamma^{1-\sigma_c}\iota^p}$$

$$\pi_3 = \frac{1}{1 + \beta\gamma^{1-\sigma_c}\iota^p} \frac{(1 - \beta\gamma^{1-\sigma_c}\xi_p)(1 - \xi_p)}{\xi_p((\phi_p - 1)\epsilon_p + 1)}$$

for ι_p denoting a degree of indexation of not re-optimized prices to past inflation ($\iota_p = 0$ denotes no indexation), ξ_p denotes a degree of price stickiness ($\xi_p = 0$ denotes no stickiness), ϵ_p denotes a parameter of curvature of the Kimball goods market aggregator. Price mark-up shock is assumed to follow:

$$\epsilon^p_t = \rho_p \epsilon^p_{t-1} + \eta^p_t - \mu_p \eta^p_{t-1}$$

Rental rate of capital is governed by:

$$r^k_t = -(k_t - l_t) + w_t$$

A stochastic wage mark-up (real wage less marginal rate of substitution between working and consuming) is given by:

$$\mu^w_t = w_t - (\sigma_l l_t + \frac{1}{1 - \lambda}(c_t - \lambda c_{t-1}))$$

where σ_L denotes an elasticity of labour supply with respect to the real wage and λ denotes a habit formation parameter for consumption.

Nominal wages, like prices, are also assumed to be sticky and not re-optimized wages are indexed to past inflation. Real wages are governed by:

$$w_t = w_1 w_{t-1} + (1 - w_1)(\mathrm{E}_t(w_{t+1} + \mathrm{E}_t(\pi_{t+1}) - w_2 \pi_t + w_3 \pi_{t-1} - w_4 \mu^w_t + \epsilon^w_t$$

where:

$$w_1 = \frac{1}{1 + \beta\gamma^{1-\sigma_c}}$$

$$w_2 = \frac{1 + \beta\gamma^{1-\sigma_c}\iota_w}{1 + \beta\gamma^{1-\sigma_c}}$$

$$w_3 = \frac{\iota_w}{1+\beta\gamma^{1-\sigma_c}}$$
$$w_4 = \frac{1}{1+\beta\gamma^{1-\sigma_c}}\frac{(1-\beta\gamma^{1-\sigma_c}\xi_w)(1-\xi_w)}{\xi_w((\phi_w-1)\epsilon_w+1)}$$

for ι_w denoting a degree of indexation of not re-optimized wages to past inflation ($\iota_w = 0$ denotes no indexation), ξ_w denotes a degree of price stickiness ($\xi_w = 0$ denotes no stickiness), ϵ_w denotes a parameter of curvature of the Kimball labour market aggregator. Price mark-up shock is assumed to follow:

$$\epsilon_t^w = \rho_w \epsilon_{t-1}^w + \eta_t^w - \mu_w \eta_{t-1}^w$$

Finally, monetary policy equation is give by:

$$r_t = \rho r_{t-1}(r_\pi \pi_t + r_y(y_t - y_t^p)) + r_{\Delta y}((y_t - y_t^p) - (y_{t-1} - y_{t-1}^p)) + \epsilon_t^r$$

where y_t^p denotes a potential product, which is defined as a product that would prevail under no price and wage rigidity and without mark-up shocks (to prices and wages). Monetary policy shock is assumed to follow:

$$\epsilon_t^r = \rho_r \epsilon_{t-1}^r + \eta_t^r$$

4 Estimation

In this section we present estimates of the Smets and Wouters (2007) model, first under assumption of normally distributed shocks and then under assumption that shocks follow a skewed distribution. We start by presenting how priors for the estimation were assumed. As far as structural parameters and shocks' persistence parameters are concerned, in both cases prior assumptions are the same and follow Smets and Wouters (2007). As for shocks scale parameters, i.e. the σ's, in the first case these are shocks' standard deviations whereas in the second case these are shocks scale parameters, which, along with their skewness parameters imply values of shocks' standard deviations. Therefore, prior assumptions regarding σ's are different in the two discussed cases. Moreover, in the second case shocks' skewness parameters emerge, i.e. the d's, which were not present in the first case. On the other hand, the first case can be thought of as a special case of the second one, just with shocks' skewness parameters set at zero. In such a situation shocks' scale parameters become their standard deviations.

4.1 Estimation with normally distributed shocks

A priori distributions of structural parameters were assumed as shown in Table (1). A priori distributions of disturbances' persistence parameters are provided in Table (2). A priori distributions of shocks' volatilities (standard deviations) are provided in Table (3).

4.2 Estimation with skewed shocks

In what follows, we assume that shocks η_a, η_b, η_g, η_I, η_r, η_p and η_w follow a closed skew-normal distribution:

$$(\eta_{A,t}, \eta_{b,t}, \eta_{g,t}, \eta_{I,t}, \eta_{r,t}, \eta_{p,t}, \eta_{w,t}) \sim csn_{7,7}\left(\mu_\eta, \Sigma_\eta, D_\eta, \vartheta_\eta, \Delta_\eta\right)$$

where:

$$\mu_\eta = (\mu_a, \mu_b, \mu_g, \mu_I, \mu_r, \mu_p, \mu_w)$$

$$\Sigma_\eta = diag\left(\sigma_a, \sigma_b, \sigma_g, \sigma_I, \sigma_r, \sigma_p, \sigma_w\right)$$

$$D_\eta = diag\left(d_a, d_b, d_g, d_I, d_r, d_p, d_w\right)$$

$$\vartheta_\eta = (0, 0, 0, 0, 0, 0, 0)$$

$$\Delta_\eta = diag\left(1, 1, 1, 1, 1, 1, 1\right)$$

where μ_a, μ_b, μ_g, μ_I, μ_r, μ_p, μ_w denote location parameters of shocks, σ_a, σ_b, σ_g, σ_I, σ_r, σ_p, σ_w denote scale parameters of shocks (which are not their variances), and finally d_a, d_b, d_g, d_I, d_r, d_p, d_w denote skewness parameters of shocks. Since we want shocks to have a mean value equal to zero, value of $\mu_\eta = (\mu_a, \mu_b, \mu_g, \mu_I, \mu_r, \mu_p, \mu_w)^T$ is set at:

$$\mu_\eta = -\sqrt{\frac{2}{\pi}} \frac{\Sigma d}{\sqrt{1 + d^T \Sigma d}}$$

where:

$$\Sigma = diag(\sigma_a^2, \sigma_b^2, \sigma_g^2, \sigma_I^2, \sigma_r^2, \sigma_p^2, \sigma_w^2)$$

and:

$$d = (d_a, d_b, d_g, d_I, d_r, d_p, d_w)^T$$

Note that for $d_a = d_b = d_g = d_I = d_r = d_p = d_w = 0$ the normal distribution case is also contained in the above specification.

parameter	distribution	mean	std. dev.
ψ	Normal	4.00	1.50
σ_c	Normal	1.50	0.37
h	Beta	0.70	0.10
ξ_w	Beta	0.50	0.10
σ_L	Normal	2.00	0.75
ξ_p	Beta	0.50	0.10
ι_w	Beta	0.50	0.15
ι_p	Beta	0.50	0.15
Ψ	Beta	0.50	0.15
Φ	Normal	1.25	0.12
r_π	Normal	1.50	0.25
ρ	Beta	0.75	0.10
r_y	Normal	0.125	0.05
$r_{\Delta y}$	Normal	0.125	0.05
$\bar{\pi}$	Gamma	0.625	0.10
$100(\beta^{-1} - 1)$	Gamma	0.25	0.10
$\bar{L}$	Normal	0.00	0.20
$\bar{\gamma}$	Normal	0.40	0.10
α	Normal	0.30	0.05

Table 1: Priors for estimation of structural parameters.

parameter	distribution	mean	std. dev.
ρ_a	Beta	0.50	0.20
ρ_b	Beta	0.50	0.20
ρ_g	Beta	0.50	0.20
ρ_I	Beta	0.50	0.20
ρ_r	Beta	0.50	0.20
ρ_p	Beta	0.50	0.20
ρ_w	Beta	0.50	0.20
μ_p	Beta	0.50	0.20
μ_w	Beta	0.50	0.20

Table 2: Priors for estimation of disturbances' persistence parameters.

parameter	distribution	mean	std. dev.
σ_a	Inv. Gamma	0.10	2.00
σ_b	Inv. Gamma	0.10	2.00
σ_g	Inv. Gamma	0.10	2.00
σ_I	Inv. Gamma	0.10	2.00
σ_r	Inv. Gamma	0.10	2.00
σ_p	Inv. Gamma	0.10	2.00
σ_w	Inv. Gamma	0.10	2.00

Table 3: Priors for estimation of shocks' volatilities.

parameter	distribution	mean	std. dev.
$\sqrt{\sigma_a}$	Inv. Gamma	0.10	2.00
$\sqrt{\sigma_b}$	Inv. Gamma	0.10	2.00
$\sqrt{\sigma_g}$	Inv. Gamma	0.10	2.00
$\sqrt{\sigma_I}$	Inv. Gamma	0.10	2.00
$\sqrt{\sigma_r}$	Inv. Gamma	0.10	2.00
$\sqrt{\sigma_p}$	Inv. Gamma	0.10	2.00
$\sqrt{\sigma_w}$	Inv. Gamma	0.10	2.00

Table 4: Priors for estimation of shocks' volatilities.

To set priors for scale and skewness parameters (location parameters is implied *via* a relation $\mu_\eta = -\sqrt{\frac{2}{\pi}}\frac{\Sigma d}{\sqrt{1+d^T\Sigma d}}$), we first run estimation with normal shocks using a standard Kalman filter based maximum likelihood estimation procedure and then fit filtered (identified) shocks with univariate closed skew-normal distributions. Estimates of scale and skewness parameters of these univariate distributions provide modes for *a priori* distributions for skew-normal filter based estimation.

A priori distributions of shocks' scale and skewness parameters are provided in Table (4) and Table (5) respectively. In Table (4) square roots of scale parameters were presented to make them at the level comparable with levels of standard deviations. For skewness parameters relatively flat priors were assumed. This was the case, because we wanted the filter to identify skewness parameters without restrictions imposed on them *a priori*. Posterior modes for structural parameters, shocks' persistence parameters, shocks' standard deviations (normal distribution case) and scale parameters (skewed case), and of shocks' skewness parameters are presented in Tables (6) - (9) respectively. Letter 'N' denotes, that results refer to the normal distribution case, whereas letter 'S' refers to the skewed case. In case of skewness parameters (Table (9)) no abbrevations were used.

parameter	mode-N	std-N	mode-S	std-S
ψ	4.7427	1.2263	0.8644	0.1729
σ_c	0.9333	0.088	0.9721	0.0564
h	0.6701	0.0545	0.5369	0.0591
ξ_w	0.6599	0.1334	0.7699	0.0638
σ_L	1.122	0.6773	2.4902	0.6492
ξ_p	0.8294	0.0407	0.8171	0.0332
ι_w	0.5493	0.1694	0.4939	0.1712
ι_p	0.2444	0.0951	0.297	0.1151
Ψ	0.7277	0.1114	0.8169	0.073
Φ	1.3708	0.0964	1.3893	0.0885
r_π	1.6137	0.1967	1.6889	0.1997
ρ	0.823	0.0291	0.84	0.0249
r_y	0.0072	0.0109	0.007	0.009
r_{Δ_y}	0.1033	0.0184	0.07	0.0129
$\bar{\pi}$	0.6266	0.0806	0.6945	0.0749
$100(\beta^{-1}-1)$	0.1677	0.0618	0.1421	0.0547
$\bar{L}$	1.7878	1.0396	2.4758	0.7275
$\bar{\gamma}$	0.3555	0.0437	0.3643	0.0374
α	0.4749	0.0671	0.6846	0.0303

Table 5: Modes of estimation of structural parameters (N-normal shocks) and scale parameters (S-skewed shocks).

parameter	distribution	mean	std. dev.
d_a	Normal	7	6.00
d_b	Normal	-60	60.00
d_g	Normal	-2	6.00
d_I	Normal	-3	6.00
d_r	Normal	40	60.00
d_p	Normal	15	20.00
d_w	Normal	-1.5	6.00

Table 6: Priors for estimation of shocks' skewness parameters.

parameter	mode-N	std-N	mode-S	std-S
ρ_a	0.9511	0.0171	0.9224	0.0192
ρ_b	0.8821	0.032	0.8205	0.035
ρ_g	0.972	0.0118	0.9373	0.0157
ρ_I	0.491	0.1316	0.4077	0.0982
ρ_r	0.452	0.076	0.4458	0.0694
ρ_p	0.9481	0.0363	0.9121	0.0329
ρ_w	0.9448	0.0292	0.8573	0.0577
μ_p	0.8001	0.092	0.8049	0.0611
μ_w	0.8773	0.0534	0.7977	0.0612

Table 7: Modes of estimation of disturbances' persistence parameters (N-normal shocks) and scale parameters (S-skewed shocks).

parameter	mode-N	std-N	mode-S	std-S
σ_a	0.8903	0.0764	0.8864	0.0773
σ_b	0.042	0.009	0.1143	0.0222
σ_g	0.4196	0.0342	0.4054	0.0315
σ_I	0.5093	0.1003	0.9357	0.1108
σ_r	0.0981	0.009	0.0983	0.0084
σ_p	0.0843	0.0133	0.1258	0.0147
σ_w	0.5043	0.0521	0.4911	0.0449

Table 8: Modes of estimation of shocks' volatilities (N-normal shocks) and scale parameters (S-skewed shocks).

parameter	mode	std
d_a	-0.0349	0.6992
d_b	-225.3905	40.0875
d_g	-0.3619	0.6417
d_I	-10.5349	2.7389
d_r	79.128	18.4032
d_p	88.0833	12.7671
d_w	0.6119	0.6133

Table 9: Modes of estimation of shocks' skewness parameters

4.3 Forecasting experiment

The forecasting experiment was conducted for the time period starting in 2007Q1 and ending in 2012Q2, so that the recent financial crisis is involved in the sample. Forecasting properties of the Smets and Wouters (2007) model are well known, but in the original paper the experiment was conducted up till 2004Q4, so that our results extend the sample by almost 8 years. Importantly, the experiment in the original paper does not involve the financial crisis. We start by presenting graphs of out-of-sample forecasts obtained using a model with normally distributed shocks, which are presented on Figures (1) - (7), and of forecasts obtained using a model with skewed shocks, which are presented on Figures (8) - (14).

Figure 1: Recursive forecasts of output - model with normal shocks.

Table (10) presents Root Mean Square Errors (RMSE) for observed variables of forecasts obtained from the model with normally distributed shocks. The following Table (11) presents analogical results of forecasts obtained from the model with skewed shocks. To make the comparison easier two following tabes, i.e. Table (12) and Table (13), present difference in respective RMSE, i.e. RMSE of forecasts with normal shocks less RMSE of forecasts with of skewed shocks, and percentage gain of

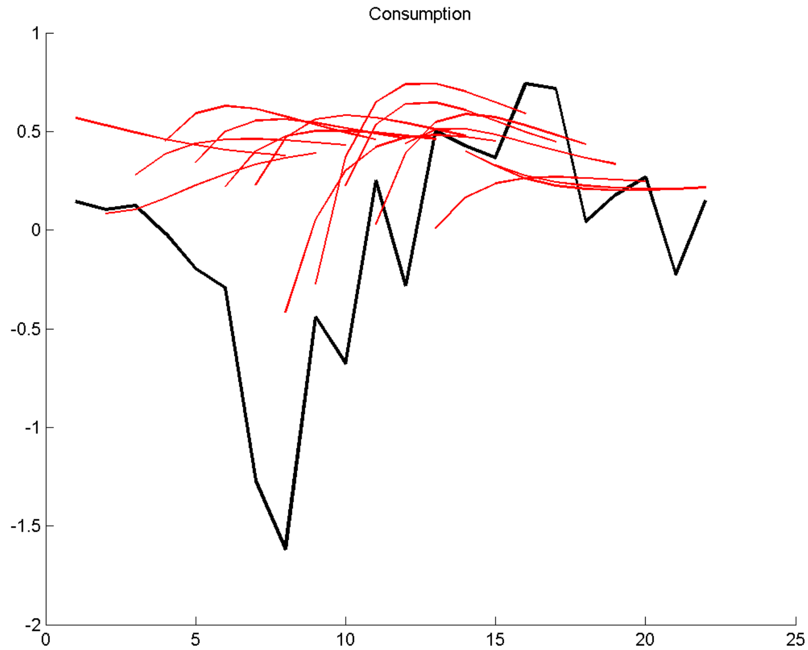

Figure 2: Recursive forecasts of consumption - model with normal shocks.

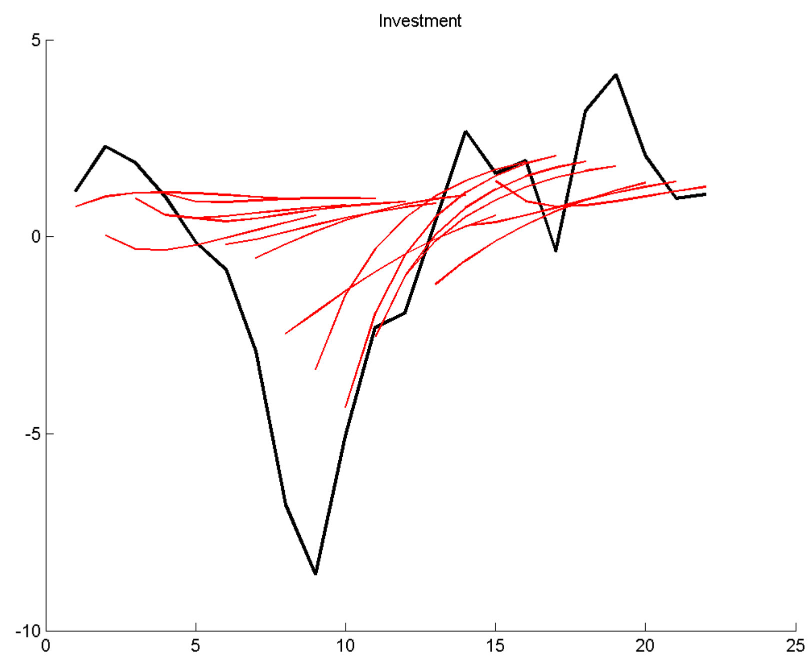

Figure 3: Recursive forecasts of investment - model with normal shocks.

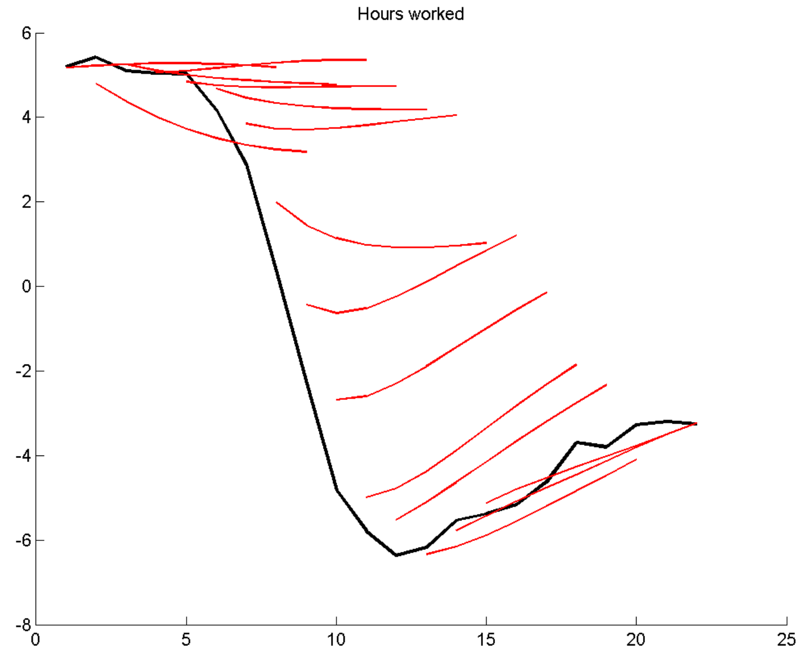

Figure 4: Recursive forecasts of hours worked - model with normal shocks.

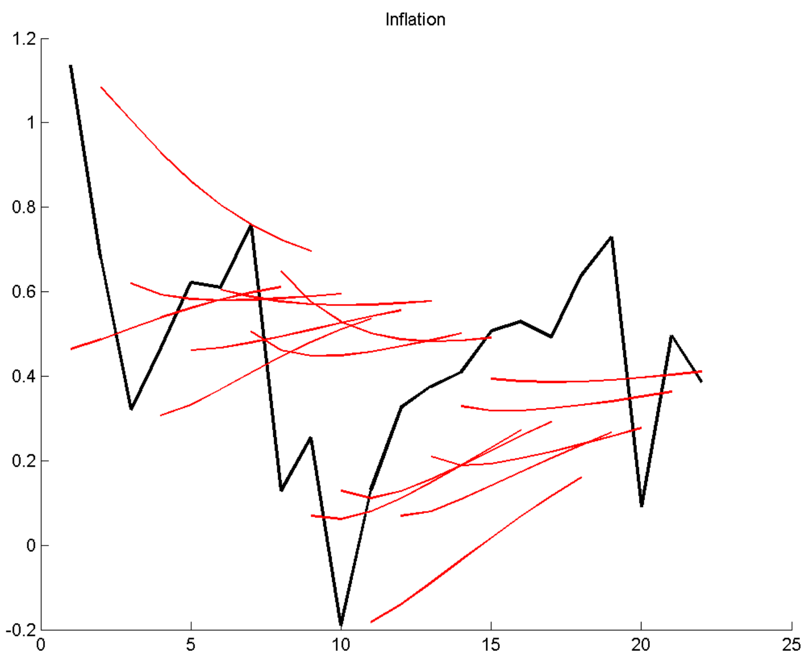

Figure 5: Recursive forecasts of inflation - model with normal shocks.

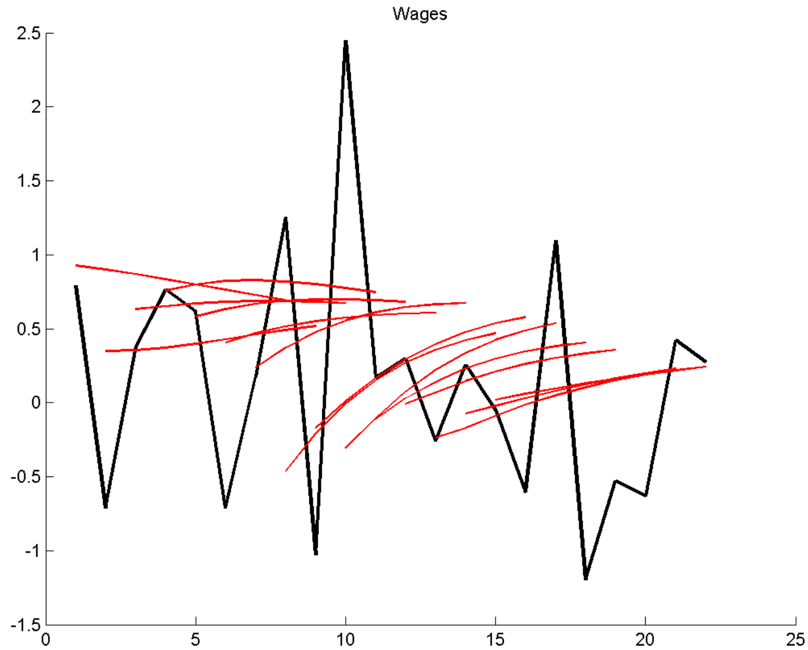

Figure 6: Recursive forecasts of real wages - model with normal shocks.

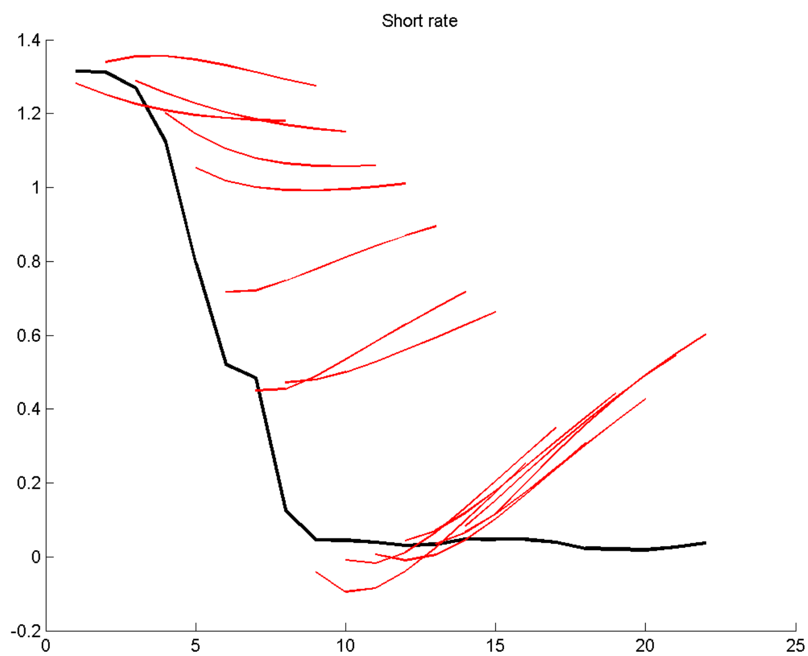

Figure 7: Recursive forecasts of short rate - model with normal shocks.

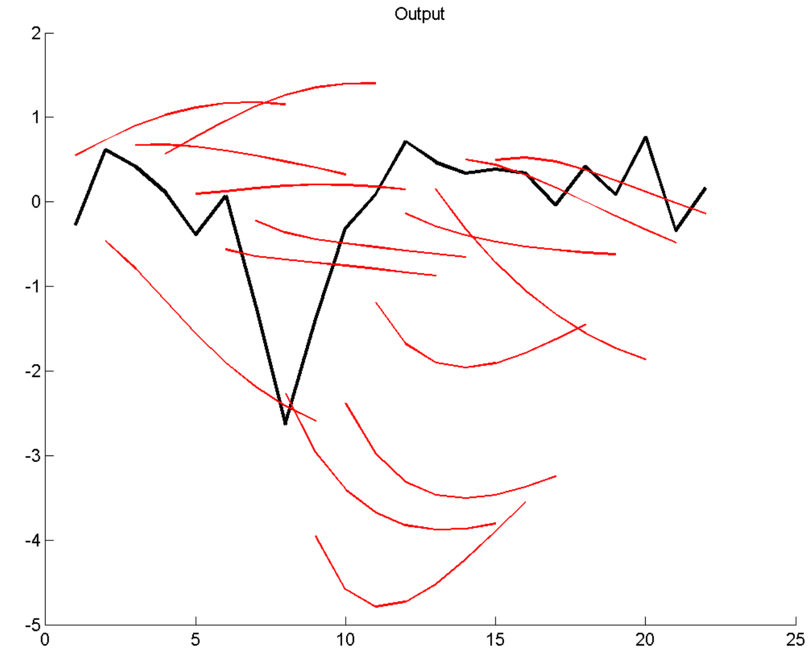

Figure 8: Recursive forecasts of output - model with skewed shocks.

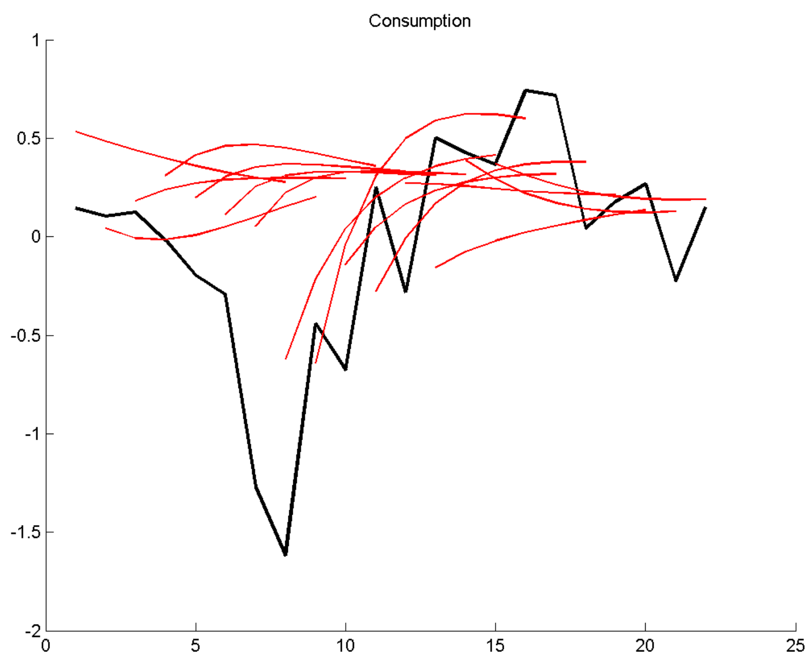

Figure 9: Recursive forecasts of consumption - model with skewed shocks.

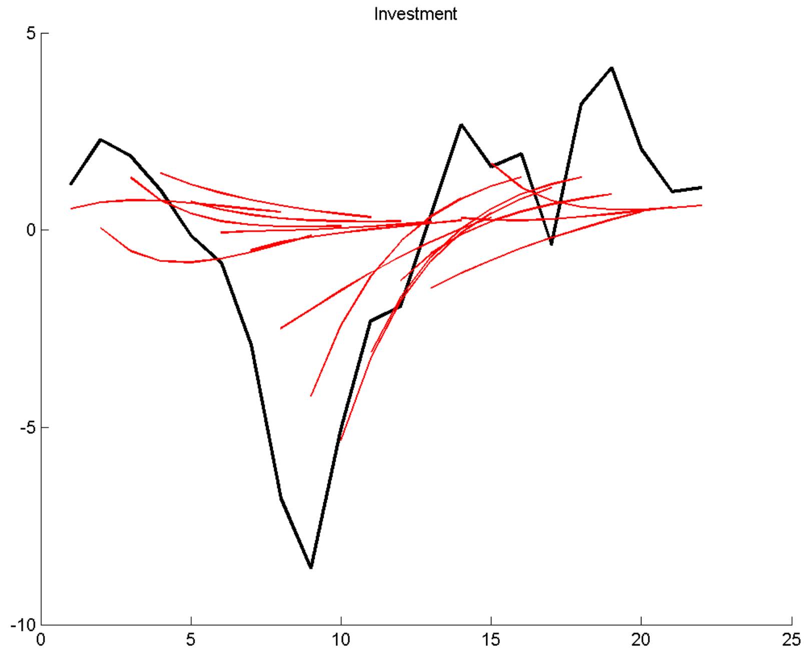

Figure 10: Recursive forecasts of investment - model with skewed shocks.

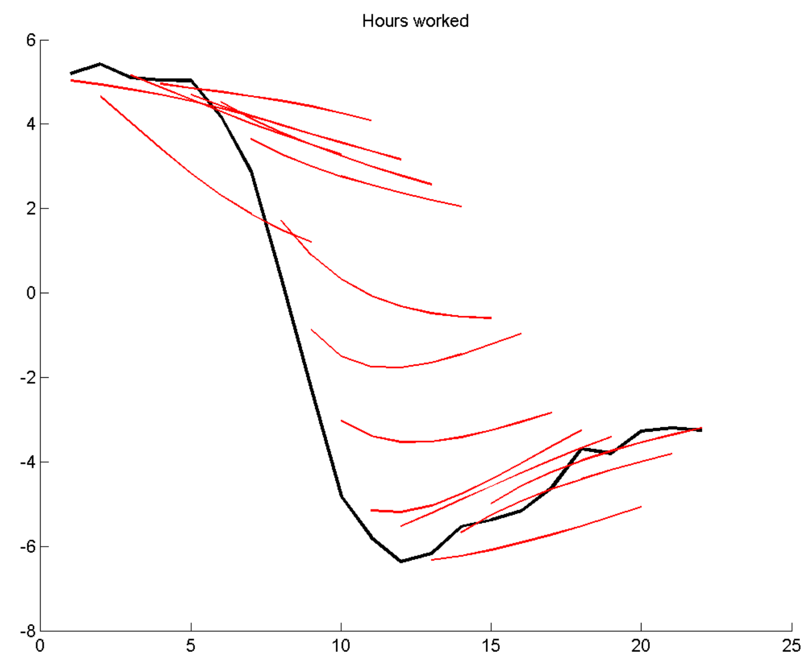

Figure 11: Recursive forecasts of hours worked - model with skewed shocks.

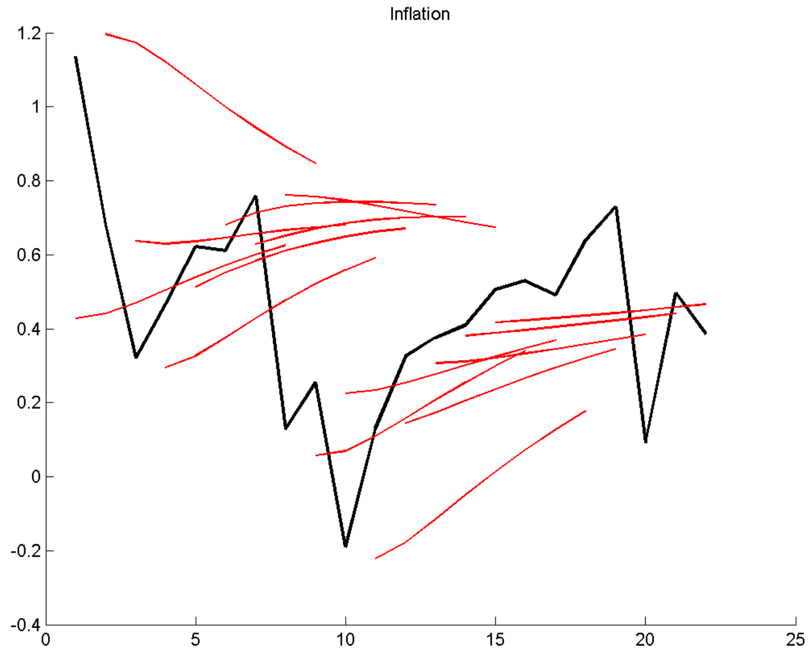

Figure 12: Recursive forecasts of inflation - model with skewed shocks.

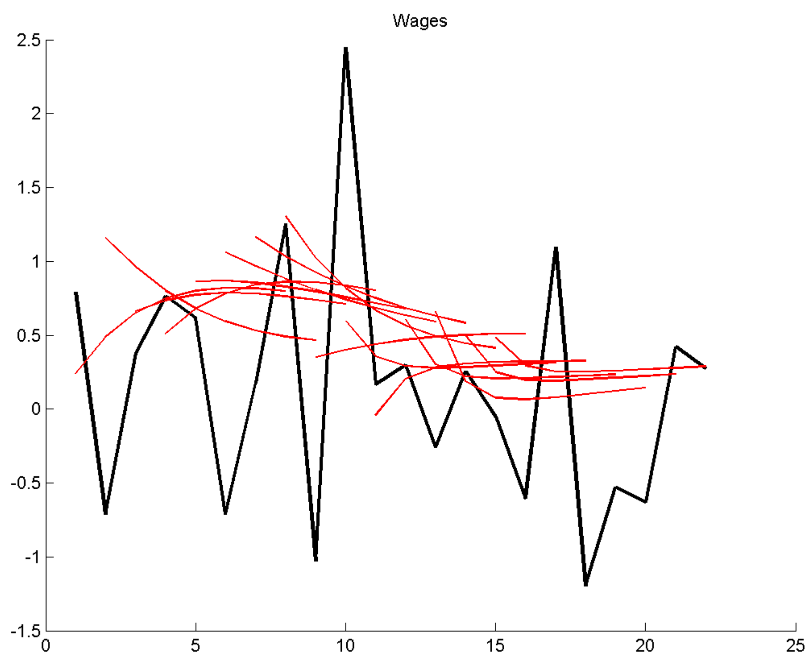

Figure 13: Recursive forecasts of real wages - model with skewed shocks.

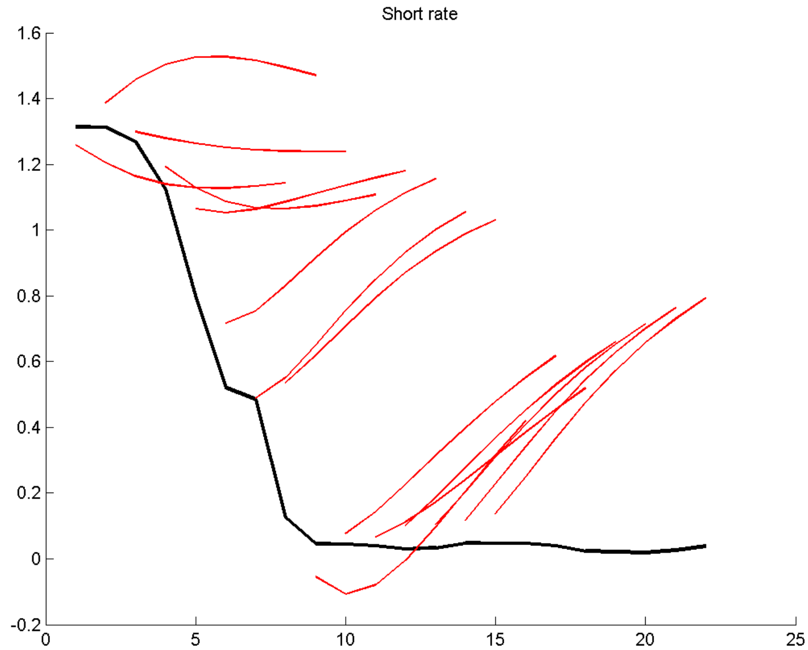

Figure 14: Recursive forecasts of short rate - model with skewed shocks.

the model with skewed shocks relative to the one with normally distributed shocks. From the tables it can be seen, that inclusion of skewness resulted in improvement of the forecasting performance of some variables - of consumption, investment and hours worked, but, on the other hand, reduced accuracy in case of the others - output, inflation, wages and the short rate. In other words, we do not find a uniform improvement over all the variables, but we find and improvement over some of them. Largest improvement is observed in the series of hours worked and it concerns, quite uniformly, all the forecasting horizons (increase in accuracy of $15\% - 19\%$). In case of consumption, improvement is more pronounced in the first three quarters ($14\% - 15\%$) and then falls to the level of $9\% - 10\%$. As far as investment is concerned, the improvement takes a U-shaped form, since in the first quarter it reaches 5% and then, over quarters from $Q2$ to $Q5$ falls to $1\% - 3\%$ and finally reaches higher levels again ($5\% - 8\%$). Deterioration of the forecasting performance is most pronounced in case of the short rate (-25% - -14%) and, apart from the first quarter, is quite uniform at the level of -25% - -22%. For wages and inflation, the level of deterioration takes an inverted U-shape (apart from an outlier in $Q3$ for wages), reaching a through at the level of -20% in $Q4$ in case of inflation and the

Forecast horizon	1Q	2Q	3Q	4Q	5Q	6Q	7Q	8Q
Output	1.03	1.60	1.87	2.05	2.18	2.01	1.80	1.80
Consumption	0.65	0.85	0.91	0.94	0.89	0.85	0.82	0.68
Investment	2.13	2.98	3.38	3.66	3.84	3.92	3.91	3.77
Hours worked	0.96	2.00	3.07	4.08	4.95	5.71	6.33	6.83
Inflation	0.30	0.29	0.33	0.32	0.33	0.34	0.36	0.35
Wages	0.96	0.92	0.94	0.90	0.93	0.97	0.94	0.94
Short rate	0.12	0.23	0.34	0.45	0.55	0.63	0.69	0.75

Table 10: RMSE for forecasts of observed variables for normally distributed shocks.

Forecast horizon	1Q	2Q	3Q	4Q	5Q	6Q	7Q	8Q
Output	1.07	1.74	2.12	2.36	2.52	2.39	2.21	2.23
Consumption	0.55	0.71	0.78	0.84	0.81	0.75	0.74	0.62
Investment	2.01	2.95	3.34	3.57	3.72	3.72	3.67	3.54
Hours worked	0.80	1.66	2.55	3.42	4.15	4.76	5.20	5.54
Inflation	0.33	0.34	0.39	0.39	0.38	0.39	0.41	0.40
Wages	0.99	0.98	0.86	0.92	0.96	0.99	0.92	0.91
Short rate	0.14	0.28	0.42	0.57	0.68	0.78	0.86	0.92

Table 11: RMSE for forecasts of observed variables for skewed shocks.

level of -7% in $Q2$ in case of wages. For wages in $Q3$ and in $Q7$ - $Q8$ we observe an improvement in fact. For output, level of forecasting performance deterioration increases with the forecasting horizon, starting from -4% in the first quarter and reaching -25% in the last one.

Forecast horizon	1Q	2Q	3Q	4Q	5Q	6Q	7Q	8Q
Output	-0.04	-0.14	-0.25	-0.31	-0.34	-0.37	-0.40	-0.43
Consumption	0.09	0.13	0.12	0.09	0.08	0.09	0.07	0.06
Investment	0.12	0.02	0.03	0.08	0.12	0.19	0.23	0.23
Hours worked	0.15	0.34	0.51	0.65	0.80	0.95	1.12	1.28
Inflation	-0.03	-0.05	-0.06	-0.06	-0.05	-0.05	-0.05	-0.05
Wages	-0.03	-0.06	0.08	-0.02	-0.03	-0.02	0.019	0.03
Short rate	-0.02	-0.05	-0.08	-0.11	-0.13	-0.15	-0.16	-0.17

Table 12: Difference in RMSE for forecasts of observed variables obtained using a model with normally distributed shocks and using a model with skewed shocks.

Forecast horizon	1Q	2Q	3Q	4Q	5Q	6Q	7Q	8Q
Output	-0.04	-0.09	-0.13	-0.15	-0.15	-0.18	-0.22	-0.24
Consumption	0.14	0.15	0.14	0.10	0.09	0.10	0.09	0.09
Investment	0.05	0.01	0.01	0.02	0.03	0.05	0.06	0.08
Hours worked	0.15	0.17	0.16	0.16	0.16	0.17	0.18	0.19
Inflation	-0.09	-0.16	-0.18	-0.20	-0.16	-0.13	-0.13	-0.13
Wages	-0.03	-0.07	0.09	-0.02	-0.03	-0.02	0.02	0.03
Short rate	-0.14	-0.22	-0.24	-0.25	-0.24	-0.23	-0.24	-0.23

Table 13: Relative improvement of RMSE of forecasts of observed variables obtained when using a model with skewed shocks (positive value represents an improvement in relative terms).

After having investigated the RMSE statistics we move on to log predictive scores of the forecasts. The state-space form of the reduced, estimated Smets and Wouters (2007) model, takes the following form:

$$\begin{aligned} y_t &= Fx_t \\ x_t &= Ax_{t-1} + B\xi_t \end{aligned}$$

for $t \in \mathcal{T} = \{1, 2, ..., T\}$, where $x_t \in \mathrm{R}^p$ and $y_t \in \mathrm{R}^n$ denote states and observed variables respectively and ξ_t follows a 7-dimensional either normal or *csn* distribution. Let $p_N(\xi_t|Y_{t-1})$ denote a normal distribution of ξ_t with parameters estimated using observed variables up to period $t-1$ and, analogically, let $p_S(\xi_t|Y_{t-1})$ denote a *csn* distribution of ξ_t with parameters estimated up to period $t-1$. Let also A_{t-1} and B_{t-1} denote model matrices with parameters estimated using observed variables up to period $t-1$ and let $p_x(x_t|Y_{t-1})$ and $p_y(x_t|Y_{t-1})$ denote distributions of states and observables, as estimated using data up to period $t-1$. In order not to abuse notation in what follows, let empirical observations be denoted in this section by v_t, $t \in \mathcal{T}$.

Let τ_1 denote a first period of the first forecasting horizon, i.e. $2007Q1$ and let τ_2 denote a first period of the last forecasting horizon, i.e. $2003Q1$. Using a state-space model form, we evaluate $p_x(x|Y_{t-1})$, $x = x_t, x_{t+1}, ..., x_{t+h-1}$, and $p_y(y_t|Y_{t-1})$, $y = y_t, y_{t+1}, ..., y_{t+h-1}$, for $t = \tau_1, \tau_1 + 1, ..., \tau_2$, where $h = 8$ denotes the forecasting horizon, in two cases - when ξ_t follows either distribution p_N or distribution p_S. This is achieved by means of a Monte Carlo simulation of draws from distribution $p_S(\xi_t|Y_{t-1})$ and from distribution $p_N(\xi_t|Y_{t-1})$, respectively, for $t = \tau_1, \tau_1 + 1, ..., \tau_2, \tau_2 + 1, ..., \tau_2 + h - 1$, with 10^4 draws for each t. Result of a Monte Carlo draw of each ξ_t is then propagated according to $x_t = A_t x_{t-1} + B_t \xi_t$, for $t, t+1, ..., t+h-1$, which in turn is plugged into $y_t = Fx_t$, for $t, t+1, ..., t+h-1$. This results in Monte Carlo estimation of distributions $p_x(x_t|Y_{t-1})$ and $p_y(y_t|Y_{t-1})$ for $t = \tau_1, \tau_1 + 1, ..., \tau_2, \tau_2 + 1, ..., \tau_2 + h - 1$.

Using simulated distribution $p_y(y_t|Y_{t-1})$ we evaluate a version of a log-predictive score statistic of forecasts using data v_t for $t = \tau_1, \tau_1 + 1, ..., \tau_2, \tau_2 + 1, ..., \tau_2 + h - 1$, for each variable $i = 1, 2, ..., 7$, i.e. for each dimension of y_t, and for each forecast horizon $j = 1, 2, ..., 8$, according to:

$$lps_{i,j} = \sum_{t=\tau_1}^{\tau_2} \ln(\int_{(v_{t+j-1},-i)} p_y(v_{t+j-1}|Y_{t-1}) d(v_{t+j-1}, -i))$$

where $(v_t, -i)$ denotes observed variable v_t without the i-th dimension. Table (14) and Table (15) presents the results. What can be observed is, that in terms of the score, normal distribution consequently tends to dominate the csn, even for variables and horizons, for which csn consequently produces better RMSE statistics. The latter one achieves better score only for terminal horizons in case of output, consumption, investment and hours worked.

5 Concluding remarks

In this paper we conducted a forecasting experiment which consists in estimating a medium scale DSGE economy, allowing the structural shocks to follow a skewed distribution. In particular, since we work within a state-space setting (a first order approximation around a steady-state), we used a closed skew-normal distribution, which is closed under most state-space setting transformations. Using data ranging from $1990Q1$ to $2012Q2$ we estimate the model and recursively verify its out-of sample forecasting properties for time period $2007Q1$ - $2012Q2$. Results are mixed in the sense, that inclusion of skewness via a CSN distribution can help forecasting some variables (consumption, investment and hours worked), but, on the other hand, results in deterioration in the others (output, inflation, wages and the short rate). Moreover, the log predictive score of a model with normally distributed shocks tends to uniformly dominate the one with skewed shocks.

References

[Genz and Bretz (2009)] Genz, A., Bretz, F. (2009): *Computation of Multivariate Normal and t Probabilities.*, Springer.

[Genton (2004)] Genton, M. G. (ed.): *Skew-elliptical distributions and their applications. A journel beyond normality.*, A CRC Press Company (2004).

[González-Farías et al. (2004)] González-Farías, G., Domínguez-Molina, J., Gupta, A. (2004): *Additive properties of skew normal random vectors.*, Journal of Statistical Planning and Inference, Vol. 126, pp. 521-534.

[Smets and Wouters (2007)] Smets, F., Wouters, R.: *Shocks and frictions in US business cycles. A Bayesian approach.*, ECB Working Paper Series No 722 (2007).