Paulsen, Kenneth Sæterhagen

Working Paper

Conditional Forecasting with DSGE Models - a Conditional Copula Approach

Working Paper, No. 4/2017

Provided in Cooperation with:
Norges Bank, Oslo

Suggested Citation: Paulsen, Kenneth Sæterhagen (2017) : Conditional Forecasting with DSGE Models - a Conditional Copula Approach, Working Paper, No. 4/2017, ISBN 978-82-7553-970-8, Norges Bank, Oslo,
https://hdl.handle.net/11250/2495589

This Version is available at:
https://hdl.handle.net/10419/210114

WORKING PAPER

Conditional forecasting with DSGE models - A conditional copula approach

NORGES BANK RESEARCH

4 | 2017

KENNETH SÆTERHAGEN PAULSEN

NORGES BANK

Working papers fra Norges Bank, fra 1992/1 til 2009/2 kan bestilles over e-post:
FacilityServices@norges-bank.no

Fra 1999 og senere er publikasjonene tilgjengelige på www.norges-bank.no

Working papers inneholder forskningsarbeider og utredninger som vanligvis ikke har fått sin endelige form. Hensikten er blant annet at forfatteren kan motta kommentarer fra kolleger og andre interesserte. Synspunkter og konklusjoner i arbeidene står for forfatternes regning.

Working papers from Norges Bank, from 1992/1 to 2009/2 can be ordered by e-mail
FacilityServices@norges-bank.no

Working papers from 1999 onwards are available on www.norges-bank.no

Norges Bank's working papers present research projects and reports (not usually in their final form) and are intended inter alia to enable the author to benefit from the comments of colleagues and other interested parties. Views and conclusions expressed in working papers are the responsibility of the authors alone.

ISSN 1502-819-0 (online)
ISBN 978-82-7553-970-8 (online)

Conditional forecasting with DSGE models - A conditional copula approach*

Kenneth Sæterhagen Paulsen†

Norges Bank,

April, 2017

Abstract

DSGE models may be misspecified in many dimensions, which can affect their forecasting performance. To correct for these misspecifications we can apply conditional information from other models or judgment. Conditional information is not accurate, and can be provided as a probability distribution over different outcomes. These probability distributions are often provided by a set of marginal distributions. To be able to condition on this information in a structural model we must construct the multivariate distribution of the conditional information, i.e. we need to draw multivariate paths from this distribution. One way to do this is to draw from the marginal distributions given a correlation structure between the different marginal distributions. In this paper we use the theoretical correlation structure of the model and a copula to solve this problem. The copula approach makes it possible to take into account more flexible assumption on the conditional information, such as skewness and/or fat tails in the marginal density functions. This method may not only improve density forecasts from the DSGE model, but can also be used to interpret the conditional information in terms of structural shocks/innovations.

Keywords: DSGE model, conditional forecast, copula

JEL classification: C53, E37, E47, E52

*This working paper should not be reported as representing the views of Norges Bank. The views expressed are those of the author and do not necessarily reflect those of Norges Bank. I have received useful comments from colleagues at Norges Bank, especially Leif Brubakk, Karsten R. Gerdrup, Junior Maih and Leif Anders Thorsrud.

†Correspondence to: Kenneth Sæterhagen Paulsen, Monetary policy department, Norges Bank, P.O. Box 1179 Sentrum, 0107 Oslo, Norway. E-mail address: Kenneth.Paulsen@Norges-Bank.no.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have been found useful to produce forecasts, see for example Adolfson *et al.* (2006), but models such as direct forecast models, ARIMAs, factor models, VARs, B-VARs, FA-VARs and others are found at some horizons and variables to have superior forecasting properties.[1] As all models are likely to be misspecified along different dimensions it may improve forecast performance if we use a larger information set than accessible to the model itself, i.e. the larger information set can be used to produce point forecasts or density forecasts of the variables of the DSGE model, be it from a pool of different models[2], judgment or a combination of the two. Then these point or density forecasts can be incorporated into the DSGE model using conditional forecasting procedures. The DSGE model can then be used to decompose the distribution of the conditional information into distributions of structural shocks/innovations identified by the DSGE model.

In this paper we suggest a method to incorporate inaccuracy in the conditional information represented by a general form of a multivariate distribution using a copula with a known correlation matrix, taken from the model, and a set of marginal distributions. The set of marginal distributions can exhibit skewness, fat tails and/or be truncated. The conditional information can be made conditional on history or not. This makes the copula based approach more flexible then other alternative approaches suggested by the literature. The resulting multivariate distribution will then be the marginal distribution of the multidimensional path of a set of variables over all horizons of the forecasts.

Incorporation of conditional information into structural models has been investigated before. Doan *et al.* (1984) use post-sample information and exploit the covariance structure of a VAR to produce conditional forecasts. Waggoner and Zha (1999) use a Bayesian rejection sampling method to compute exact finite-sample density forecasts from VARs, also taking into account parameter uncertainty. Banbura *et al.* (2014) instead use a Kalman filter(smoother) approach, which they argue improve considerably the speed of the algorithm compared to Waggoner and Zha (1999). Andersson *et al.* (2008) also extend Waggoner and Zha (1999) to take inaccurate conditional information into account and its effect on the density forecast of the unrestricted variables. They find that these densities are too narrow, if only conditioning on the central tendency.

Maih (2010) shows a way to incorporate conditional information from other models into a DSGE model.[3] What he refers to as soft conditioning, makes it possible to take into account inaccuracy in the conditional information. A strict assumption with his approach is that this inaccuracy must be represented by a truncated multivariate normal distribution.

[1] The class of competing models may also be a weighted density forecast from a pool of models, where the weights can be calculated based on out-of-sample forecasting performance, as in Aastveit *et al.* (2011).

[2] The pool of models may include the DSGE model itself.

[3] Models can also refer to judgment.

Furthermore, his approach does not give the possibility to add judgment on the unconditional marginal distributions of the information to condition on.[4] This is a limitation, as we may want to change the variance or skewness of the unconditional distribution of the conditional information without having knowledge about how this will effect the conditional distribution. When doing this we need to secure that the sampling of the paths from the multivariate distribution is conditional on past history.

The approach in this paper is also related to the approach put forward by Smith and Vahey (2016). They use a Gaussian copula model to take into account cross-sectional and serial dependence in time series. The copula model is set up using a correlation matrix that is parameterized using a latent stationary Markov vector autoregression (MVAR) model and marginal distributions that are estimated using a kernel density estimator or a skew t distribution. They document that their model compares well, in terms of out-of-sample real time forecast, with Bayesian vector autoregression models that assume symmetric marginal distribution of the data. In this paper we uses the correlation matrix from a DSGE model to parameterized the copula instead of the correlation matrix from a latent MVAR model, but in contrast to Smith and Vahey (2016), we are not able to take into account possible non-normality in data during estimation of the DSGE model.

Yet another related approach is given by Robertson *et al.* (2005). They show, using a relative entropy method, that they can impose moments restriction taken from theory on the conditional information. By using the relative entropy as a distance measure they minimizes the distance between the distribution of the conditional information and the final forecast that impose the moments restrictions. We, on the other hand, sample from the distribution of the conditional information that are in line with the correlation matrix of the DSGE model.

In section 2 we go through the general theory of copulas, and how they can be used to draw from any type of multivariate distribution with a known correlation matrix. Section 3 will outline how conditional information can be incorporated into a DSGE model for forecasting purposes. Section 4 will give some application, while we will conclude in section 5.

2 Copula Theory

A copula can be used to decompose a multivariate distribution into two parts: the marginal distributions of each variable which describes the randomness in each variable and a copula which describes the dependence between the random variables. A copula is defined as a

[4]Let us present an example to make it clear what we mean about the unconditional distribution. For an AR process we have $y_t = \lambda y_{t-1} + u_t$, where $u_t \sim N(0, \sigma)$ is the disturbance to the process. Then the distribution of y_t conditional on information up to time $t-1$ is given by $N(0, \sigma)$. This is what we refer to as the conditional distribution of y, whereas the unconditional variance of y_t is given by $N(0, \sqrt{\frac{1}{1-\lambda^2}}\sigma)$.

multivariate distribution where each marginal distribution is uniform. There are many such copulas, but in this paper we will only focus on the Gaussian copula. The cumulative distribution function (CDF) of this copula is given by

$$C_\Sigma = \Phi_\Sigma \left(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_N) \right), \tag{1}$$

where Φ_Σ is the multivariate normal CDF with correlation matrix Σ of size $N \times N$ and Φ is the univariate standard normal CDF. This means that

$$u_n \sim U(0,1) \text{ for } n \in [1, N], \tag{2}$$

where $U(0,1)$ is the uniform distribution on the interval $[0,1]$. Let the marginal distributions of the N variables be given by

$$x_n \sim F_n \text{ for } n \in [1, N], \tag{3}$$

where F_n can be any marginal CDF. By Sklar's theorem we can then represent the full multivariate CDF as

$$G_\Sigma = \Phi_\Sigma \left(\Phi^{-1}(F_1(x_1)), ..., \Phi^{-1}(F_N(x_N)) \right). \tag{4}$$

The corresponding multivariate probability density function (PDF) is given by

$$g_\Sigma = c_\Sigma \cdot f_1(x_1) \cdot \ldots \cdot f_N(x_N), \tag{5}$$

where the PDF of the copula is given by

$$c_\Sigma = \frac{1}{\sqrt{|\Sigma|}} \exp \left(-\frac{1}{2} \begin{pmatrix} \Phi^{-1}(F_1(x_1)) \\ \vdots \\ \Phi^{-1}(F_N(x_N)) \end{pmatrix}' \left(\Sigma^{-1} - \mathbf{I} \right) \begin{pmatrix} \Phi^{-1}(F_1(x_1)) \\ \vdots \\ \Phi^{-1}(F_N(x_N)) \end{pmatrix} \right), \tag{6}$$

and where $f_n(x_n)$ for $n \in [1, N]$ are the marginal PDFs.

We may also be interested in the conditional multivariate CDF or PDF. Given the Gaussian copula and the assumed marginal distributions the conditional CDF is given by

$$G_{\Sigma,\Sigma_X} | x_i, ..., x_j \in X = \frac{\Phi_\Sigma \left(\Phi^{-1}(F_1(x_1)), ..., \Phi^{-1}(F_N(x_N)) \right)}{\Phi_{\Sigma_X} \left(\Phi^{-1}(F_i(x_i)), ..., F_j(x_j)) \right))}, \tag{7}$$

where X is the set of the variables with a known value to condition on, and Σ_X is the correlation matrix of the variables in X. And the conditional PDF is given by

$$g_{\Sigma,\Sigma_X} | x_i, ..., x_j \in X = \frac{c_\Sigma}{c_{\Sigma_X}} \prod_{x_v \notin X} f_v(x_v), \tag{8}$$

where

$$c_{\Sigma_X} = \frac{1}{\sqrt{|\Sigma_X|}} \exp\left(-\frac{1}{2} \begin{pmatrix} \Phi^{-1}(F_i(x_i)) \\ \vdots \\ \Phi^{-1}(F_j(x_j)) \end{pmatrix}' \left(\Sigma_X^{-1} - \mathbf{I} \right) \begin{pmatrix} \Phi^{-1}(F_i(x_i)) \\ \vdots \\ \Phi^{-1}(F_j(x_j)) \end{pmatrix} \right) \tag{9}$$

is the PDF of the conditional copula.

If on the other hand we have the conditional marginal distributions $F_m(x_m)|x_i, ..., x_j \in X$, or in short hand notation $\Psi_m(x_m)$, and an adjusted correlation matrix $\Sigma_{\underline{X}}$, where $\underline{X}$ is the set of variables not conditioned on. Then the conditional multivariate CDF is given by

$$G_{\Sigma_{\underline{X}}}|x_i, ..., x_j \in X = \Phi_{\Sigma_{\underline{X}}} \left(\Phi^{-1}(\Psi_1(x_1)), ..., \Phi^{-1}(\Psi_M(x_M)) \right), \tag{10}$$

where $x_m \in \underline{X}$ for all $m \in [1, M]$. The corresponding PDF is

$$g_{\Sigma_{\underline{X}}}|x_i, ..., x_j \in X = c_{\Sigma_{\underline{X}}}|x_i, ..., x_j \in X \cdot \psi_1(x_1) \cdot ... \cdot \psi_M(x_M) \tag{11}$$

$$c_{\Sigma_{\underline{X}}}|x_i, ..., x_j \in X = \frac{1}{\sqrt{|\Sigma_{\underline{X}}|}} \exp\left(-\frac{1}{2} \beth' \left(\Sigma_{\underline{X}}^{-1} - \mathbf{I} \right) \beth \right) \tag{12}$$

$$\beth = \begin{bmatrix} \Phi^{-1}(\Psi_1(x_1)) \\ \vdots \\ \Phi^{-1}(\Psi_M(x_M)) \end{bmatrix}, \tag{13}$$

where $c_{\Sigma_{\underline{X}}}|x_i, ..., x_j \in X$ is the PDF of the conditional copula in this case and $\psi_m(x_m)$ is the PDF of $\Psi_m(x_m)$.

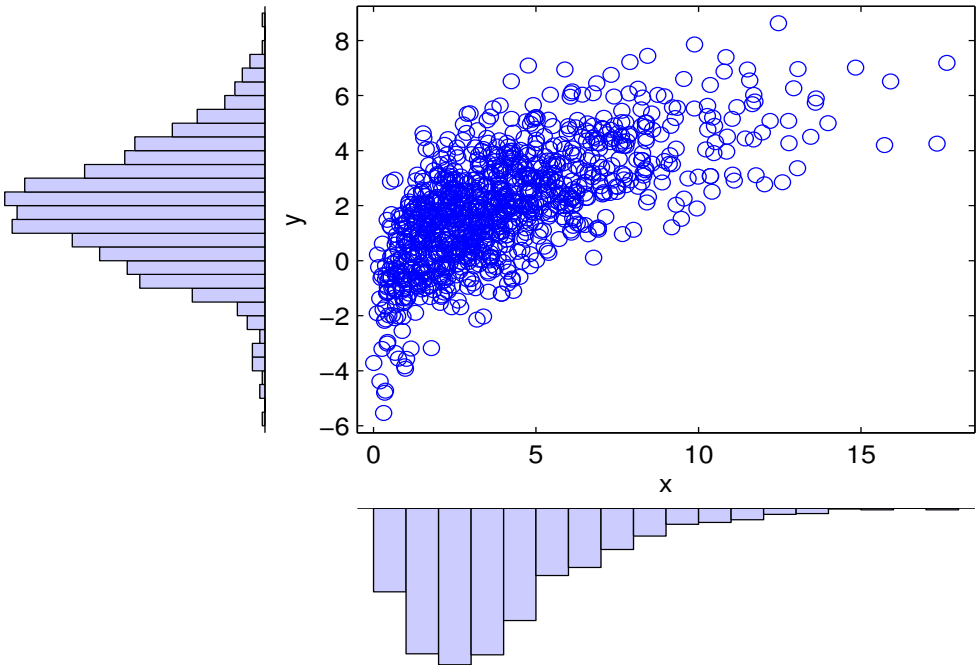
2.1 Example

Let $y \sim N(2, 2)$, $x \sim GAMMA(2, 2)$ and the linear correlation between y and x be given by

$$\Sigma = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}. \tag{14}$$

Then we can construct a multivariate distribution in y and x by using a copula. By using the algorithm presented in appendix A we can draw random numbers from this distribution. In figure 1 you can see that the draws of y is coming from the $N(2, 2)$ distribution, and that the draws of x is coming from the $GAMMA(2, 2)$ distribution. The PDF and CDF of the multivariate distribution are plotted in figures 2 and 3.

Figure 1: Drawing from a multivariate distribution using a Gaussian copula with correlation matrix Σ and marginal distributions $y \sim N(2,2)$ and $x \sim GAMMA(2,2)$.



3 Forecasting with DSGE models

As in Maih (2011) we assume that the DSGE model can be written in its log linearized form

$$E_t\left[\Theta_{-1}y_{t-1} + \Theta_0 y_t + \Theta_{+1}y_{t+1} + \Psi\varepsilon_t | I_t\right] = 0, \tag{15}$$

where I_t is the information set of the agents at time t, y_t is a $m \times 1$ vector of endogenous variables, ε_t is a $m_\varepsilon \times 1$ vector of exogenous innovations to the shock processes, which are assumed to be $\sim IID$, Θ_{-1}, Θ_0 and Θ_{+1} are $m \times m$ matrices, Ψ is an $m \times m_\varepsilon$ matrix. If this problem has a solution it can be written in a state space representation

$$y_t = Ay_{t-1} + \sum_{j=0}^{J} B_j \varepsilon_{t+j}, \tag{16}$$

where A is a $m \times m$ matrix and B_j is a $m \times m_\varepsilon$ matrix. This solution assumes that $I_t = \{\varepsilon_{t+j}, y_{t-s} | s \in [1:\infty], j \in [0:J]\ \}$, i.e. we may assume that the agents of the model react to anticipated future innovations, where J is the number of anticipated steps. The derivation of the matrices B_j can be found in Maih(2011) appendix (A).

Figure 2: PDF of a multivariate distribution that is constructed by a Gaussian copula with correlation matrix Σ and marginal distributions $y \sim N(2,2)$ and $x \sim GAMMA(2,2)$.

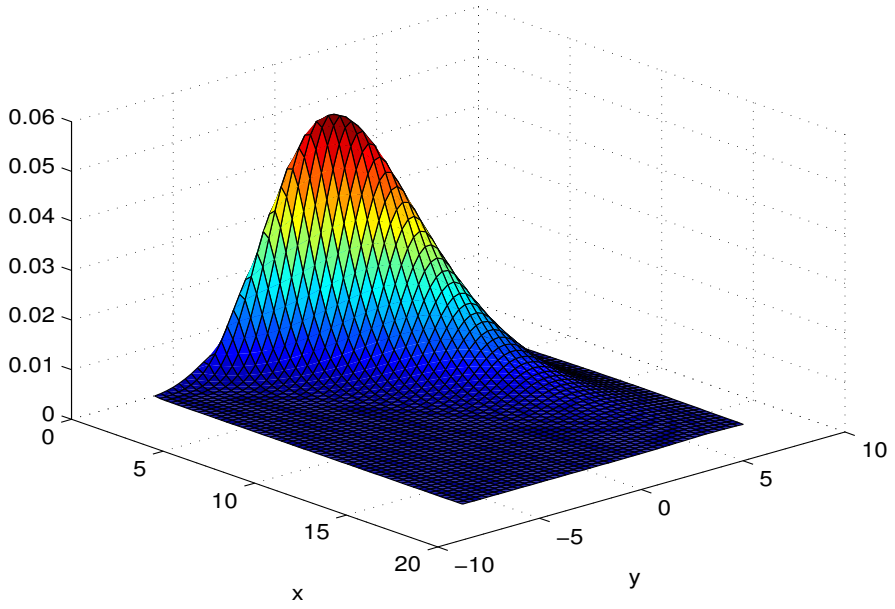
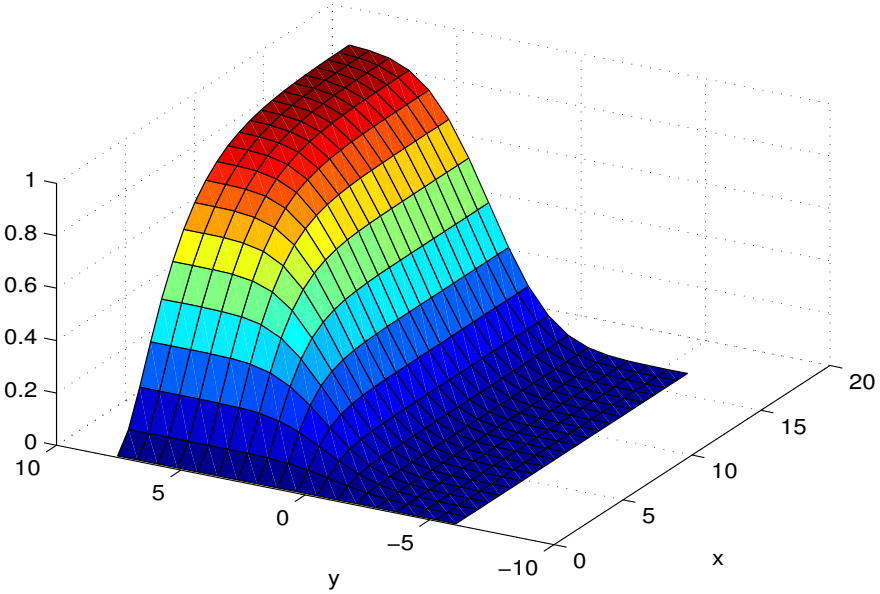


Figure 3: CDF of a multivariate distribution that is constructed by a Gaussian copula with correlation matrix Σ and marginal distributions $y \sim N(2,2)$ and $x \sim GAMMA(2,2)$.



A k step ahead forecast at time T can then be found from

$$\begin{aligned} y_{T+k} &= A^k y_T + \sum_{j=0}^{J} \sum_{s=1}^{k} A^{k-s} B_j \varepsilon_{T+j+s-1} \\ &= A^k y_T + \sum_{t=0}^{J+k-1} \Upsilon_{k,t} \varepsilon_{T+t}. \end{aligned} \tag{17}$$

Stacking all the forecast up to period $T+k$ and taking into account that conditional information may also be put on innovations, we get

$$\underbrace{\begin{bmatrix} y_{T+1} \\ \vdots \\ y_{T+k} \\ \varepsilon_{T+1} \\ \vdots \\ \varepsilon_{T+J} \\ \varepsilon_{T+J+1} \\ \vdots \\ \varepsilon_{T+J+k-1} \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} A \\ \vdots \\ A^k \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\overline{Y}} y_T + \underbrace{\begin{bmatrix} \Upsilon_{1,1} & \cdots & \Upsilon_{1,J} & 0 & \cdots & 0 \\ \Upsilon_{2,1} & \cdots & \Upsilon_{2,J} & \Upsilon_{2,J+1} & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \Upsilon_{k,1} & \cdots & \Upsilon_{k,J} & \Upsilon_{k,J+1} & \cdots & \Upsilon_{k,J+k-1} \\ \mathbf{I} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \mathbf{I} & 0 & \cdots & \cdots & \vdots \\ \vdots & 0 & \ddots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \mathbf{I} \end{bmatrix}}_{\Upsilon} \underbrace{\begin{bmatrix} \varepsilon_{T+1} \\ \vdots \\ \varepsilon_{T+J} \\ \varepsilon_{T+J+1} \\ \vdots \\ \varepsilon_{T+J+k-1} \end{bmatrix}}_{\varepsilon}, \tag{18}$$

where Y and $\overline{Y}$ has size $mk+(J+k-1)m_\varepsilon \times 1$, Υ has size $mk+(J+k-1)m_\varepsilon \times (J+k-1)m_\varepsilon$ and ε has size $(J+k-1)m_\varepsilon \times 1$. If $\varepsilon_t \sim N(0, \mathbf{I})$ as in Maih(2011) equation 26 implies that

$$Y \sim N(\overline{Y}, \Upsilon\Upsilon'). \tag{19}$$

3.1 Hard conditioning

Later we need a way to identify the innovations that should match the conditional information we put on the model. Let D be a selection matrix of size $q \times mk + (J+k-1)m_\varepsilon$, then

$$\begin{aligned} Y_c - D\overline{Y} &= D\Upsilon\widehat{\varepsilon} \\ &= R\widehat{\varepsilon}, \end{aligned} \tag{20}$$

where Y_c is the q observations of Y that is being conditioned on, i.e. it has size $q \times 1$. To be able to solve this problem the matrix R must have full rank, or else it means that there are too few innovations to match the conditional information. We can find the identified innovations to match the conditional information as

$$\widehat{\varepsilon} = R^{-1}(Y_c - D\overline{Y}). \tag{21}$$

Maih(2011) shows in proposition 1 that this estimator has the smallest variance among all linear estimators. The intuition behind equation 21 is that the difference between the conditional information, Y_c, and the unconditional forecast for the restricted variables and periods, $D\overline{Y}$, is mapped into the innovations with the smallest possible variance.

3.2 Soft conditioning using a copula

3.2.1 Conditional marginal distributions

Maih(2011) assumes that

$$Y_c \sim TN|I_T(\mu, \Omega^{I_T}, [L, H]), \tag{22}$$

where $TN|I_T$ is the truncated multivariate normal distribution with mode μ, lower bound L and upper bound H, and $\Omega^{I_T} = RR'$ is the (auto)covariance matrix and has size $q \times q$. Both are assumed to be conditional on I_T.

In this paper we assume that

$$Y_c \sim G_{\Sigma^{I_T}}|I_T, \tag{23}$$

where $\Sigma^{I_T} = RR' \oplus \sigma^{I_T}$ is the (auto)correlation matrix and has size $q \times q$.[5] σ^{I_T} is the normalization factor given by

$$\sigma^{I_T} = \sqrt{diag(RR')}\sqrt{diag(RR')}^T. \tag{24}$$

With each observation i of Y_c, call it Y_c^i, being distributed as

$$Y_c^i \sim F_i|I_T. \tag{25}$$

For any marginal distribution $F_i|I_T$ the steps of the algorithm are as follows

1. Make a draw from the distribution $G_{\Sigma^{I_T}}|I_T$ as explained in appendix A.
2. Identify the innovations to match the draw from step 1, using hard conditioning.
3. Replicate step 1 and 2 Q number of times.

[5] $\oplus$ is the element wise division operator.

4. After step 3 the distributions of the innovations to match the conditional information are identified, and forecasts can be made based on the Q number of simulated points from these distributions.

3.2.2 Unconditional marginal distributions

On the other hand, if the marginal distributions to condition on are not conditioned on the information set I_T, then we need to find the (auto)correlation matrix not conditioned on I_T. Let us abbreviate it as Σ. It can be found by using the theoretical counterpart from the model or it can be calculated empirically using historical data

$$\Lambda = \begin{bmatrix} \Lambda_0 & \Lambda_1 & \cdots & \cdots & \cdots & \Lambda_{k+j-1} \\ \Lambda_1 & \ddots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \Lambda_0 & \Lambda_1 & \cdots & \Lambda_{k-1} \\ \vdots & \vdots & \Lambda_1 & \Lambda_0 & \cdots & \Lambda_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \Lambda_{k+j-1} & \cdots & \Lambda_{k-1} & \Lambda_{k-2} & \cdots & \Lambda_0 \end{bmatrix}, \tag{26}$$

where j is the number of periods back in time to condition on, Λ_i for $i \in [0, k+j-1]$ is the autocorrelation matrix between the variables to condition on at lag i[6], and $\Sigma = D\Lambda D'$. The theoretical formula of the contemporaneous covariance matrix ($\beth_0$) from the model can be found by solving the Lyapunov equation

$$\beth_0 = A\beth_0 A' + BB'. \tag{27}$$

Given $\beth_0$ we get

$$\Lambda_0 = \beth_0 \oplus \lambda_0, \tag{28}$$

where

$$\lambda_0 = \sqrt{diag(\Lambda_0)}\sqrt{diag(\Lambda_0)}'. \tag{29}$$

Λ_i for $i \in [1, k+j-1]$ can be found by

$$\Lambda_i = A^i \Lambda_0. \tag{30}$$

Let us define Y_H as all the historical given observation we want to condition on and Σ_{Y_H} be the $j \times j$ symmetric upper part of Σ. By this we can find the final conditional multivariate distribution

[6] Λ_0 is then the contemporaneous correlation matrix.

$$Y_c \sim G_{\Sigma,\Sigma_{Y_H}} | y_h \in Y_H. \tag{31}$$

With each observation i of Y_c, call it Y_c^i being distributed as

$$Y_c^i \sim F_i. \tag{32}$$

The algorithm from the last section is also valid in this case, except that the distribution $G_{\Sigma^{I_T}}|I_T$ is substituted by $G_{\Sigma,\Sigma_{Y_H}}|y_h \in Y_H$.

4 Application

4.1 Lubik and Schorfheide (2007)

In this paper we focus on a slightly modified version of the small open economy DSGE model developed by Lubik and Schorfheide (2007). It is a model of aggregate output (y_t), domestic inflation (π_t), first difference of the nominal exchange rate (Δe_t), nominal interest rate (r_t), output abroad (y_t^*) and inflation abroad (π_t^*). The main equation of the model are:

The demand equation:

$$y_t = E_t y_{t+1} - \gamma(r_t - E_t\pi_{t+1}) - \rho_z z_t - \alpha\gamma z_{t+1}^q + \theta\Delta y_{t+1}^*. \tag{33}$$

The Phillips curve:

$$\pi_t = \beta E_t\pi_{t+1} + \alpha\beta z_{t+1}^q - \alpha z_t^q + \frac{\kappa}{\gamma}(y_t - \overline{y_t}) + z_t^\pi. \tag{34}$$

Purchasing power parity:

$$\pi_t = \Delta e_t + (1-\alpha)z_t^q + \pi_t^*. \tag{35}$$

Monetary policy:

$$r_t = \alpha_r r_{t-1} + (1-\alpha_r)(\phi_1\pi_t + \phi_2 y_t) + z_t^r. \tag{36}$$

Definitions:

$$\Delta y_t^* = y_t^* - y_{t-1}^* \tag{37}$$

$$\overline{y_t} = -\theta y_t^*. \tag{38}$$

Expectations:

$$E_t y_{t+1} = y_{t+1} \tag{39}$$

$$E_t\pi_{t+1} = \psi\pi_{t+1} + (1-\psi)\pi_{t-1}. \tag{40}$$

The variables π_t^* and y_t^* are seen as exogenous from the view of the domestic economy. We model these variables as AR(1) processes:

$$\pi_t^* = \rho_{\pi^*}\pi_{t-1}^* + \sigma_{\pi^*}\epsilon_t^{\pi^*} \tag{41}$$

$$y_t^* = \rho_{y^*}y_{t-1}^* + \sigma_{y^*}\epsilon_t^{y^*}. \tag{42}$$

There are also 4 domestic shocks in the model, productivity shock (z_t), price markup shock (z_t^π), terms of trade shock (z_t^q) and monetary policy shock (z_t^r). All shocks follows AR(1) processes:

$$z_t = \rho_z z_{t-1} + \sigma_z\epsilon_t^z \tag{43}$$

$$z_t^\pi = \rho_{z^\pi} z_{t-1}^\pi + \sigma_{z^\pi}\epsilon_t^{z^\pi} \tag{44}$$

$$z_t^q = \rho_{z^q} z_{t-1}^q + \sigma_{z^q}\epsilon_t^{z^q} \tag{45}$$

$$z_t^r = \rho_{z^r} z_{t-1}^r + \sigma_{z^r}\epsilon_t^{z^r}. \tag{46}$$

Where ρ_x is the autocorrelation coefficient, $\epsilon_t^x \sim N(0,1)$, and σ_x is the the standard deviation of the innovations of the AR(1) processes.

The parameters of the models are as follows, $1-\psi$ is the degree of how backward looking the agents of the model are when forming inflation expectations, τ is the intertemporal substitution elasticity, $1 > \alpha > 0$ is the import share, $\kappa > 0$ is a function of underlying structural parameters, such as labor supply and demand elasticities and parameters capturing the degree of price stickiness. α_r, ϕ_1 and ϕ_2 are monetary policy parameters of the Taylor type rule, and r_{ss} is the steady-state level of the real interest rate. The other parameters are function of the others:

$$\beta = e^{\frac{-r_{ss}}{400}} \tag{47}$$

$$\gamma = \tau + \alpha(2-\alpha)(1-\tau) \tag{48}$$

$$\theta = \frac{\alpha(2-\alpha)(1-\tau)}{\tau}. \tag{49}$$

See Lubik and Schorfheide (2007) for more details.

4.1.1 Data and estimation

The model is estimated using Norwegian and trade weighted data on 6 variables (the observables). The observables are listed in table 1, where also the transformation of each series is documented.

Table 1: Data description and transformation.

Variable	Description	Transformation
GDP (y_t)	GDP, Norway. Data from Statistics Norway.	Taken log and HP-filtered with lambda 3000.
GDP abroad (y_t^*)	Trade weighted GDP. Data from Thomson Reuter, Statistics Norway and Norges Bank.	Taken log and HP-filtered with lambda 3000.
Inflation (π_t)	Core CPI index. Data from Statistics Norway.	Taken the log-approximated growth rate of the and subtracted the mean.
Inflation abroad (π_t^*)	Import weighted core CPI index for 25 of Norway's trading partners. Data from Statistics Norway, Thomson Reuter and Norges Bank.	Taken the log-approximated growth rate of the and subtracted the mean.
Key policy rate (r_t)	Data from Norges Bank.	Filtered by a linear trend.
Money market interest rate abroad (r_t^*)	Import weighted money market interests for 7 of Norway's trading partners. Data from Statistics Norway, Thomson Reuter and Norges Bank.	Filtered by a linear trend.
Exchange rate (e_t)	Import weighted nominal exchange rate. Data from Thomson Reuter and Norges Bank.	Taken the log-approximated growth rate and subtracted the mean.

First some parameters of the model are calibrated. $\alpha = 0.3$ to fit the import share observed in the data, while $r_{ss} = 1.5$ consistent with the observed productivity in Norway. The rest of the parameters of the model are estimated using Bayesian techniques as in Smets and Wouters (2007). See the table 2 for the selected priors and estimation results.

Table 2: Parameter estimation.

Name	Prior	Prior mean	Prior std.	Post. mode	Post. std.
α_r	BETA	0.7	0.15	0.7677	0.0327
κ	NORMAL	0.5	0.1	1.2238	0.2069
ϕ_1	NORMAL	1.5	0.25	2.1048	0.1814
ϕ_2	BETA	0.3	0.15	0.0854	0.0431
ψ	BETA	0.3	0.15	0.1629	0.1824
τ	BETA	0.5	0.2	0.8973	0.0725
σ_z	INVGAMMA	0.01	INF	0.0036	0.0007
σ_{z^π}	INVGAMMA	0.01	INF	0.0152	0.0022
σ_{z^q}	INVGAMMA	0.02	INF	0.0308	0.0023
σ_{z^r}	INVGAMMA	0.003	INF	0.0014	0.0001
σ_{π^*}	INVGAMMA	0.005	INF	0.0034	0.0002
σ_{y^*}	INVGAMMA	0.01	INF	0.0050	0.0004
ρ_z	BETA	0.5	0.2	0.8063	0.0472
ρ_{z^π}	BETA	0.5	0.2	0.8541	0.0799
ρ_{z^q}	BETA	0.5	0.2	0.1402	0.0704
ρ_{z^r}	BETA	0.5	0.2	0.3944	0.0747
ρ_{π^*}	BETA	0.5	0.2	0.2294	0.0884
ρ_{y^*}	BETA	0.5	0.2	0.8885	0.0372

4.2 Benchmark

To keep the exercises as simple as possible we discard parameter uncertainty in this paper.[7] As a benchmark we will in this section present density forecasts of some key variables. We create these by simulating the vector of innovations to the shock processes from the distribution $\varepsilon_{t+h} \sim N(0, \mathbf{I})$, see figures 4 and 5 for the resulting forecasts.[8]

To show that the algorithm given in section 3.2.1 works we will condition on the density forecast of the observables from the benchmark forecast, and show that the identified densities of innovations in this exercise replicate the distribution of ε_{t+h}. From figures 6 and 7 we see that we are able to do just that.

Seen from figures 4 and 5, it may be argued that the density forecasts are to wide for some variables, e.g. inflation, as the 90 percentile is much wider than the historical variation in the series.[9]

4.3 Condition on forecasts from other models

In this section we illustrate how we can use the algorithm to interpret the forecasts from a pool of vector autoregression (VAR) models in terms of structural shocks/innovations in

[7] To account for parameter uncertainty you can introduce an additional loop over the algorithm presented in this paper using the posterior draws.

[8] 4000 draws are used for all simulations.

[9] This is the same as saying that the model generates too much volatility compared to the historical variation in the data.

the DSGE model. We proceed in the following steps. In section 4.3.1 we describe a way to produce combined forecasts of many VAR models using a forecast performance criterion. Next we present the results in section 4.3.2.

4.3.1 Pooled VAR forecast

For each observable of our DSGE model, we create a pool of VAR models to select from. The pool of VAR models are created by adding 0-6 other variables from the set of the other observables plus the Money market interest rate abroad (r_t^*) (without replications). The reason to pick only 6 is to keep the problem parsimoniously. We also replicate all these models for 1-3 lags. That makes 192 models for each observable. We then evaluate the out-of-sample recursive forecast performance of each model using mean log scores. The 20 ($= M$) best performing models at forecasting horizon 1 are selected for each observable using the following score

$$score_{h,v,T,m} = exp\left(\frac{\sum_{s=t}^{T} log(\varphi_{h,v,s,m}(Y_{s+h,v}))}{T-t+1}\right), \tag{50}$$

where we have defined t to be the start period of the evaluation, T the end period of the evaluation, the density forecast at time s at horizon h for model m for observable v by $\varphi_{h,v,s,m}$ and the actual data at time s for variable v as $Y_{s,v}$.[10] $\varphi_{h,v,s,m}$ is a distribution, while $Y_{s,v}$ has size 1 x 1. $v \in [1, O]$, where O is the number of observables. We then combine the forecasts using a linear opinion pool

$$\varphi_{h,v,s}^{CF} = \sum_{m=1}^{M} w_{h,v,s,m}\varphi_{h,v,s,m}. \tag{51}$$

$w_{h,v,s,m}$ is the weight on each model in the combined density, and is calculated as

$$w_{h,v,s,m} = \frac{score_{h,v,s,m}}{\sum_{m=1}^{M} score_{h,v,s,m}}. \tag{52}$$

We are now finished with producing the density forecast we want to decompose using the DSGE model.

4.3.2 Results

As noted in section 4.2 some of the density forecasts from the DSGE model seem to be too wide compared with the historical distributions of the variables. In figure 8 you can see the corresponding density forecasts of the pool of VAR models. Figure 9 display the identified distribution of each innovation to match the conditional information. As seen from the figure there are especially some innovations that have much lower variance; monetary policy ($\epsilon_t^{z^r}$),

[10] $m \in [1, M]$

price markup ($\epsilon_t^{z^\pi}$) and inflation abroad ($\epsilon_t^{z^{\pi*}}$). While on the other hand the terms of trade ($\epsilon_t^{z^q}$) and productivity (ϵ_t^z) innovations has been identified to have higher variance. When it comes to interpreting the difference in the mean of the identified distributions, we see that the those are quite small. This is an indication that the point forecasts from the pool of VARs are not too different to those stemming from the DSGE model.

It is still unclear how much of the change in the distribution of the density forecasts of the observables that can be attributed to each innovation. Forecast error variance decomposition can be used to illustrate this, but as the identified distribution of the innovation may not be symmetric, it is necessary to use another approach. In figures 10 and 11 we decompose the Xth percentile of the density forecasts of the key policy rate for the benchmark and the pool of VARs respectively. The decomposition is constructed by calculating the Xth percentile of the variable of interest based on simulations from one innovation at the time. Then the sum of the contributions are scaled to sum to the Xth percentile of density forecasts of the variable of interest. In our example, $X = 90$.

The difference in the decomposition of the pool of VARs and DSGE forecasts can be found in figure 12. From the figure we get a clear picture of which innovations that contributes to the reduced variance of the density forecasts, and it corresponds well to earlier discussion of how the identified distributions of the innovation changed.

Care should be taken when interpreting the results using the Lubik Schorfheide (2007). The reason is that it is important for the identification of the distribution of the structural innovation that the model can replicate the correlation in the data it wants to describe. Table 4 show that this is not the case. This fact is not only important when interpreting conditional information, but should be the prime objective for any exercise that uses a DSGE model. The focus of this paper is not to build the best DSGE model for the Norwegian economy, but to illustrate the algorithm in the simplest framework possible.

5 Conclusion

In this paper we have presented an algorithm that can be used to incorporate conditional information in terms of distributions into a structural model. Few restrictions on the conditional information are assumed, as the approach handle conditional information with fat tailed, skewed or truncated marginal distributions.

We have tested the algorithm by setting up a pool of different VARs and conditioned on the the combined density forecasts from these models in a DSGE model. Doing this we have shown that we can decompose these forecasts into probability distributions of the innovations of the DSGE model, and give the forecasts a structural interpretation in terms of shocks. This can be useful for assessment of risk in policy analysis, as it is important to identify the important risk factors when forming policy decisions.

A caveat, however, is that the analysis in this paper assumes that agents of the model

do not take uncertainty into account when forming their decisions. This important question is beyond the reach of this paper as that require a departure from the linearization of the DSGE model.

References

Aastveit, K. A., Gerdrup, K., Jore, A. S. and Thorsrud, L. A. (2011) Nowcasting GDP in real-time: A density combination approach, Working Paper 2011/11, Norges Bank.

Adolfson, M., Lindè, J. and Villani, M. (2006) Forecasting performance of an estimated open economy DSGE model for the euro area, Working Papers 203, Sveriges Riksbank.

Andersson, M. K., Palmqvist, S. and Waggoner, D. F. (2008) Density-conditional forecasts in dynamic multivariate models, Working Paper 247, Sveriges Riksbank.

Banbura, M., Giannone, D. and Lenza, M. (2014) Conditional forecast and scenario analysis with vector autoregressions for large corss-sections, Working Paper No 1733, European Central Bank.

Doan, T., Litterman, R. and Sims, C. (1984) Forecasting and conditional projection using realistic prior distributions, *Econometric Reviews*, **3(1)**, pp. 1–100.

Lubik, T. A. and Schorfheide, F. (2007) Do central banks respond to exchange rate movements? a structural investigation, *Journal of Monetary Economics*, **54(4)**, pp. 1069–1087.

Maih, J. (2010) Conditional forecasts in dsge models, Working Paper 2010/07, Norges Bank.

Robertson, J. C., Tallman, E. and Whiteman, C. H. (2005) Forecasting using relative entropy, *Journal of Money, Credit and Banking*, **37(3)**, pp. 383–401.

Smets, F. and Wouters, R. (2007) Shocks and frictions in us business cycles, a bayesian dsge approach, Working Paper No 722, European Central Bank.

Smith, M. S. and Vahey, S. P. (2016) Asymmetric forecast densities for u.s. macroeconomic variables from a gaussian copula model of cross-sectional and serial dependence, *Journal of Business & Economic Statistics*, **34(3)**, pp. 416–434.

Waggoner, D. F. and Zha, T. (1999) Conditional forecasts in dynamic multivariate models, *The Review of Economics and Statistics*, **81(4)**, pp. 639–651.

A Drawing random numbers from a multivariate distribution using a Copula

To make draws from the multivariate distribution G_Σ in N variables the following algorithm may be used:

1. Draw Q number of observations from the multivariate normal distribution Φ_Σ. Abbreviate the draws from this distribution for variable $n \in [1, N]$ as y_n, which then has size $Q \times 1$.

2. For each variable $n \in [1, N]$ map the observation found in step 1 to the interval $[0, 1]$ using $u_n = \Phi(y_n)$, where Φ is the standard normal CDF.

3. Map to the final draws from the marginal distribution of variable $n \in [1, N]$ by $x_n = F_n^{-1}(u_n)$, where F_n is the CDF of the marginal distribution of variable n.

If you want to draw from the multivariate distribution $G_{\Sigma_{\underline{X}}}|x_i, ..., x_j \in X$ instead of G_Σ you must substitute Φ_Σ with $\Phi_{\Sigma_{\underline{X}}}$ and F_n with $F_n(x_n)|x_i, ..., x_j \in X$ in the above algorithm.

To make draws from the conditional multivariate distribution $G_{\Sigma,\Sigma_{22}}|x_k \in X_2$, where the hard conditional information is given by a, the following algorithm may be used:

1. Partition the variables into the sets X_1 and X_2. Where X_2 is the set of variables to condition on with dimension I, and X_1 is the set of variable not to condition on with dimension $N - I$. If we re-order them according to $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, i.e. with all $x_k \in X_2$ stacked in a vector x_2 and with all $x_i \in X_1$ stacked in a vector x_1, we can partition the correlation matrix as $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}$. Σ_{11} has size $N - I \times N - I$, Σ_{21} has size $I \times N - I$, Σ_{12} has size $N - I \times I$ and Σ_{22} has size $I \times I$.

2. Map the conditional information a to the normal distribution by $\mu_k = \Phi^{-1}(F_k(a_k))$ for all $a_k \in a$, and stack the μ_ks in a vector μ with the same order as in a.

3. Adjust the mean and the correlation matrix of the variables not conditioned on. $\overline{\mu} = \Sigma_{12}\Sigma_{22}^{-1}\mu$ and $\overline{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

4. Draw Q number of observations from the multivariate normal distribution $\Phi(\overline{\mu}, \overline{\Sigma})$. Abbreviate the draws from this distribution for variable $n \in [1, N - I]$ as y_n, which then has size $Q \times 1$.

5. For each variable $n \in [1, N - I]$ map the observation found in step 3 to the interval $[0, 1]$ using $u_n = \Phi(y_n)$, where Φ is the standard normal CDF.

6. Map to the final draws from the marginal distribution of variable $n \in [1, N - I]$ by $x_n = F_n^{-1}(u_n)$, where F_n is the CDF of the marginal distribution of variable n.

B Tables

Table 3: Second order moments generated by the modified Lubik and Schorfheide (2007) model versus empirical

	Δe_t	r_t	π_t	π_t^*	y_t	y_t^*
Standard errors	2.58	0.89	1.44	0.58	1.58	1.15
Standard errors (empirical)	2.33	0.37	0.24	0.35	1.53	1.26

See table 1 for the definitions of Δe_t, r_t, π_t, π_t^*, y_t and y_t^*. The empirical second order moments of the data used to estimate the model is given in the lines where (empirical) is provided.

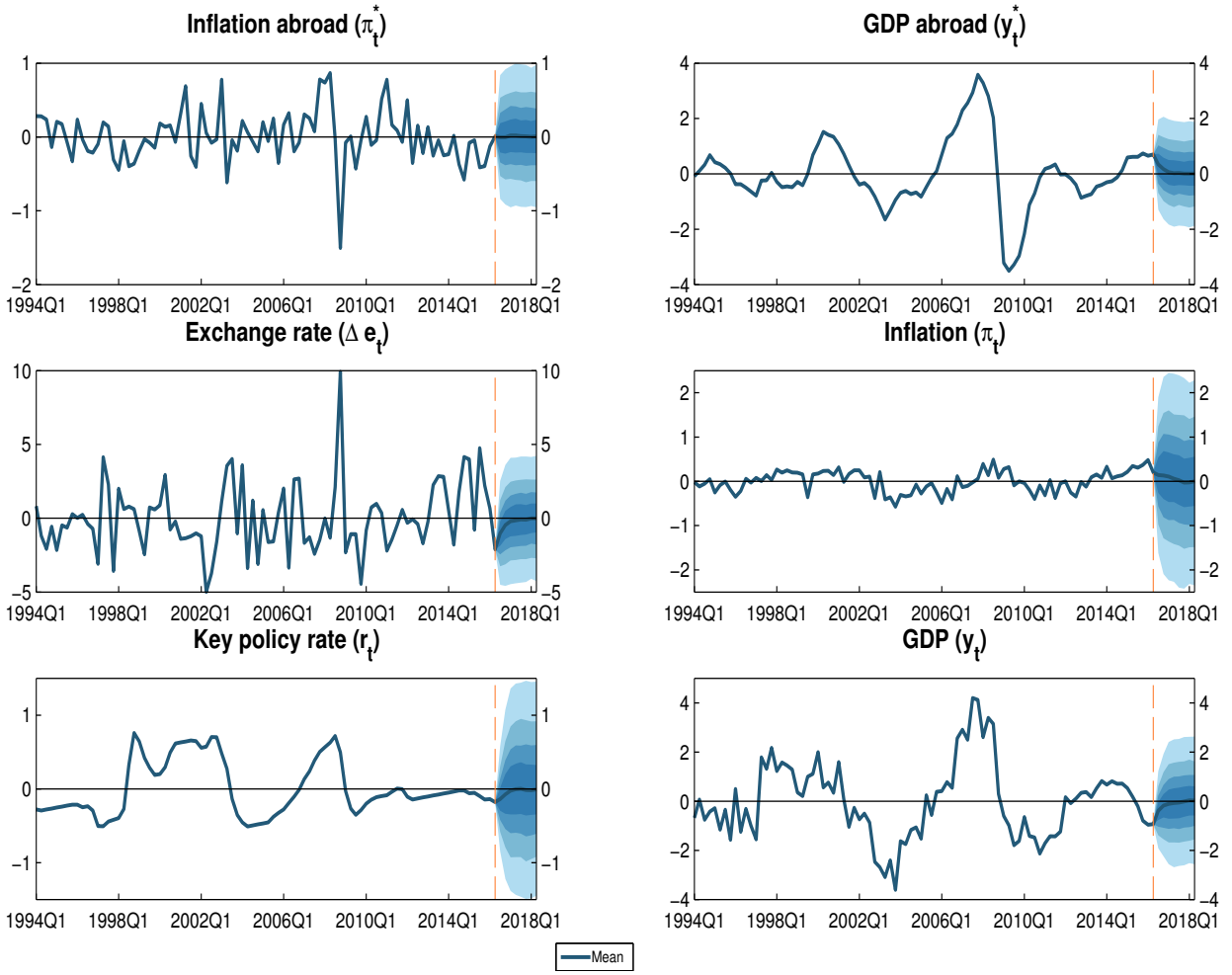
Table 4: Correlations generated by the modified Lubik and Schorfheide (2007) model versus empirical

	Δe_t	r_t	π_t	π_t^*	y_t	y_t^*
Δe_t	1.00	0.65	0.78	-0.22	-0.28	0.04
Δe_t (empirical)	1.00	-0.03	0.06	-0.41	0.06	0.03
r_t	0.65	1.00	0.75	0.00	-0.68	0.01
$r_t(empirical)$	-0.03	1.00	0.38	0.09	0.31	0.37
π_t	0.78	0.75	1.00	0.00	-0.39	0.08
π_t (empirical)	0.06	0.38	1.00	-0.06	0.40	0.18
π_t^*	-0.22	0.00	0.00	1.00	0.00	0.00
π_t^* (empirical)	-0.41	0.09	-0.06	1.00	0.14	0.34
y_t	-0.28	-0.68	-0.39	0.00	1.00	-0.33
y_t (empirical)	0.06	0.31	0.40	0.14	1.00	0.67
y_t^*	0.04	0.01	0.08	0.00	-0.33	1.00
y_t^* (empirical)	0.03	0.37	0.18	0.34	0.67	1.00

See table 1 for the definitions of Δe_t, r_t, π_t, π_t^*, y_t and y_t^*. The empirical correlations of estimation sample is given in the lines where (empirical) is provided.

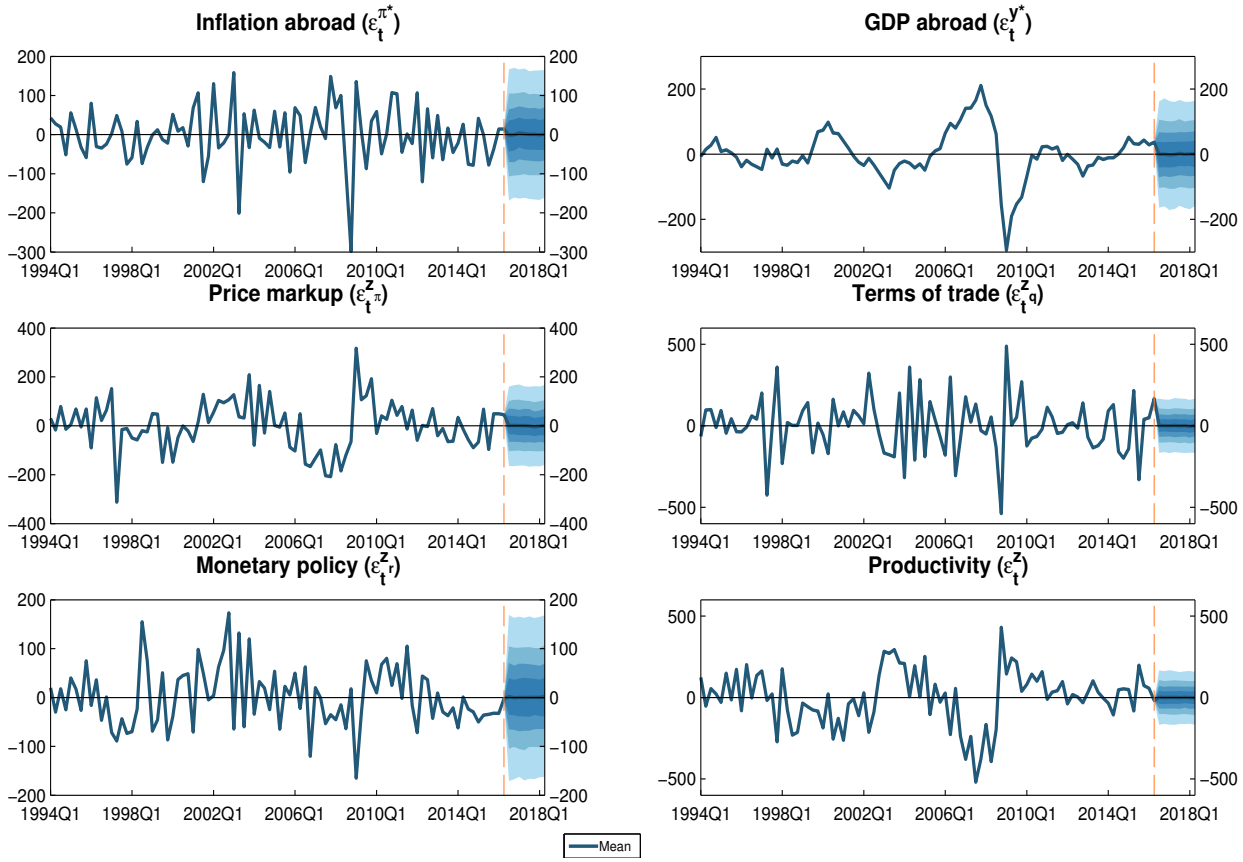
C Graphs

Figure 4: Density forecasts produced by the modified Lubik and Schorfheide (2007) model.



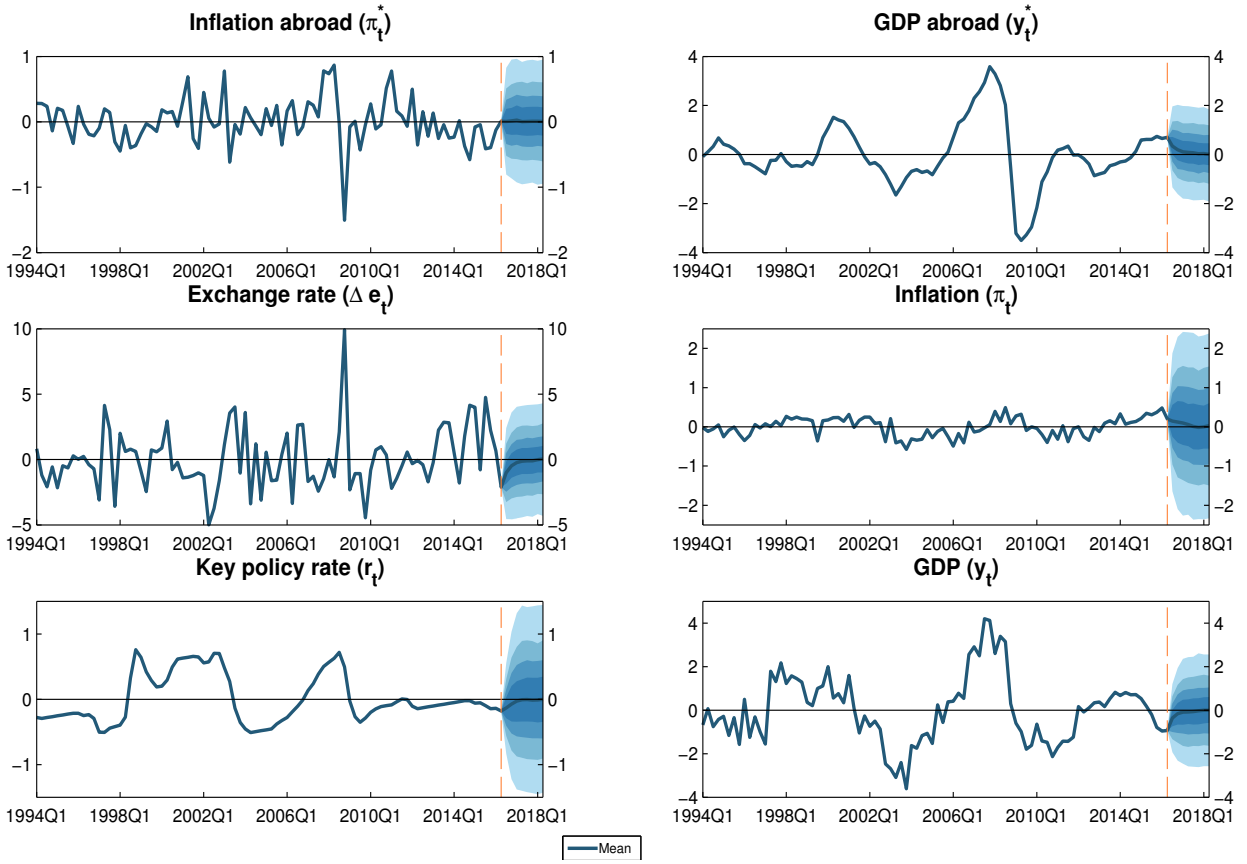
Density forecasts are produced by simulating 4000 draws from the vector of innovations at each forecasting step. See table 1 for the definitions of π_t^*, y_t^*, Δe_t, π_t, r_t and y_t. y_t^* and y_t are both measured as percentage deviation from steady-state, while the rest are measured as percentage points deviation from steady-state.

Figure 5: Density forecasts produced by the modified Lubik and Schorfheide (2007) model. Innovations (multiplied by 100).



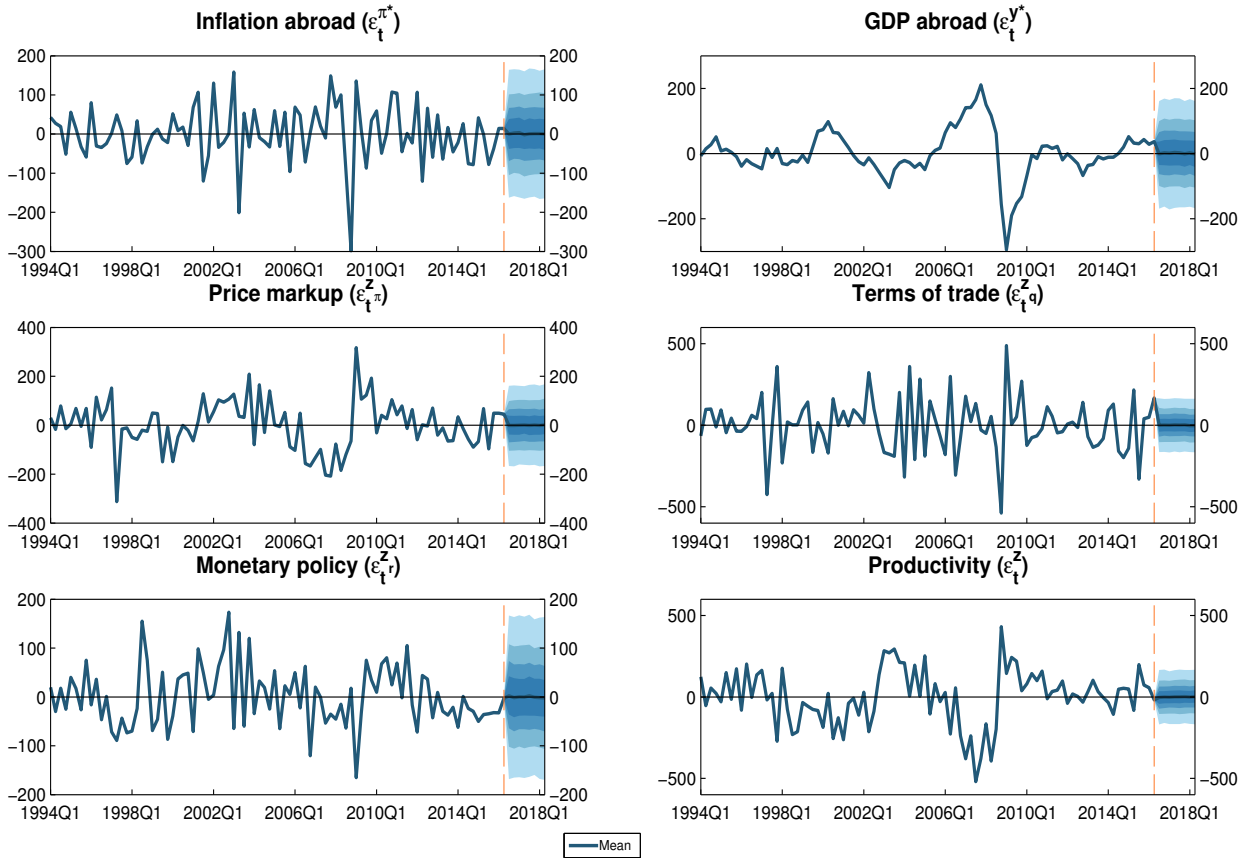
Density forecasts are produced by simulating 4000 draws from the vector of innovations at each forecasting step. See section 4.1 for the definitions of $\epsilon_t^{\pi^*}$, $\epsilon_t^{y^*}$, $\epsilon_t^{z^\pi}$, $\epsilon_t^{z^q}$, $\epsilon_t^{z^r}$ and ϵ_t^{z}.

Figure 6: Density forecasts produced by conditioning on the forecasts from the modified Lubik and Schorfheide (2007) model.



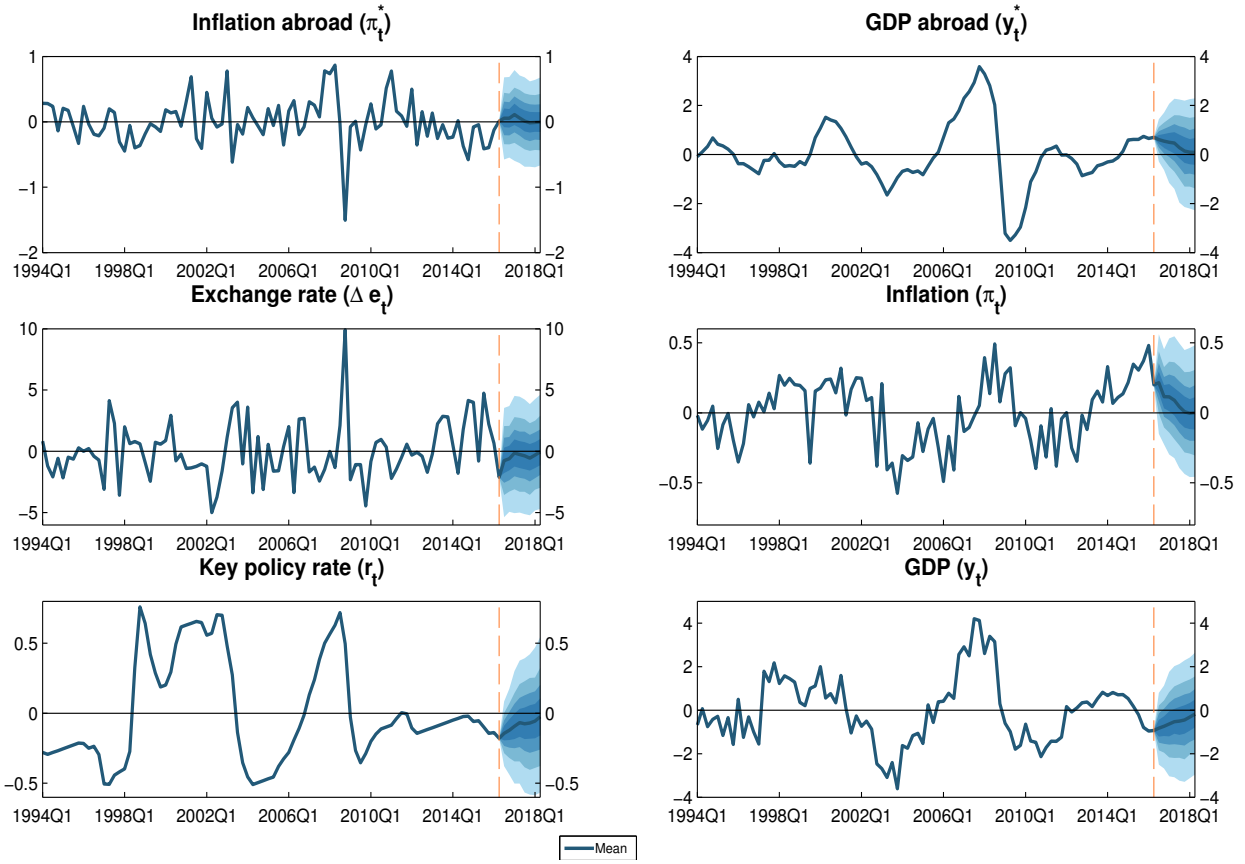
Density forecasts is produced by simulating 4000 draws from the multivariate distribution of the density forecasts of the endogenous variables over the full forecasting horizon produced by the modified Lubik and Schorfheide (2007) model using the algorithm given in section 3.2.1. The multivariate distribution is constructed by a Gaussian copula that is parameterized using the correlation structure of the modified Lubik and Schorfheide (2007) model and the marginal distributions of the density forecasts shown in figure 4. The theoretically calculated marginal distributions from the model are used. See table 1 for the definitions of π_t^*, y_t^*, Δe_t, π_t, r_t and y_t. y_t^* and y_t are both measured as percentage deviation from steady-state, while the rest are measured as percentage points deviation from steady-state.

Figure 7: Identified distributions of the innovation when conditioning on the forecasts from the modified Lubik and Schorfheide (2007) model (multiplied by 100).



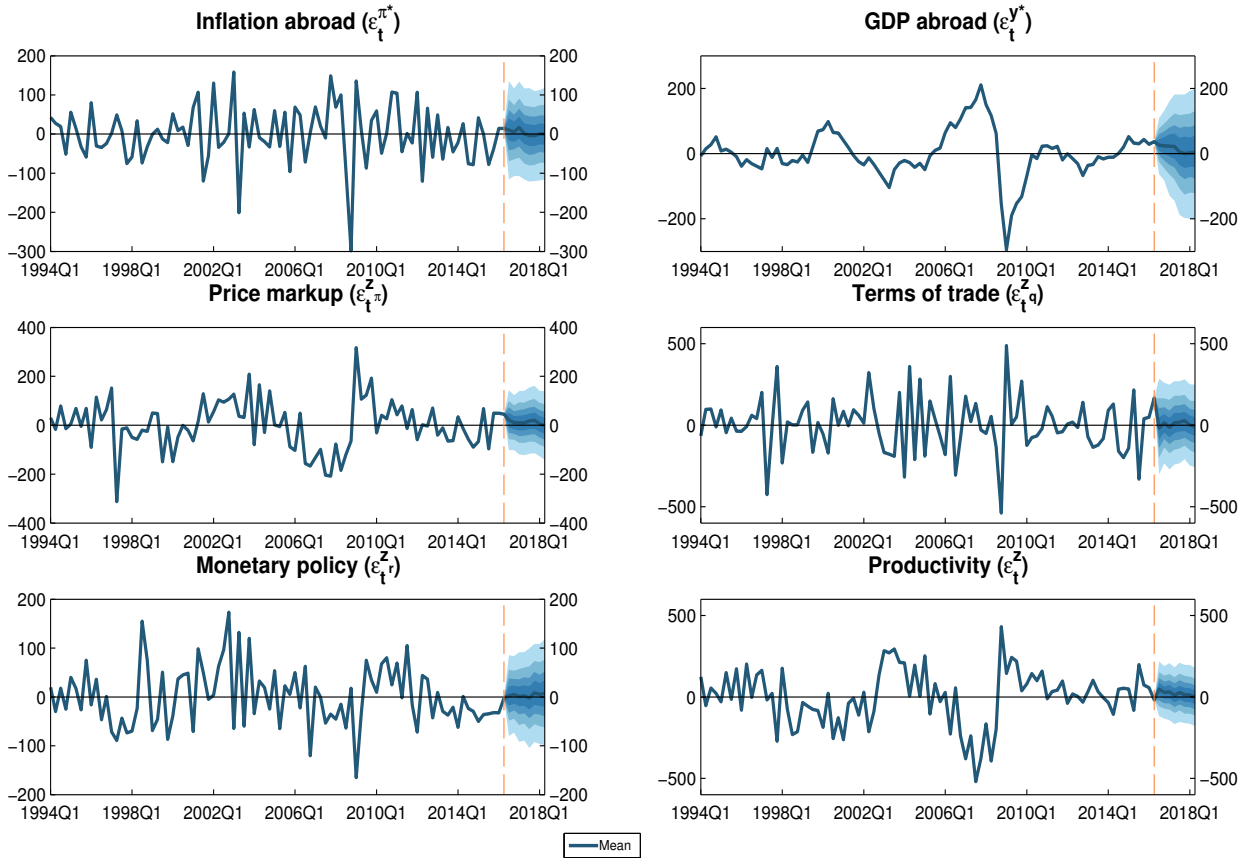
Density forecasts is produced by simulating 4000 draws from the multivariate distribution of the density forecasts of the endogenous variables over the full forecasting horizon produced by the modified Lubik and Schorfheide (2007) model using the algorithm given in section 3.2.1. The multivariate distribution is constructed by a Gaussian copula that is parameterized using the correlation structure of the modified Lubik and Schorfheide (2007) model and the marginal distributions of the density forecasts shown in figure 4. The theoretically calculated marginal distributions from the model are used. See section 4.1 for the definitions of $\epsilon_t^{\pi^*}$, $\epsilon_t^{y^*}$, $\epsilon_t^{z^\pi}$, $\epsilon_t^{z^q}$, $\epsilon_t^{z^r}$ and ϵ_t^{z}.

Figure 8: Density forecasts produced by conditioning on the forecasts from the pool of VAR models.



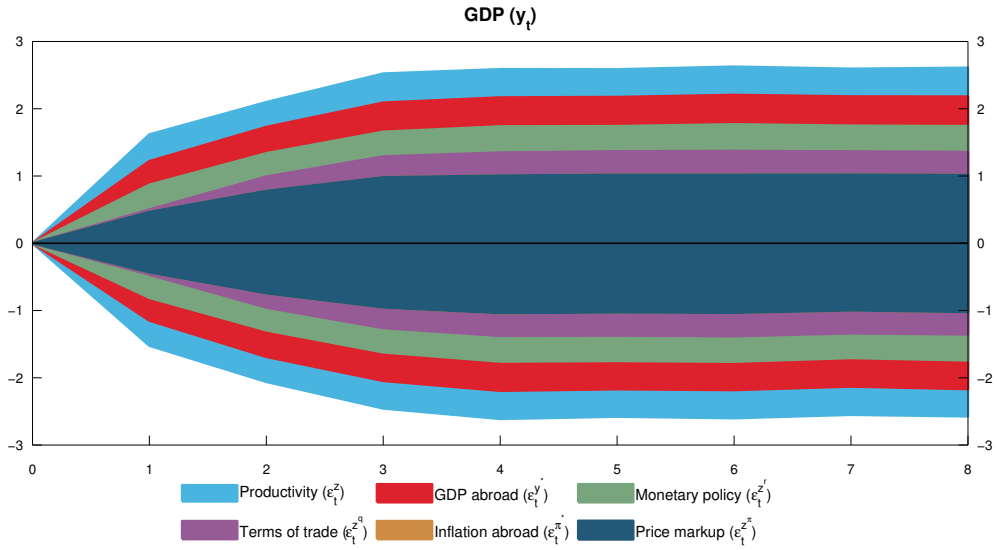
Density forecasts are produced by simulating 4000 draws from the multivariate distribution of the density forecasts of the endogenous variables over the full forecasting horizon produced by the pool of VAR models using the algorithm given in section 3.2.1. The multivariate distribution is constructed by a Gaussian copula that is parameterized using the correlation structure of the modified Lubik and Schorfheide (2007) model and the marginal distributions of the density forecasts of the pool of VAR models. A Gaussian kernel density estimator is used to estimate these marginal distributions. See table 1 for the definitions of π_t^*, y_t^*, Δe_t, π_t, r_t and y_t. y_t^* and y_t are both measured as percentage deviation from steady-state, while the rest are measured as percentage points deviation from steady-state.

Figure 9: Identified distributions of the innovation when conditioning on the forecasts from the pool of VAR models (multiplied by 100).



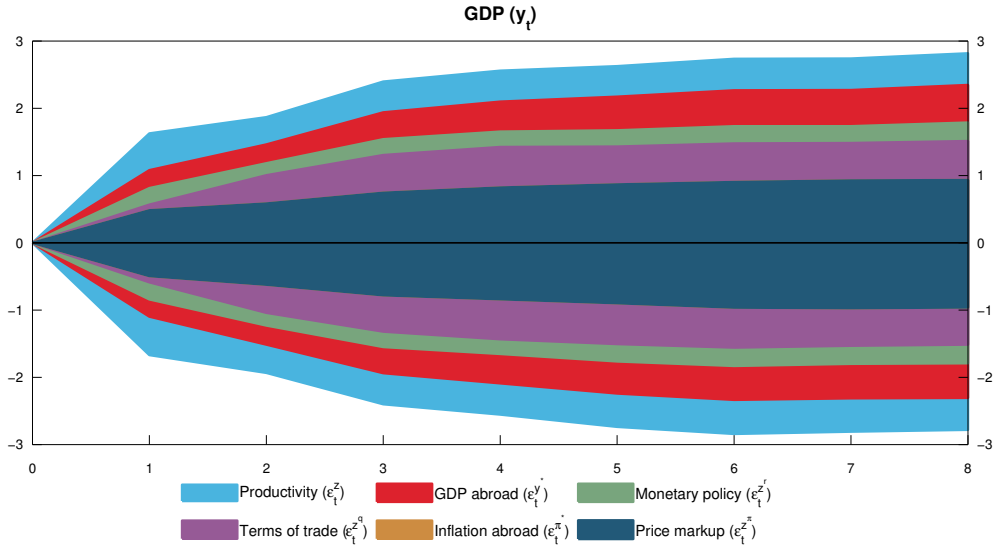
Density forecasts are produced by simulating 4000 draws from the multivariate distribution of the density forecasts of the endogenous variables over the full forecasting horizon produced by the pool of VAR models using the algorithm given in section 3.2.1. The multivariate distribution is constructed by a Gaussian copula that is parameterized using the correlation structure of the modified Lubik and Schorfheide (2007) model and the marginal distributions of the density forecasts of the pool of VAR models. A Gaussian kernel density estimator is used to estimate these marginal distributions. See section 4.1 for the definitions of $\epsilon_t^{\pi^*}$, $\epsilon_t^{y^*}$, $\epsilon_t^{z^\pi}$, $\epsilon_t^{z^q}$, $\epsilon_t^{z^r}$ and ϵ_t^{z}.

Figure 10: Decomposition of the density forecasts as percentage deviations from the mean forecast. Modified Lubik and Schorfheide (2007) model.



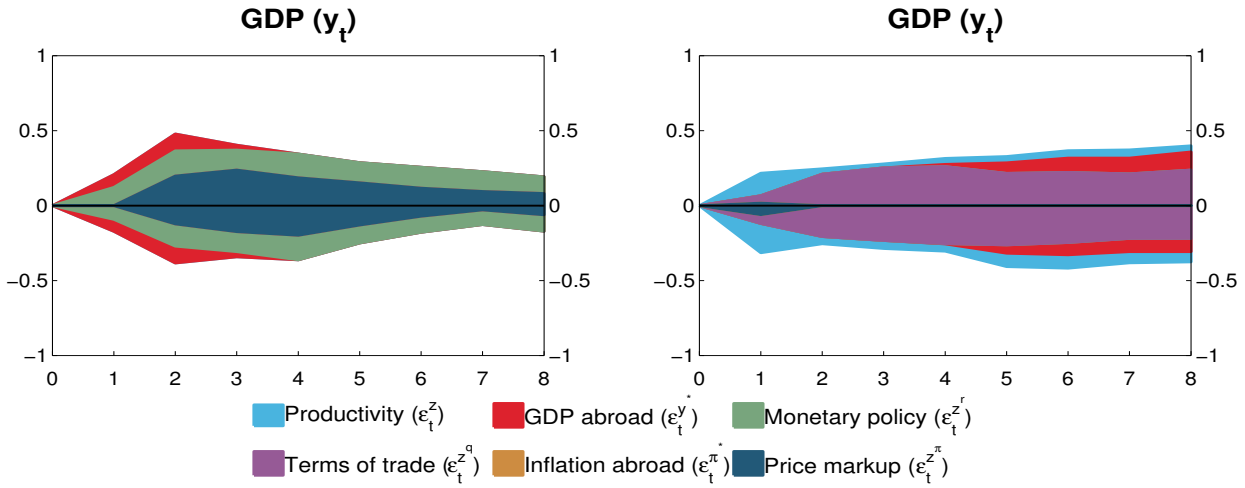
The 90th percentile is used. See table 1 for the definition of y_t.

Figure 11: Decomposition of the density forecasts as percentage deviations from the mean forecast. Pool of VAR models.



The 90th percentile is used. See table 1 for the definition of y_t.

Figure 12: Percentage points difference in the decomposition of the density forecasts from the modified Lubik and Schorfheide (2007) model and the pool of VAR models of the GDP gap (y_t).



Difference between the decomposition in figure 10 and figure 11. The innovations given in the left hand side figure indicates that those innovations leads to a wider density forecast of the GDP gap, at a given horizon, in the modified Lubik and Schorfheide (2007) model versus the pool of VAR models. The opposite is the case in the right hand side figure. See table 1 for the definition of y_t.