Conference on

Forecasting and Monetary Policy

Berlin, 23-24 March 2009

Junior Maih
Central Bank of Norway

„Conditional Forecasts in DSGE Models“

Conditional Forecasts in DSGE Models*

Junior Maih[†]
Norges Bank

February 20, 2009

Abstract

New-generation DSGE models are sometimes misspecified in dimensions that matter for their forecasting performance. The paper suggests one way to improve the forecasts of a DSGE model using a conditioning information that need not be accurate. The technique presented allows the forecaster to apply a continuum of degrees of uncertainty around the mean of the conditioning information, making hard-conditional and unconditional forecasts special cases. Using a small open-economy DSGE model, the paper contrasts conditional forecasts based on anticipated events to those based on unanticipated ones.

Keywords: DSGE model, conditional forecast
JEL classification: C53, F47

*This project was started in April 2008 when I was visiting the Federal Reserve Bank of Atlanta. The hospitality of the Atlanta Fed is gratefully acknowledged. I would like to thank seminar participants at the Atlanta Fed, Norges Bank, the Norwegian school of Management and Statistics Norway, for useful comments. I also thank Will Roberds, John Robertson, Dan Waggoner and Tao Zha for the many helpful suggestions. The views expressed herein at my own and do not necessarily reflect the opinions of Norges Bank. All errors are my own fault.

†Monetary Policy Department, Norges Bank, Bankplassen 2, P.O. Box 1179 Sentrum, NO-0107 Oslo, Norway. E-mail: junior.maih@norges-bank.no, tel: +47 22 31 62 20

Contents

1 Introduction

In addition to being useful for policy analysis, new-generation DSGE models have also been shown to compare well with models such as VARs and BVARs in terms of forecast accuracy (Smets and Wouters (2004), Adolfson et al. (2005)). Nevertheless, DSGE models are sometimes misspecified in some dimensions that affect their forecasting performance (see e.g. Del-Negro et al. (2005)). In the event of huge and unexpected shocks, models that lack the flexibility to adapt are very likely to deliver poor (short-term) forecasts. This paper shows how the theory of conditional forecasts can be extended to DSGE models. It argues that to the extent that leading information is available, relevant and reliable, conditioning on it may reduce the uncertainty in the endogenous variables and thereby improve the forecasting performance of a DSGE model without necessarily having to change its structure[1].

The need to incorporate conditioning information into a forecast comes naturally in circumstances in which observations on some variables are released before others, or in cases where it is believed that some other model may be superior to the DSGE model of interest when it comes to forecasting the variable to be used as conditioning information. In any case, having a systematic way to incorporate that information in the forecasts from a model more easily allow the tracking of systematic forecast errors than in the case of judgemental forecasts where there is no formal model of how the data are used (see e.g. Robertson et al. (2005)).

Conditional forecasts methods have typically been developed and applied for models with fewer theoretical underpinnings than DSGE models. To mention a few, Doan et al. (1984) exploit the covariance matrix structure in a VAR to account for the impact of conditioning a forecast on post-sample values for some variables in their model. Waggoner and Zha (1999) extend Doan, Litterman and Sims and use Bayesian methods to compute the exact finite-sample distribution of conditional forecasts in both structural and reduced-form VARs, accounting for the uncertainty in the parameters. Robertson et al. (2005) develop a relative entropy procedure for imposing restrictions on simulated forecasts distributions.

DSGE models offer a better structural interpretation than VARs and from a policy standpoint, we need more than mere forecasts, we need them

[1] Changing the structure of the model may imply that one understand the microfoundations of some observed phenomenon, which is not always obvious. The current financial crisis and the problems associated with volatile oil prices are cases in point.

to be economically interpretable. Only few papers have attempted to compute constrained forecasts in a DSGE model. Christoffel et al. (2007) construct conditional forecasts for the New Area-Wide Model of the Euro Area. Benes et al. (2008) interpret the "off-model information" they condition on as judgement and compute forecasts based on KITT (the RBNZ's DSGE model). Both papers as well as the aforementioned typically assume no uncertainty around the information they condition on, which is known as hard conditioning, even if the conditioning information may represented by forecasts coming from other models[2].

This paper is close in spirit with Andersson et al. (2008), who extend Waggoner and Zha (1999) and develop a procedure for density-conditional forecasts for a SVAR, which they estimate on Swedish data. Interestingly, they show in an example that the distribution of the unrestricted variables may be too narrow if the model is conditioned only upon central tendencies. We take that idea one step further and argue that this holds true even if there is no uncertainty about the conditioning information. The technique we present allows the forecaster to apply a continuum of degrees of uncertainty around the mean of the conditioning information, making hard-conditional and unconditional forecasts special cases. Because it does not take it for granted that conditioning will necessarily improve forecasts, given that the models are inherently misspecified and that the conditioning information itself need not be accurate (forecasts from other models, data revisions, etc.), the paper aims at shedding light into the conditions for which hard conditions are superior to soft conditions or to no conditions and vice-versa[3]. This has the advantage of pointing out the variables for which the cross-equations restrictions of the model may be too tight.

A further contribution of the paper is the discussion of difference concepts that are not present in VARs. Unlike in VARs, the type of conditioning method employed in a DSGE depends on whether the conditioning information is anticipated or not. In a rational expectations setting, anticipated future shocks affect the current decisions of economic agents. In this case, the

[2] The exception to this is Waggoner and Zha (1999), who also discuss soft conditioning. However, they use an inefficient rejection sampling procedure to do soft conditioning.

[3] For tightly parameterized models, model misspecification is certainly an issue if the hypothesized relationships are not supported by the data. Such misspecifications may be pushed into the shock processes, resulting in an uncertainty that may be too large to be useful for policy analysis. In that case, a further advantage of conditioning then, is to reduce that uncertainty.

conditioning variable is exogenous and the most econometrically meaningful exercise is a perfect foresight simulation. The emphasis of the paper, however, will be on the case where the conditioning variable is endogenous. Unanticipated shocks hitting the system over the forecast horizon are such that the resulting values for the conditioning endogenous variables match the restrictions imposed by the conditioning scheme. As a consequence the unrestricted variables have to be determined simultaneously using the reduced-form solution of the DSGE model[4]. We also show how anticipated and unanticipated events can be combined in the same framework.

The rest of the paper proceeds as follows: section 2 illustrates conditioning in a bivariate normal distribution. While this simple example serves the purpose of building some intuition, it also helps us draw conclusions that will reappear when we turn to DSGE models. Section 3 then presents the general framework for forecasting with DSGE models. Using that framework, section 4 proceeds to deriving the formulas for reduced-form conditional forecasts. Section 5 considers an application of the techniques derived in section 4 to the Lubik and Schorfheide (2007) model estimated on Canadian data. The application shows the benefits of conditioning when the dynamics of the data is adequately nailed by the model. In section 6, reduced-form forecasts are contrasted with those the perfect foresight ones. Section 7 concludes.

2 Conditioning in a bivariate normal distribution

In order to build some intuition for the type of analysis we will be doing in the next sections, consider the following conditional distribution for some variable y:

$$f(y|x) = N\left[\mu_y - \frac{\rho\sigma_y}{\sigma_x}(\mu_x - x), \sigma_y^2(1-\rho^2)\right]$$

[4]In a rational expectations setting, agents exploit any available information that improves forecasts. One possible interpretation of reduced-form forecasts in this context is that, among all the possible future paths generated by unconditional forecasts we restrict ourselves to the ones that fall into a target region. Another way of looking at such forecasts is in terms of scenario analysis for communication: how do we reinterpret unconditional forecasts when we have a prior (strong or agnostic) on the future development of some variable.

where μ_y is the (marginal) mean of y, σ_y its (marginal) standard deviation and ρ the coefficient of correlation between x and y. Likewise, μ_x denotes the mean of x and σ_x its variance.

Assume $x = 0$, $\mu_y = 0$, $\sigma_y = 1$, and $\rho = .5$. Then for various values of μ_x and σ_x, we can compute the marginal density $f(y|x)$. We pick those values from a truncated normal distribution for x

$$f(x|x \in [\underline{x}, \overline{x}]) = \frac{\phi\left(\frac{x-\mu_{x0}}{\sigma_{x0}}\right)}{\sigma_{x0}\left[\Phi\left(\frac{\overline{x}-\mu_x}{\sigma_{x0}}\right) - \Phi\left(\frac{\underline{x}-\mu_{x0}}{\sigma_{x0}}\right)\right]}$$

where μ_{x0} and σ_{x0} are such that $\lim\limits_{[\underline{x},\overline{x}]\longrightarrow(-\infty,+\infty)} f(x|x \in [\underline{x}, \overline{x}]) = N(\mu_{x0}, \sigma_{x0})$, $\phi(\cdot)$ is the standard normal probability density function and $\Phi(\cdot)$ the associated cumulative density function. It is easy to derive $\mu_{x|x\in[\underline{x},\overline{x}]}$ and $\sigma_{x|x\in[\underline{x},\overline{x}]}$. The boundaries of x will be given by $\underline{x} = .5 - \delta\sigma_x$ and $\overline{x} = .5 + \delta\sigma_x$, where we will let $\delta = 3, 2.5, 2, 1.5, 1, .5, .1$.

Figure 1 shows how the density of y conditional on x changes as we vary δ. For large values of δ such as 3, the conditional distribution does not change much, but as we decrease δ, we become more and more informative about the location of x and the conditional distribution of y shifts to the left. This implies that some of the areas of the support of y that were unlikely under large values of δ become more and more likely.

We can think of x as some conditioning information that helps improve our inference about the density of y. If the model is correct, and the conditioning information is good, then we can expect the mean of y to be around -4 when δ, which can be interpreted here as the degree of tightening, is .1. If the model is incorrect, or if the conditioning information is bad, there is no guarantee that we can make good predictions about y. Even if the model is not good, it may still capture the correlation between x and y in such a way that a good information on x implies a good prediction on y.

Although this is a simple example, the conclusions derived here extend to more complicated settings as we will see later. But before generalizing those ideas to the case where x and y are matrices of containing observations of several variables over time, we first turn to the framework for which the formulas will be derived.

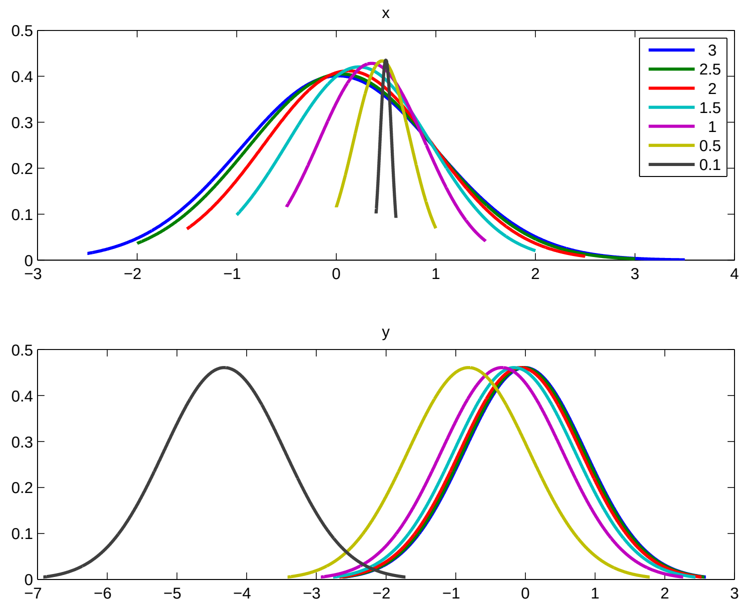

Figure 1: Bivariate normal example

3 General framework for forecasting with DSGE models

Let the DSGE model in linearized form be given by

$$E\left[\Theta_{-1}\left(\theta\right)y_{t-1}+\Theta_{0}\left(\theta\right)y_{t}+\Theta_{1}\left(\theta\right)y_{t+1}+\Psi\left(\theta\right)\varepsilon_{t}\right]=0 \tag{1}$$

where y_t is a $m\times 1$ vector of endogenous variables (including both states and controls), ε_t is a $x\times 1$ vector of exogenous shocks, Θ_{-1}, Θ_0, Θ_1 are $m\times m$ matrices, Ψ is an $m\times x$ matrix. Those matrices are a function of θ, the vector of deep structural parameters.

The solution of the of this system takes the form

$$y_{t}=A\left(\theta\right)y_{t-1}+B\left(\theta\right)\varepsilon_{t},\,\varepsilon_{t}\sim N\left(0,I\right) \tag{2}$$

where A is a $m\times m$ matrix, B is $m\times x$. This is the state space representation.

The model, if estimated, is estimated using data up to period T. We are interested in conditional forecasts h periods ahead. The n-step forecast at time T can be written as

$$y_{T+n} = A^n y_T + \sum_{i=1}^n A^{n-i} B \varepsilon_{T+i}$$

We can decompose the forecast errors of the variables up to period $T+n$, and re-write them in a system as:

$$\begin{bmatrix} B & 0 & \cdots & \cdots & 0 \\ AB & B & \ddots & \ddots & 0 \\ A^2B & AB & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ A^{n-1}B & A^{n-2}B & \cdots & AB & B \end{bmatrix} \begin{bmatrix} \varepsilon_{T+1} \\ \varepsilon_{T+2} \\ \vdots \\ \varepsilon_{T+n-1} \\ \varepsilon_{T+n} \end{bmatrix} = \begin{bmatrix} y_{T+1} - Ay_T \\ y_{T+2} - A^2 y_T \\ y_{T+3} - A^3 y_T \\ \vdots \\ y_{T+n} - A^n y_T \end{bmatrix}$$

or more compactly

$$\mathcal{R}(\theta)\,\varepsilon = r(\theta) \subseteq \Pi(\theta)$$

where M is $mn \times nx$ and ε is $nx \times 1$.

4 Conditional Forecasts and Probability Distributions

Consider a condition that restraints the value of the j-th endogenous variable in the model at time $T+n$, denoted by $y_{T+n}(j)$ in the range of $\left[\underline{y}_{T+n}(j), \overline{y}_{T+n}(j)\right]$. Then this condition implies that

$$\mathcal{R}_{j,\circ}(\theta)\,\varepsilon \in \left[\underline{y}_{T+n}(j) - (A^n y_T)_j, \overline{y}_{T+n}(j) - (A^n y_T)_j\right]$$

where $\mathcal{R}_{j,\circ}$ is the j^{th} line of matrix $\mathcal{R}$, ε is a $k \times 1$ vector of stacked ε_ts.

The condition above is easily generalized to more conditioning variables and periods. A compact form of such a generalization is

$$R(\theta)\,\varepsilon \in C(\theta) \subseteq \mathbb{R}^q,\ q \leq k = h \times x$$

where h is the maximum number of forecast horizons, so that k is the total number of future shocks, R is a $q \times k$ matrix stacking the relevant rows of $\mathcal{R}_{j,\circ}$, and $C(\theta)$ is the restricted set of outcomes.

More explicitly, we consider conditions such that

$$\underline{r}(\theta) \leq R(\theta)\varepsilon \leq \overline{r}(\theta)$$

and assume that $R(\theta)$ is of rank $q \leq k = h \times x$.

4.1 Decomposition of shocks

Because in general $q < k$, the covariance matrix ε conditional on the restriction will be singular. It is possible, however, to partition the space of ε into disturbances that are crucial for meeting the restrictions and those that are not[5]. Consider the decomposition

$$\varepsilon = M_1\gamma_1 + M_2\gamma_2, \text{ with } \gamma_1 \sim N(0, I_{k-q})$$

M_1 is a $k \times (k-q)$ matrix chosen to be an orthonormal basis for the null space of R, that is

$$M_1 = \left\{X \in \mathbb{R}^k | RX = 0 \wedge X'X = I\right\}$$

M_2 which is $k \times q$ could be chosen either as an orthonormal basis of the null space of M_1' or the orthonormal basis for the column space of R'. In both cases, RM_2 will be invertible so that the restriction above simplifies to

$$(RM_2)^{-1}\underline{r} \leq \gamma_2 \leq (RM_2)^{-1}\overline{r}$$

Hence draws from γ_1 can be made from an unrestricted normal distribution while draws for γ_2 will have to lie within the specified boundaries.

4.2 Conditioning Methods

Here we are interested in characterizing the distribution of and ε, given the restrictions, and thereby that of the conditional forecasts y_{T+n}.

$$E(\varepsilon|C,\theta) = M_2 E\gamma_2$$

$$V(\varepsilon|C,\theta) = M_1M_1' + M_2E\gamma_2\gamma_2'M_2'$$

and it follows that

$$p(y_{T+n}|Y_{T+n-1},\theta) = N(Ay_{T+n-1} + BE\varepsilon_{T+n}; BV(\varepsilon_{T+n})B')$$

[5]Thanks to Dan Waggoner at the Atlanta Fed for suggesting this approach.

More explicit expressions for both $E(\varepsilon|C,\theta)$ and $V(\varepsilon|C,\theta)$ depend on the type of conditioning.

4.2.1 Hard conditioning

Here we consider the case where $\underline{y}_{T+n}(j) = \overline{y}_{T+n}(j)$ so that $C(\theta)$ collapses to a $q \times 1$ vector of values $r^*(\theta) = \underline{r}(\theta) = \overline{r}(\theta)$. The set of conditions in the constraint can then be re-written as:

$$R\varepsilon = r^*$$

In this case,

$$\gamma_2 = (RM_2)^{-1} r^*$$

and the distribution of $\varepsilon|C$ is as follows[6][7]

$$E(\varepsilon|C,\theta) = M_2 (RM_2)^{-1} r^*, \text{ and } V(\varepsilon|C,\theta) = M_1 M_1'$$

4.2.2 Soft and no conditioning

In this case, γ_2 is drawn from a truncated normal distribution. Define $\alpha_1 \equiv (RM_2)^{-1} \underline{r}(\theta)$ and $\alpha_2 \equiv (RM_2)^{-1} \overline{r}(\theta)$. Those two vectors are $q \times 1$

We have for $i = 1, 2, ..., q$

$$E(\gamma_{2i}|\gamma_{2i} \in [\alpha_{1i}, \alpha_{2i}]) = -\frac{\phi(\alpha_{2i}) - \phi(\alpha_{1i})}{\Phi(\alpha_{2i}) - \Phi(\alpha_{1i})}$$

$$V(\gamma_{2i}|\gamma_{2i} \in [\alpha_{1i}, \alpha_{2i}]) = 1 - \frac{\alpha_{2i}\phi(\alpha_{2i}) - \alpha_{1i}\phi(\alpha_{1i})}{\Phi(\alpha_{2i}) - \Phi(\alpha_{1i})} - \left[\frac{\phi(\alpha_{2i}) - \phi(\alpha_{1i})}{\Phi(\alpha_{2i}) - \Phi(\alpha_{1i})}\right]^2$$

[6]In general, for a stochastic process $\epsilon \sim N(0, \Sigma_\epsilon)$ with restriction $R\epsilon = r$, it can be shown that the linear estimator with smallest variance is given by $\hat{\epsilon} \equiv E(\epsilon|r) = \Sigma_\epsilon R' (R\Sigma_\epsilon R')^{-1} r$ and that its variance is $\Sigma_\epsilon - \Sigma_\epsilon R' (R\Sigma_\epsilon R')^{-1} R\Sigma_\epsilon$. In the special case where $\Sigma_\epsilon = I$, we retrieve the formulas in Waggoner and Zha (1999). It can also be shown that this estimator corresponds to the hard condition derived above. The proofs are available from the author upon request.

[7]On the assumption that ε is normally distributed, we can compute a compatibility test along the lines of Guerrero and Peña (2000). In particular, the statistic $K = r' (RR')^{-1} r$ follows a χ^2 with q degrees of freedom. Such a test can be useful to gauge whether the dynamics of the model is at odds with the restrictions or not.

It follows that with $\gamma_2 \equiv \left[\gamma_{21}, \gamma_{22}, ..., \gamma_{2q}\right]'$,

$$E\left(\varepsilon|C,\theta\right) = M_2 E\gamma_2$$

$$V\left(\varepsilon|C,\theta\right) = M_1 M_1' + M_2 V\left(\gamma_2\right) M_2'$$

All the γ_{2i} are independent and they are independent from γ_{1i} and $V\left(\gamma_2\right)$ is a diagonal matrix. This implies that no Gibbs (or other multivariate procedure) is required for sampling.

Letting $\left(\alpha_{1i}, \alpha_{2i}\right) \longrightarrow \left(-\infty, \infty\right)$ in the soft condition case above, we have for $i = 1, 2, ..., q$

$$E\left(\gamma_{2i}\right) = 0 \qquad V\left(\gamma_{2i}\right) = 1$$

It follows that

$$\begin{array}{rclcrcl} E\left(\varepsilon|C,\theta\right) & = & 0 & \quad & V\left(\varepsilon|C,\theta\right) & = & M_1 M_1' + M_2 I_q M_2' \\ & & & & & = & I_k \end{array}$$

which is the unconditional forecast case.

Then if the information is accurate and the model coincides with the process generating the data, hard conditions imply a lower variance than soft conditions, which in turn imply a lower variance than no conditions.

4.3 Sampling

To the extent that the conditioning information is accurate, future observations may contain some relevant information about the location of the parameters to be estimated and those parameters are potentially better estimated when including that information. Formally, then we can write

$$y_{T+h} = f_h\left(y_T\right) \tag{3}$$

where $f_h\left(\cdot\right)$ reflects the fact that we may need to update the estimate of θ before computing the forecasts. Hence a Gibbs sampling technique could be designed along the lines suggested by Waggoner and Zha (1999). In a DSGE context, this would consist in:

- initializing an arbitrary value of θ, typically the peak (mode) of $p\left(\theta|Y_T\right)$ or any value randomly drawn $p\left(\theta|Y_T\right)$,

- solving the model in reduced form and recovering the starting values of the unobservables
- generating forecasts $y^{(i)}_{T+1}, y^{(i)}_{T+2}, ..., y^{(i)}_{T+h}$, from $p\left(y_{T+1}, y_{T+2}, ..., y_{T+h}|\theta^{(i-1)}\right)$
- augmenting the original data set with the forecasts and estimating a new value of θ.

But this process of sampling from one vector of parameter conditional on the data (and the forecasts) and then sampling forecasts conditional on the data and the forecasts might be infeasible due to several difficult and expensive steps involved in the process of estimating DSGE models and using the estimates for computing forecasts[8]. One could potentially circumvent this problem by adopting a less computationally intensive method of estimation as GMM, but this would come at the cost of having to select only a few moments of the variables.

The approach suggested here involves estimating the posterior distribution of the parameters only once, that is using only the information available up to period T and leaving open the possibility of computing forecasts based on partially or totally calibrated models. In this way, forecasts can be based on calibration, prior draws, posterior draws or even draws around the mode. The estimated (and or calibrated) parameters are assumed to remain invariant to additional information[9]. In this case, the sampling technique is the following:

For $i = 1, 2, ...N$

1. draw $\theta^{(i)}$ from a chosen distribution (calibration, prior, posterior, mode). If not admissible redraw $\theta^{(i)}$
2. Solve the model for matrices A, B, M_1, M_2, R and r^* (or $\underline{r}(\theta)$ and $\overline{r}(\theta)$ where it applies), and recover the starting values for the unobservable variables by the Kalman smoother

[8]The first step, which consists of finding the peak of the distribution of the parameters typically implies evaluating the likelihood function or the posterior at various admissible vectors of parameters in the parameter space. At each such evaluation, the steady state of the model has to be found and the model has to be solved. Once the peak is found, the posterior distribution has to be constructed through long simulations that are required to compute an unknown distribution.

[9]This is mostly a simplifying assumption. It might well be the case that over the forecasting horizon, the conditioning variables take on values that are far away from the process having generated the observations up to the initial conditions for the forecasts.

3. draw γ_1 from $N(0, I_{k-q})$, and draw γ_2 if necessary

4. Construct a draw of $\varepsilon^{(i)}$ and generate forecasts $\left\{y_{T+1}^{(i)}, y_{T+2}^{(i)}, ..., y_{T+h}^{(i)}\right\}$.

The generated sequence $\left\{y_{T+1}^{(1)}, y_{T+2}^{(1)}, ..., y_{T+h}^{(1)}, ..., y_{T+1}^{(N)}, y_{T+2}^{(N)}, ..., y_{T+h}^{(N)}\right\}$ constitute the distribution of the conditional forecasts.

5 Application to the Lubik Schorfheide (2007) model

5.1 The Lubik-schorfheide model

The various applications we will consider are based on the Lubik and Schorfheide (2007) (LS07) model. It is a small scale open economy DSGE model in Aggregate output (y_t), CPI inflation(π), nominal interest rate(R_t), terms of trade (q_t), exogenous world output (y_t^*), potential output in the absence of nominal rigidities ($\bar{y}_t$), growth rate of the underlying technological progress (z_t) and exchange rate (e_t).

The main equations given in (4) include a demand equation, a Phillips curve, an equation defining domestic inflation as a function of the exchange rate, terms of trade and foreign inflation, a monetary policy reaction function and an equation for potential output.

$$
\begin{array}{rcl}
y_t & = & E_t y_{t+1} - [\tau + \alpha(2-\alpha)(1-\tau)](R_t - E_t \pi_{t+1}) - \rho_z z_t \\
 & & -\alpha[\tau + \alpha(2-\alpha)(1-\tau)] E_t \Delta q_{t+1} + \frac{\alpha(2-\alpha)(1-\tau)}{\tau} E_t \Delta y_{t+1}^* \\
\pi_t & = & \beta E_t \pi_{t+1} + \alpha\beta E_t \Delta q_{t+1} - \alpha \Delta q_t + \frac{\kappa}{\tau+\alpha(2-\alpha)(1-\tau)}(y_t - \bar{y}_t) \\
\pi_t & = & \Delta e_t + (1-\alpha)\Delta q_t + \pi_t^* \\
R_t & = & \rho_R R_{t-1} + (1-\rho_R)[\psi_1 \pi_t + \psi_2 y_t + \psi_3 \Delta e_t] + \varepsilon_t^R \\
\bar{y}_t & = & -\frac{\alpha(2-\alpha)(1-\tau)}{\tau} y_t^*
\end{array}
\tag{4}
$$

Technological progress, foreign output, terms of trade and foreign inflation are exogenous AR(1) processes

$$
\begin{aligned}
z_t &= \rho_z z_{t-1} + \varepsilon_{z,t} & y_t^* &= \rho_{y^*} y_{t-1}^* + \varepsilon_{y^*,t} \\
\Delta q_t &= \rho_q \Delta q_{t-1} + \varepsilon_{q,t} & \pi_t^* &= \rho_{y^*} \pi_{t-1}^* + \varepsilon_{\pi^*,t}
\end{aligned}
\tag{5}
$$

As for the parameters, τ is the intertemporal substitution elasticity, $0 < \alpha < 1$ 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. ψ_1, ψ_2 and ψ_3 and monetary policy parameters as described by the Taylor rule. See Lubik and Schorfheide (2007) for more details.

5.2 The conditional forecast exercise and the data

In order to gauge the potential usefulness of conditioning on a variable in improving the predictions of other variables, we will in turn use the observed interest rate and then the exchange rate as the conditioning information. Although in practice, accurate information on the conditioning variables may not be available, conditioning on actual realizations helps us analyze the possible dangers of conditioning. In addition, conditioning on inaccurate information would just strengthen the results of the paper, that generalize to larger models and different conditioning information[10]. More to the point, even if accurate information is not available, one can always apply soft conditioning for a given measure of uncertainty about the conditioning information.

The parameters of the LS07 model are estimated recursively. Then conditional forecasts are computed for various degrees of uncertainty. For simplicity uncertainty is measured by the standard deviation of conditioning variable over history[11]. We consider 6 degrees of uncertainty: 0, 10, 30, 50, 70 and 100% , where a value of 0 implies hard conditioning, while a value of 100 implies unconditional forecasts.

The data used for estimation and in the subsequent analysis are Canadian data, available from Schorfheide's website. The vector of observables

[10] One can condition on market information as done by Andersson et al. (2008), or on forecasts coming from other models as in Benes et al. (2008).

[11] Andersson et al. (2008) suggest ways of generating a prior uncertainty that can be used in computing density forecasts, including using past forecast errors and or the properties of the model at hand. But there is no perfect way of generating the uncertainty measure to use for the computation of soft-conditional forecasts.

comprises Annual interest rate (R_{At}^{data}), annual inflation (π_{At}^{data}), quarterly output growth (ΔY_t^{data}), exchange rate changes (Δe_t^{data}) and terms of trade changes (Δq_t^{data}). The data set runs from 1970Q1 to 2002Q4 and the priors are the same ones used by Lubik and Schorfheide. The sample from 1970Q1 to 1993Q4 used for the first estimation. All the results are based on the estimated posterior mode of the parameters.

Measures of forecast accuracy The measures of forecast accuracy we consider are the traditional mean absolute error (MAE) and the root mean square error (RMSE) presented in equation (6).

$$MAE_i\left(h\right) = \frac{1}{N_h}\sum_{d=1}^{N_h}\left|e_{i,d}\left(h\right)\right|$$

$$RMSFE_i\left(h\right) = \sqrt{\frac{1}{N_h}\sum_{d=1}^{N_h} e_{i,d}^2\left(h\right)} \tag{6}$$

We also consider a multivariate measure of point forecast accuracy based on the scaled h-step-ahead Mean Squared Error matrix (see equation (7)) used by Adolfson et al. (2005)

$$\Omega_M\left(h\right) = \frac{1}{N_h}\sum_{d=1}^{N_h}\tilde{e}_{\cdot,d}\left(h\right)\tilde{e}_{\cdot,d}'\left(h\right), \text{ with } \tilde{e}\left(h\right) \equiv M^{-\frac{1}{2}}e\left(h\right) \tag{7}$$

M is a scaling matrix that accounts for the differing scales of the forecasted variables and for the fact that the time series may be more or less intrinsically predictable in absolute terms. In this case the measure of forecast accuracy will be the log determinant statistic $\ln\left(\left|\Omega_M\left(h\right)\right|\right)$, which is invariant to the choice of the scaling matrix. Note that since the conditioning variable is matched exactly, its forecast error is 0. In that case the determinant $\left|\Omega_M\left(h\right)\right| = 0$ even if the forecast errors for the other variables are different from 0. For that reason, we remove the row and column corresponding to the conditioning variable before computing the statistic.

6 Results

The question we try to answer here is if the econometrician had known future values of the condition variables, how well would he have done in forecasting other variables applying different degrees of conditioning. Given that RMSE and MAE lead to qualitatively similar results, only the results from RMSE are reported for reasons of brevity.

In this exercise, we demean the data prior to estimation and focus the estimation on the parameters that control the dynamics of the system. The means are added back to the forecast before computing the forecast errors. We condition on the observed interest rate. The RMSEs presented in figure 2 reveals optimistic results. For the interest rate, hard conditioning is superior to all degrees of conditioning. This is also the case with output growth.

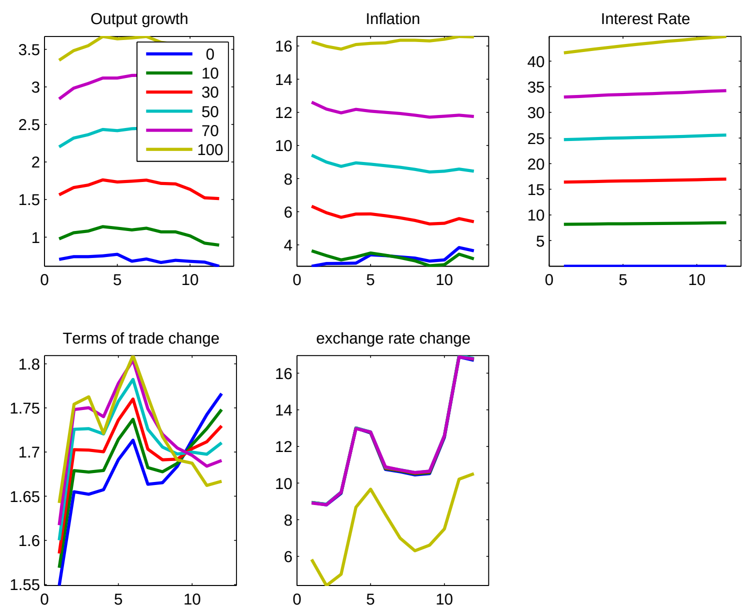

Figure 2: RMSE for model with demeanded data and conditioning on Interest rates

As for inflation, we also have good news: conditioning information on the interest rate improves the forecasts of inflation. However, we see that hard conditions no longer uniformly dominate soft conditions. While hard conditions dominate the 10% soft conditions up to 5 quarters, from then onwards hard conditions are dominated. This suggests that with a misspecified model hard conditions do not necessarily dominate soft conditions even if the conditioning information is accurate. This is echoed in the RMSEs for the exchange rate for which conditioning continues to deteriorate the forecasting performance. All in all, this also suggests that being able to apply different degrees of soft conditioning may help gauge, in a specific model, how tight the cross-equation restrictions are and thereby point the variables whose specification should be improved.

An interesting case is that of terms of trade. Remember from the equations in (5), that terms of trade are modeled as an exogenous process, implying that there is no feedback from the other equations or variables of the system onto terms of trade. But still, conditioning on the interest rate, albeit marginally, seems to improve the forecasts of terms of trade up to about 10 quarters.

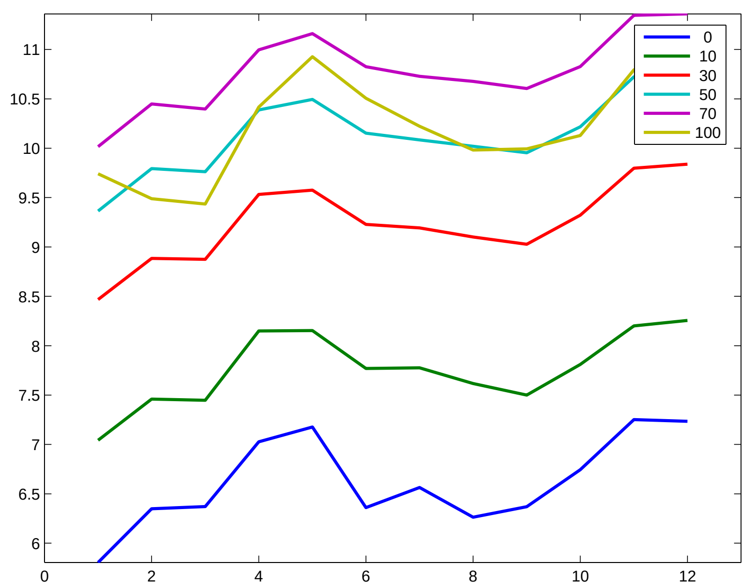

Figure 3: Log determinant statistic for model with demeanded data and conditioning on Interest rates

All in all, as revealed in figure 3, there are benefits to conditioning. The fact that unconditional forecasts (100% uncertainty) dominate the 70% uncertainty soft conditional forecasts stems from the forecast failure on exchange rate changes.

It seems that the variable that the LS07 model has most difficulty explaining or predicting is the exchange rate. The model is certainly misspecified for all the variables but it is more so for the exchange rate than for the others. We proceed to asking the question: what if we had information about the exchange rate instead. Figure 4 presents the RMSEs obtained when conditioning on exchange rate changes. The ranking from hard conditional to unconditional forecasts for the exchange rate is the same for inflation and interest. The RMSEs on output growth remind us that even if the information is accurate, hard conditioning may be too tight.

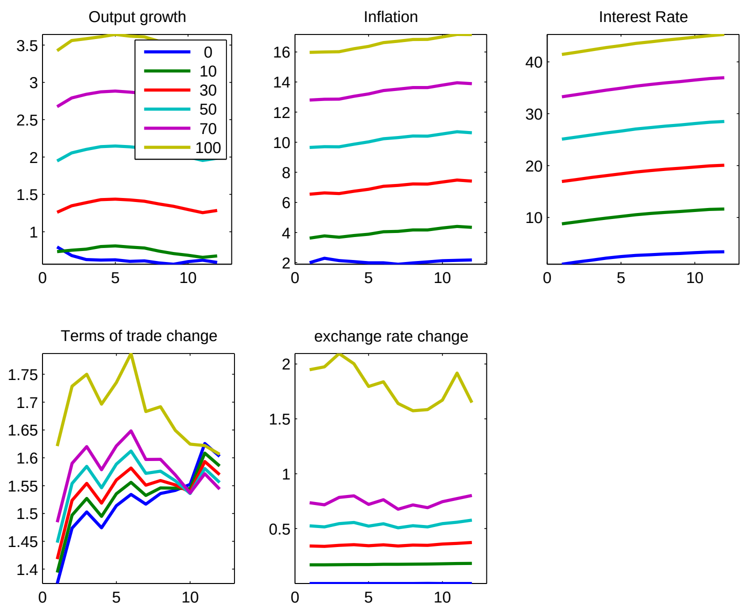

Figure 4: RMSE for model with demeanded data and conditioning on exchange rate changes

Turning to the log determinant statistic in figure 5, the ranking from hard conditional to unconditional forecasts observed in the RMSEs for exchange rate changes is preserved.

Although the model is misspecified for the exchange rate, it does capture some of the correlations between exchange rate and the other variables, such that knowledge of the exchange implies improved forecasts for the other variables in the model. This suggests that even if a model is misspecified along

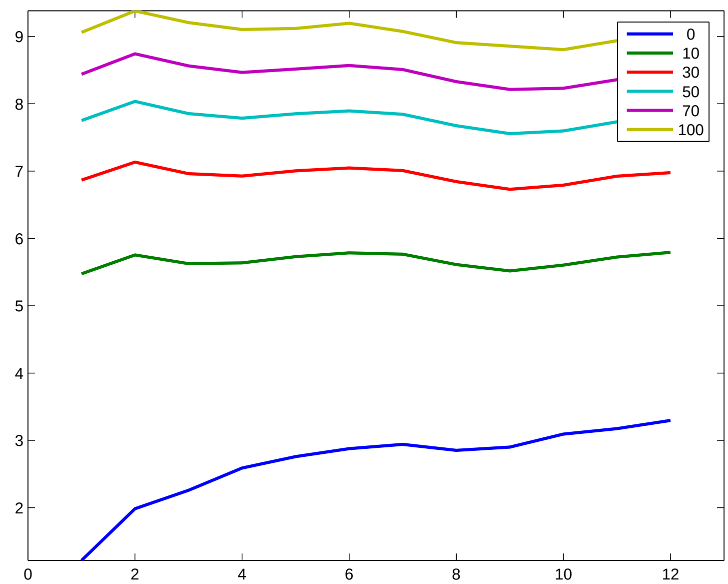

Figure 5: Log determinant statistic for model with demeanded data and conditioning on exchange rate changes

one dimension, it may still be useful in explaining the behavior of other variables.

6.1 Discussion

One reason for the success of the conditional forecast exercise above lies in the fact that the data were demeaned prior to estimation. Attempts to estimate the means in the data alongside the structural parameters resulted in very poor forecasts[12]. For the conditioning variable, the forecasting procedure interprets shifts in the mean as shocks, which is exactly what is needed to replicate the restrictions. For the unrestricted variables however, there is no way to know whether the means have shifted. We argued earlier, (see equation 3) that we may need to update the estimates of the parameters for doing forecast. Unfortunately the Gibbs sampling strategy proposed by Waggoner and Zha (1999) is too expensive to be done for DSGE models. So in trying to understand why conditioning deteriorates the forecasting performance in

[12]The results are available upon request.

this case we guess that failing to update the parameters and in particular the constant over the forecast horizon may be one of the problems. In any case, forecast failure is more likely when the means are estimated than when the uncertainty about the mean is taken away. Intuitively, the estimation of the constant interacts with the estimation of the other parameters in complicated ways and inherits the possible misspecifications of the model may have. In demeaning the data then, we insulate the mean for possible model misspecifications.

7 Deterministic vs reduced-form forecasts

So far we have assumed that the conditioning variable is endogenous, which implied that the agents in the economy observe and react only to current shocks, while expecting that there will be no shock in the future. Now we relax that assumption and allow agents to have perfect foresight over future shocks throughout the forecast horizon. This makes the conditioning variable exogenous. In the appendix, we show how a perfect-foresight solution can be computed for a rational expectations model of the form (1). It is also shown how perfect-foresight forecasts can be combined with reduced form ones.

When doing perfect foresight, the conditioning variable needs to be exogenous and it is this feature of that makes it difficult to condition on just any endogenous variable as we do in the reduced form case. If an endogenous variable is to become exogenous, then one equation has to be discarded from the system, which can be problematic. For instance if we have information on the exchange rate or on the interest rate, choosing one equation to remove from the LS07 system is not obvious. If we have information on terms of trade, however, we can discard the exogenous process for terms of trade. Naturally then, in the perfect foresight exercise we condition on 12 observations of future terms of trade and try to gauge how well the LS07 model would have done at predicting the other variables if all the agents had known future values of terms of trade. The model is estimated recursively as before and at each step the posterior mode of the distribution of the parameters is stored and the reduced-form forecasts are computed. After the computation of the reduced-form forecasts, the stored posterior modes are used to compute the forecasts for the perfect-foresight model.

Figure 6 compares the RMSEs obtained from the reduced-form forecasts to those obtained from perfect foresight. Given that there is no uncertainty

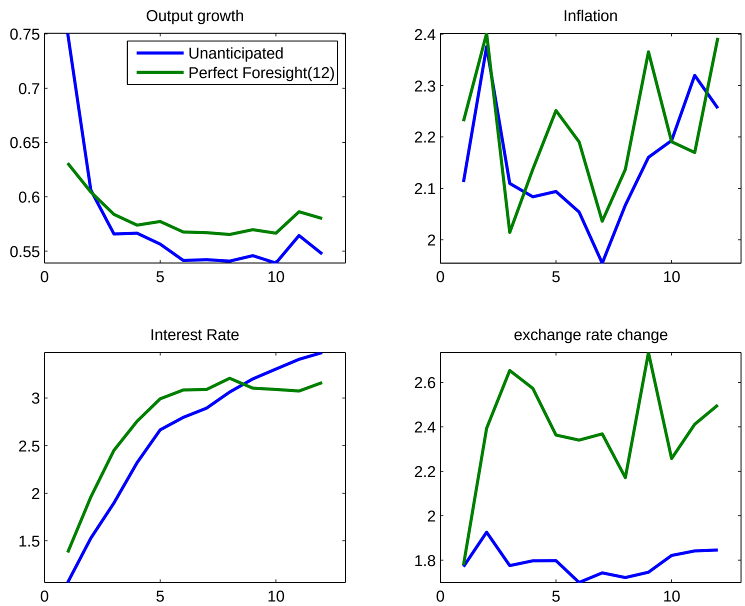

Figure 6: RMSE (conditioning on Terms of trade): Reduced-form versus perfect foresight forecasts

under perfect foresight, the reduced-form forecasts are hard-conditional forecasts. For the most part, reduced-form forecasts tend to outperform the perfect foresight forecasts. In particular, reduced-form forecasts dominate perfect foresight ones for the exchange rate throughout the forecast horizon, they dominate perfect foresight ones for output growth from the second to the end of the forecasting horizon and for the interest rate from the first to the ninth. As for inflation, reduced-form forecasts also dominate perfect foresight ones for most of the periods considered. Not surprisingly then, reduced-form forecasts also dominate perfect foresight forecasts in the multivariate measure of forecast accuracy (figure 7).

Although reduced-form forecasts tend to dominate perfect-foresight forecasts in this example, it is difficult to trace the exact reasons for such an outcome. It might be the case that the observed data are not generated with the agents having information about terms of trade in the future. But still, giving to rational agents the information about future terms of trade makes their reaction today dependent on all the information they have including all the terms of trade to happen in the future. Hence an anticipated change

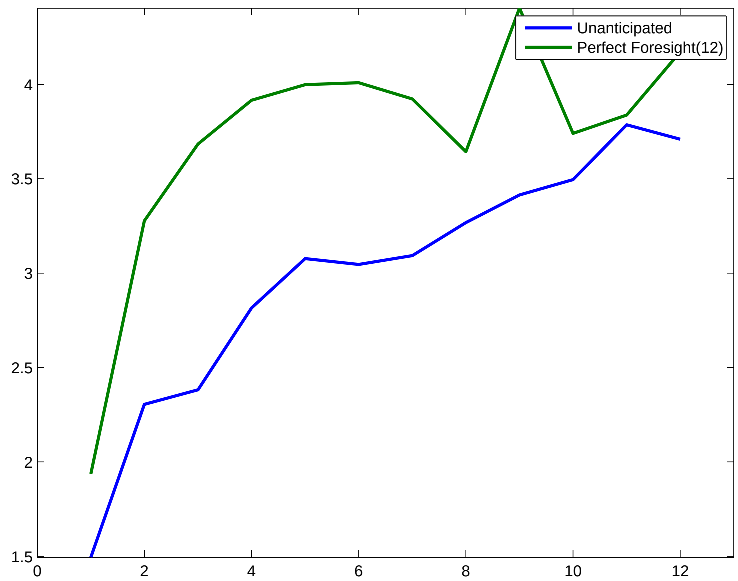

Figure 7: Log determinant statistic (conditioning on Terms of trade): Reduced-form versus perfect foresight forecasts

in the terms of trade tomorrow will trigger a jump either down or up of the other variables and not necessarily a smooth adjustment as would typically be the case in reduced-form (and backward-looking) models. To the extent that jumps are not present in the data, the perfect-foresight assumption is likely to produce poor forecasts relative to the reduced-form forecasts which mostly emphasize the persistence in the data.

Perhaps except for the exchange rate, the forecast performances of the two assumptions are not too far apart and in several instances, perfect-foresight forecasts also dominate reduced-form ones. In other words both reduced-form dominate perfect foresight forecasts are in the same ballpark and both belong in the toolbox of the DSGE model forecaster. One could easily think of examples in which perfect foresight might be more relevant than reduced-form forecasts. For instance when legislative and implementation lags in fiscal policy ensure that private agents receive clear signals about the tax rates they face in the future. In that case, as discussed by Leeper et al. (2008), an econometrician who fails to align his information set with the information set of the agents will get distorted inferences about the effects of tax policies.

8 Conclusion

The paper suggests one way to inform the forecasts of a DSGE model in the presence of conditional information. It argues that conditioning does not necessarily improve the forecasting performance of a DSGE model and in some cases it might even deteriorate forecast accuracy. This happens when the dynamics of the model is at odds with the data or when the correlation between the conditioning information and the other variables in the model is insignificant. On the other hand, in the presence of good conditioning information, even a misspecified DSGE model can still have its forecasting performance improved if it adequately nails the dynamics of the data or the correlation between the conditioning information and variables of interest. In the presence of model misspecification, hard conditioning is not necessarily the best way to go, no matter how accurate the conditioning information is. Tight cross-equation restrictions implied by the model that are forced upon the forecasts in hard conditioning can be relaxed with soft conditioning.

A A solution to the perfect foresight problem

Consider the model

$$E\left[\Theta_{-1}\left(\theta\right)y_{t-1}+\Theta_{0}\left(\theta\right)y_{t}+\Theta_{1}\left(\theta\right)y_{t+1}+\Psi_{x}\left(\theta\right)x_{t}+\Psi_{u}\left(\theta\right)u_{t}\right]=0$$

where x_t is the set of exogenous variables we have some information on and u_t, the set of exogenous shocks that are unanticipated.

With perfect foresight, we use an undetermined coefficients method to guess the solution

$$y_t = G_y y_{t-1} + G_0 x_t + \sum_{j=1}^{n} G_j x_{t+j} + G_u u_t \tag{8}$$

This guess implies that G_y solves matrix polynomial

$$\Theta_1 G_y^2 + \Theta_0 G_y + \Theta_{-1} = 0$$

Now, conditional on G_y, the other parameter matrices are given by

$$G_u = -\left[\Theta_1 G_y + \Theta_0\right]^{-1} \Psi_u$$

$$G_0 = -\left[\Theta_1 G_y + \Theta_0\right]^{-1} \Psi_x$$

$$G_j = \left(-\left[\Theta_1 G_y + \Theta_0\right]^{-1} \Theta_1\right)^j G_0 \qquad j = 1, 2, ..., n$$

The solution for the G_j depends not only on the reduced form solution through G_y, but also directly on the structural form with matrices Θ_1 and Θ_0. When $n = 0$, the solution is that of any rational expectation model as in equation (2). When $n > 0$, the solution of y_t depends not only on shocks happening in period t, but also in the anticipated shocks up to n periods into the future.

In equation (8), the term $G_0 x_t + \sum_{j=1}^{n} G_j x_{t+j}$ is deterministic decides the perfect foresight solution, in the absence of u_t shocks. If in addition, to knowing x_t, we also have information about some variables in vector y_t, then u_t can be specialized for the construction of reduced-form conditional forecasts and so it is possible to combine perfect foresight with reduced-form forecasts.

References

Adolfson, M., J. Lindé, and M. Villani (2005, April). Forecasting performance of an estimated open economy DSGE model for the euro area. Unpublished, Sveriges Riksbank.

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

Benes, J., A. Binning, and K. Lees (2008). Incorporating judgement with DSGE models. Discussion Paper 10, Reserve Bank of New Zealand.

Christoffel, K., G. Coenen, and A. Warne (2007). Conditional versus unconditional forecasting with the new area-wide model of the euro area. European Central Bank.

Del-Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2005, January). On the fit and forecasting performance of new keynesian models. Discussion Paper 4848, Center for Economic Policy Research (CEPR).

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

Guerrero, V. M. and D. Peña (2000). Linear combination of restrictions and forecasts in time series analysis. *Journal of Forecasting 19*, 103–122.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2008). Fiscal foresight: Analytics and econometrics. Working Paper W14028, NBER.

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

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

Smets, F. and R. Wouters (2004, November). Comparing shocks and frictions in US and euro area business cycles: A bayesian DSGE approach. Discussion Paper 4750, Centre for Economic Policy Research (CEPR).

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