2010 | 07

Working Paper

Monetary Policy Department

Conditional forecasts in DSGE models

By Junior Maih

NB NORGES BANK

Working papers fra Norges Bank, fra 1992/1 til 2009/2 kan bestilles over e-post:
servicesenter@norges-bank.no
eller ved henvendelse til: Norges Bank, Abonnementsservice
Postboks 1179 Sentrum
0107 Oslo
Telefon 22 31 63 83, Telefaks 22 41 31 05

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

Working papers inneholder forskningsarbeider og utredninger som vanligvis ikke har fått sin e4ndelige 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, from1992/1 to 2009/2 can be ordered by e-mail:
servicesenter@norges-bank.no
or from Norges Bank, Subscription service,
P.O.Box. 1179 Sentrum
N-0107Oslo, Norway.
Tel. +47 22 31 63 83, Fax. +47 22 41 31 05

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 1502-8143 (online)
ISBN 978-82-7553-553-3 (online)

Conditional Forecasts in DSGE Models*

Junior Maih†
Norges Bank

April 27, 2010

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 for agents to anticipate the information on the conditioning variables several periods ahead. It also 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. An application to a small open-economy DSGE model shows that the benefits of conditioning depend crucially on the ability of the model to capture the correlation between the conditioning information and the variables of interest.

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

*First version: February 2009. This project was started when I was visiting the Federal Reserve Bank of Atlanta in April 2008. The hospitality of the Atlanta Fed is gratefully acknowledged. I would like to thank seminar participants and discussants at the 15th conference on Computing in Economics and Finance in Sydney, the 2009 EEA-ESEM in Barcelona, the Atlanta Fed, Bundesbank, the ECB, Norges Bank, the Bank of Israel, the Bank of Korea, the Norwegian school of Management, the University of Oslo, and, in particular, Alon Binyamini, Christian Dreger, Domenico Giannone, Luca Guerieri, Will Roberds, John Robertson, Dan Waggoner and Tao Zha, for the many useful comments and 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

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[2]. But while improving forecasts could be an objective per se, conditional forecasts are also useful for policy simulations (e.g. effects of an increase in the interest rate or on government spending, etc.), for the assessment of risk (currency depreciation or oil price shocks), for cross-checking, cross validation of different competing models. To this end, having a systematic way to incorporate conditioning information into the forecasts from a model more easily allows 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)).

Methods of conditional forecasts 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

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

[2] This other model could also be a simple judgement.

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 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[3].

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 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[4]. This has the advantage of pointing out the variables for which the cross-equations restrictions of the

[3]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.

[4]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.

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. As rational agents exploit any available information that can improve their forecasts, anticipated events matter for their current decisions. The paper suggests a way of extending the unanticipated shocks framework to the more general case of anticipated ones.

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 conditional forecasts for both anticipated and unanticipated events. 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 evaluates the benefits of conditioning when the dynamics of the data is not adequately nailed by the model. In particular, we contrast conditional forecasts for various degrees of soft conditioning and for various numbers of anticipated steps. Section 6 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:

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

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

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

where μ_{x0} and σ_{x0} are such that $\lim_{[\underline{x},\overline{x}]\longrightarrow(-\infty,+\infty)} f\left(x|x\in[\underline{x},\overline{x}]\right)=N\left(\mu_{x0},\sigma_{x0}\right)$, $\phi(\cdot)$ 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.

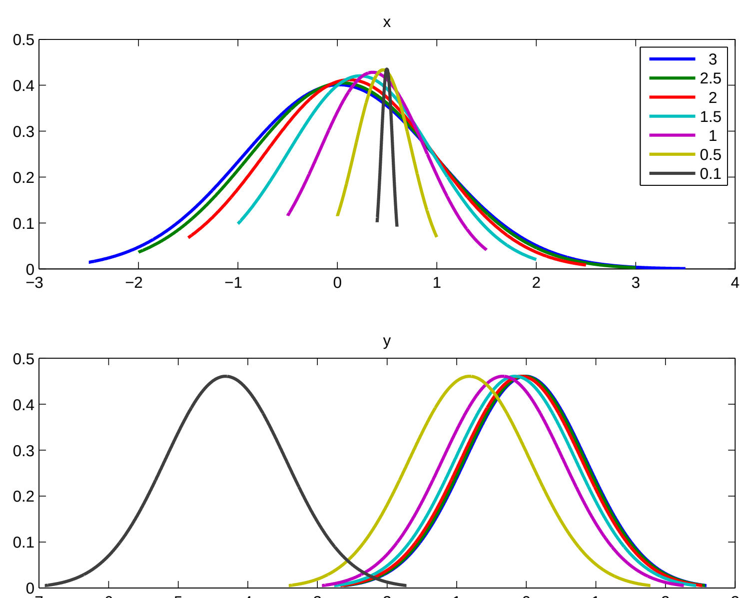

Figure 1: Bivariate normal example

We can think of x as some conditioning information that helps improve our inference about the density of y. If the model is good, 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.

3 General framework for forecasting with DSGE models

Let the DSGE model in linearized form be given by

$$E_t\left[\Theta_{-1}(\theta)\, y_{t-1} + \Theta_0(\theta)\, y_t + \Theta_1(\theta)\, y_{t+1} + \Psi(\theta)\, \varepsilon_t | \mathcal{I}_t\right] = 0,\ \varepsilon_t \sim N(0, I) \quad (1)$$

where $\mathcal{I}_t$ is the information set of the agents at time t, y_t is a $m \times 1$ vector of endogenous variables (including both states and controls), ε_t is a $m_\varepsilon \times 1$ vector of exogenous shocks, Θ_{-1}, Θ_0, Θ_1 are $m \times m$ matrices, Ψ is an $m \times m_\varepsilon$ matrix. Those matrices are a function of θ, the vector of deep structural parameters of the model.

The traditional solution to this system has a state space representation of the form

$$y_t = A(\theta)\, y_{t-1} + B(\theta)\, \varepsilon_t \quad (2)$$

where A is a $m_A \times m_A$ matrix, B is $m_A \times m_\varepsilon$. This assumes that $\mathcal{I}_t = \{\varepsilon_t, y_{t-s}, s = 1, 2, ...\}$.

A natural way to compute conditional forecasts in a DSGE model with variables that are unobservable to the econometrician is to use the Kalman filter. In practice, conditional forecasting using the Kalman filter relies on the smoother to re-estimate the initial conditions for the unobservable variables, which is an advantage. However, the Kalman filter approach would typically treat the conditioning information as accurate[5]. Moreover, using the Kalman filter, we can only do hard conditioning. We cannot condition on a density or on an interval, which precludes the use of soft conditioning.

The approach used in this paper differs from the Kalman filter approach in that it does not assume that the conditioning information is accurate. In addition, it explicitly allows for the possibility of agents reacting to anticipated future events beyond one step ahead[6]. In order to allow for this possibility of agents reacting to future events anticipated several periods ahead, we generalize equation (2) above to

$$y_t = A(\theta) y_{t-1} + \sum_{j=0}^{n} B_j(\theta) \varepsilon_{t+j}$$

and in this case the information set of the agents is expanded to include future shocks, that is $\mathcal{I}_t = \{\varepsilon_{t+j}, y_{t-s}, s = 1, 2, ..., j = 0, 1, 2, ...\}$.

Appendix (A) shows how to derive matrices $B_j(\theta)$, $j = 1, 2, ..., n$. When $j = 0$, this solution is the same as the traditional one and the shocks are unanticipated. When $j > 0$, shocks can be foreseen. Clearly, an advantage of using structural models is that we can explicitly take into account the fact that people anticipate future events. In practice, one may assume that only a subset of shocks can be anticipated and the formula above includes such information assumptions as special cases.

We assume that the state of the economy is known at time T and we are interested in conditional forecasts k periods ahead. The k-step forecast at time T can be written as

$$\begin{aligned} y_{T+k} &= A^k y_T + \textstyle\sum_{j=0}^{n} \sum_{s=1}^{k} A^{k-s} B_j \varepsilon_{T+j+s-1} \\ &= \textstyle\sum_{t=0}^{n+k-1} \Phi_{k,t} \varepsilon_{T+t} \end{aligned}$$

Stacking all the forecasts up to period $T + k$, we get the following representation:

[5] Measurement errors can be introduced to account for uncertainty in the conditioning information, but it can be difficult to obtain estimates for their standard errors without estimating the other parameters of the model jointly, which means using historical data only. If there is a formal procedure that generates the conditioning information, one way to estimate the measurement errors parameters could simply be to take the standard deviation of the forecast errors of the procedure over history. Such procedure may not always be available.

[6] This approach, unlike the Kalman filter, does not change the initial conditions of the state vector.

$$\underbrace{\begin{bmatrix} y_{T+1} \\ y_{T+2} \\ \vdots \\ y_{T+k} \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} A \\ A^2 \\ \vdots \\ A^k \end{bmatrix} y_T +}_{\bar{Y}} \underbrace{\begin{bmatrix} \Phi_{1,1} & \cdots & \Phi_{1,n} & 0 & \cdots & 0 \\ \Phi_{2,1} & \cdots & \Phi_{2,n} & \Phi_{2,n+1} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \\ \Phi_{k,1} & \cdots & \Phi_{k,n} & \Phi_{k,n+1} & \cdots & \Phi_{k,n+k-1} \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} \varepsilon_{T+1} \\ \vdots \\ \varepsilon_{T+n} \\ \varepsilon_{T+n+1} \\ \vdots \\ \varepsilon_{T+n+k-1} \end{bmatrix}}_{\varepsilon} \quad (3)$$

When $n = 1$, the shocks are unanticipated and Φ takes a form that is analogous to the one used by Waggoner and Zha (1999)

$$\Phi = \begin{bmatrix} B & 0 & \cdots & 0 \\ AB & B & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ A^{k-1}B & A^{k-2}B & \cdots & B \end{bmatrix}$$

Equation (3) implies that

$$Y \sim N\left(\bar{Y}, \Phi\Phi'\right)$$

4 Conditional Forecasts and Probability Distributions

Suppose we are given the restriction

$$DY \sim TN\left(\mu, \Omega, [L, H]\right) \Longrightarrow D\bar{Y} + R\varepsilon \sim TN\left(\mu, \Omega, [L, H]\right)$$

where matrix D is $q \times mk$ and is assumed to be of full rank, μ is the mean of the truncated multivariate normal distribution denoted by TN, Ω, the covariance matrix, L is the lower bound, H is the upper bound and $R \equiv D\Phi$ is a $q \times (n + k - 1)\, m_\varepsilon$ matrix.

Using the model properties to translate the restrictions on Y into restrictions on the shocks, the expression above implies that

$$R\varepsilon \sim TN\left(\mu - D\bar{Y}, \Omega, [\underline{r}, \overline{r}]\right)$$

where , $\underline{r} \equiv L - D\bar{Y}$ and $\overline{r} \equiv H - D\bar{Y}$ are $q \times 1$ vectors. We assume that $R(\theta)$ is of rank $q \leq h \equiv (n+k-1)\, m_\varepsilon$.

4.1 Decomposition of shocks under conditioning

Because in general $q < h$, the covariance matrix of 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[7]. Consider the decomposition

$$\varepsilon = M_1\gamma_1 + M_2\gamma_2, \text{ with } \gamma_1 \sim N\left(0, I_{h-q}\right) \tag{4}$$

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

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

M_2, which is $h \times q$, 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

$$\gamma_2 \sim TN\left\{\left(RM_2\right)^{-1}\left(\mu - D\bar{Y}\right), \left(RM_2\right)^{-1}\Omega\left(M_2'R'\right)^{-1}, \left[\alpha_{low}, \alpha_{high}\right]\right\}$$

with $\alpha_{low} \equiv \left(RM_2\right)^{-1}\underline{r}(\theta)$ and $\alpha_{high} \equiv \left(RM_2\right)^{-1}\overline{r}(\theta)$, to vectors of dimension $q \times 1$.

We are interested in characterizing the distribution of ε, given the restrictions, and thereby that of the conditional forecasts Y. The derivations above show that in order to get ε that satisfies the restrictions, one can make independent draws for γ_1 and combine them, using equation (4), with draws from a truncated normal distribution for γ_2. Then with ε in hand, the next step is simply to use equation (3) in order to make forecasts for Y. In particular, we have the following conditional distributions for ε and for Y

$$p\left(\varepsilon|\mu, \Omega, L, H, \theta\right) = N\left[M_2E\gamma_2, M_1M_1' + M_2V\left(\gamma_2\right)M_2'\right]$$

$$p\left(Y|\mu, \Omega, L, H, \theta\right) = N\left[\bar{Y} + \Phi M_2E\gamma_2, \Phi\left(M_1M_1' + M_2V\left(\gamma_2\right)M_2'\right)\Phi'\right]$$

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

Hard conditioning Here we consider the case where $L = H$, implying that $r^* (\theta) = \underline{r} (\theta) = \overline{r} (\theta)$. 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 ε is as follows

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

Proposition 1 *Under hard conditions, the estimator for* $\hat{\varepsilon} \equiv E (\varepsilon|\mu, \Omega, L, H, \theta)$ *is equal to* $R' (RR')^{-1} r^*$ *and has the smallest variance among all linear estimators in* r^*.

The proof of the proposition is given in the appendix. This alternative formula, $\hat{\varepsilon} = R' (RR')^{-1} r^*$, of the estimator is the one presented in Waggoner and Zha (1999). 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.

Soft and no conditioning When Ω and μ are unknown, they can simply be replaced by their theoretical counterparts given by the model. That is $\hat{\Omega} = RR'$ and $\hat{\mu} = D\bar{Y}$. In that case, we have simple expressions for γ_2[8].

$$\gamma_2 \sim TN \{0, I, [\alpha_{low}, \alpha_{high}]\}$$

with all the elements in the $\gamma_2 \equiv [\gamma_{21}, \gamma_{22}, ..., \gamma_{2q}]'$ vector being independent of each other. This independence of the γ_2 elements means that, no Gibbs sampling or other multivariate sampling procedure is required for computing the distribution. For $i = 1, 2, ..., q$, we have

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

[8] It can be verified that $(RM_2)^{-1} RR' \left((RM_2)^{-1}\right)' = I_{q\times q}$

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

Letting $(\alpha_{1i},\alpha_{2i})\longrightarrow(-\infty,\infty)$ 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

$$\begin{aligned} E\left(\varepsilon|C,\theta\right) &= 0 \qquad V\left(\varepsilon|C,\theta\right) &= M_1M_1'+M_2I_qM_2' \\ & & = I_k \end{aligned}$$

which is the unconditional forecast case. We see that hard conditions imply a lower variance than soft conditions, which in turn imply a lower variance than no conditions.

If Ω is known, however, one may need to resort to simulation to compute the distribution for γ_2 as Ω may introduce nonzero correlations among the elements in γ_2[9]. In any case, those derivations show that we condition with a continuum of degrees of uncertainty around a central tendency.

4.2 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

$$y_{T+h}=f_h\left(y_T\right) \tag{5}$$

where $f_h(\cdot)$ 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 $p\left(\theta|Y_T\right)$. And then for $i=1,2,...,N$, with N the number of simulations,

[9] There are different sampling algorithms for the multivariate truncated normal distribution, see for instance Geweke (1991) or Rodriguez-Yam et al. (2004).

one would have to a) solve the model in reduced form and recover the starting values of the unobservables, b) generate forecasts $y^{(i)}_{T+1}, y^{(i)}_{T+2}, ..., y^{(i)}_{T+h}$, from $p\left(y_{T+1}, y_{T+2}, ..., y_{T+h}|\theta^{(i-1)}\right)$, c) augment the original data set with the forecasts and estimate a new value of θ, d) go back to a). But this Gibbs sampling algorithm 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[10]. One could potentially circumvent this problem by adopting a less computationally intensive method of estimation such 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 then assumed to remain invariant to additional information[11]. In this case, the sampling algorithm is the following:

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

1. draw $\theta^{(i)}$ from a chosen distribution (calibration, prior, posterior, mode).
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
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 $\left\{y^{(i)}_{T+1}, y^{(i)}_{T+2}, ..., y^{(i)}_{T+h}\right\}$.

[10]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.

[11]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.

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 considered 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 (6) 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.

$$\begin{aligned}
y_t &= E_t y_{t+1} - [\tau + \alpha(2-\alpha)(1-\tau)](R_t - E_t\pi_{t+1}) - \rho_z z_t \\
&\quad -\alpha[\tau + \alpha(2-\alpha)(1-\tau)] E_t \Delta q_{t+1} + \tfrac{\alpha(2-\alpha)(1-\tau)}{\tau} E_t \Delta y_{t+1}^* \\
\pi_t &= \beta E_t \pi_{t+1} + \alpha\beta E_t \Delta q_{t+1} - \alpha \Delta q_t + \tfrac{\kappa}{\tau+\alpha(2-\alpha)(1-\tau)}(y_t - \bar{y}_t) \\
\pi_t &= \Delta e_t + (1-\alpha)\Delta q_t + \pi_t^* \\
R_t &= \rho_R R_{t-1} + (1-\rho_R)[\psi_1 \pi_t + \psi_2 y_t + \psi_3 \Delta e_t] + \varepsilon_t^R \\
\bar{y}_t &= -\tfrac{\alpha(2-\alpha)(1-\tau)}{\tau} y_t^*
\end{aligned} \tag{6}$$

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

$$\begin{aligned}
z_t &= \rho_z z_{t-1} + \varepsilon_{z,t} & y_t^* &= \rho_{y^*} y_{t-1}^* + \varepsilon_{y^*,t} \\
\Delta q_t &= \rho_q \Delta q_{t-1} + \varepsilon_{q,t} & \pi_t^* &= \rho_{y^*} \pi_{t-1}^* + \varepsilon_{\pi^*,t}
\end{aligned} \tag{7}$$

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 exchange rate, the interest rate, terms of trade and inflation 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[12]. More to the point, while it is hard to guess or obtain accurate values of an unrealized conditioning variable, it is far much easier to guess a range for the variable. With the soft conditioning technique presented we can make that range as wide as we want, reflecting the uncertainty we have about the location of the conditioning variable and so the exercise is not restrictive.

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 the conditioning variable over history[13]. We consider 3 degrees of uncertainty: 0, 50, and 100% , where a value of 0 implies hard conditioning, while a value of 100 implies unconditional forecasts. We repeat the exercise for 1, 4, 8, and 12 periods anticipated. Note that when the number of periods is 1, anticipated and unanticipated events yield the same results.

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

[12] 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).

[13] 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 as discussed in the previous section. But there is no perfect way of generating the uncertainty measure to use for the computation of soft-conditional forecasts.

prises 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 dataset runs from 1970Q1 to 2002Q4 and unlike in the Lubik-Schorfheide paper, we use uniform priors for all the parameters. The sample from 1970Q1 to 1994Q1 is used for the first estimation. 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. 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 (8).

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

$$RMSFE_i(h) = \sqrt{\frac{1}{N_h}\sum_{d=1}^{N_h} e_{i,d}^2(h)} \tag{8}$$

We also consider a multivariate measure of point forecast accuracy based on the scaled h-step-ahead Mean Squared Error matrix (see equation (9)) used by Adolfson et al. (2005)

$$\Omega_M(h) = \frac{1}{N_h}\sum_{d=1}^{N_h} \tilde{e}_{\cdot,d}(h)\,\tilde{e}'_{\cdot,d}(h), \text{ with } \tilde{e}(h) \equiv M^{-\frac{1}{2}}e(h) \tag{9}$$

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 $\ln(|\Omega_M(h)|)$, 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 $|\Omega_M(h)| = 0$ 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.

	DY_OBS	PAI_OBS	R_OBS	DQ	DE
Model					
DY_OBS	1				
PAI_OBS	-0.2538	1			
R_OBS	-0.178	0.2802	1		
DQ	0.0159	0.145	-0.1464	1	
DE	-0.1133	-0.1679	0.1892	-0.5351	1
Data					
DY_OBS	1				
PAI_OBS	-0.2249	1			
R_OBS	-0.2422	0.5518	1		
DQ	0.0987	0.1418	-0.1004	1	
DE	-0.0579	-0.2150	-0.0256	-0.3421	1

Table 1: Correlation in the model vs Correlations in the data

6 Results

The question we try to answer here is if the agents in the LS07 economy had known in advance and with various degrees of uncertainty, the shocks that would push their decisions towards the observed outcomes for each conditioning variable, how accurate would be the predictions of the model for the other variables. Before looking at the forecast performance, it is instructive to compare the correlations implied by the model to those implied by the data. The intuition is that the conditioning information is potentially useful for improving the forecasts of other variables of interest if the model is able to capture the relationship linking those variables. Table 1 displays the correlations between the variables. Despite, missing the magnitude, most of the correlations have the correct sign. Notable exceptions are the correlations between output growth (DY_OBS) and exchange rate changes (DE) and the correlation between the interest rate (R_OBS) and the exchange rate. This tells us that having information on exchange rate changes may not improve or may even worsen our forecats for output growth and interest rate and vice-versa.

We now turn to assessing the forecast performance. Given that RMSE and MAE lead to qualitatively similar results, only the results from RMSE are reported for reasons of brevity. Figures 2 to 5 present the results of the conditional forecast exercise in terms of RMSE (columns 1 to 5) for output growth, inflation, interest rate, terms of trade changes, exchange rate changes respectively and in terms of log determinant, our overall measure of forecast accuracy. Each row represents the effect of conditioning on the variable whose name is on the left. The plots are arranged such that the variables on the main diagonal represent both the conditioning variable and the effects of conditioning from the same variable. By construction then, given that the conditioning information is always accurate, hard conditions will always outperform soft conditions, which in turn will be better than no conditions.

Starting the analysis of the results with the case of unanticipated shocks (figure 2), the first row reveals that there are potential gains in terms of better forecasts for inflation from conditioning on output growth as the RMSE for inflation are lower for hard and soft conditions than for unconditional forecasts. The gains, however, from having a very accurate information on output growth are not substantial since hard conditions do not significantly outperform soft conditions. The first row also reveals some gains in terms of interest rate forecasts, but only towards the end of the forecasting period. This suggests that if the interest rate reacts to output growth, it probably does so with a delay. The same row shows that accurate information on output growth worsens the forecasts for exchange rate changes and interestingly, those results echo the implications we drew from table 1.

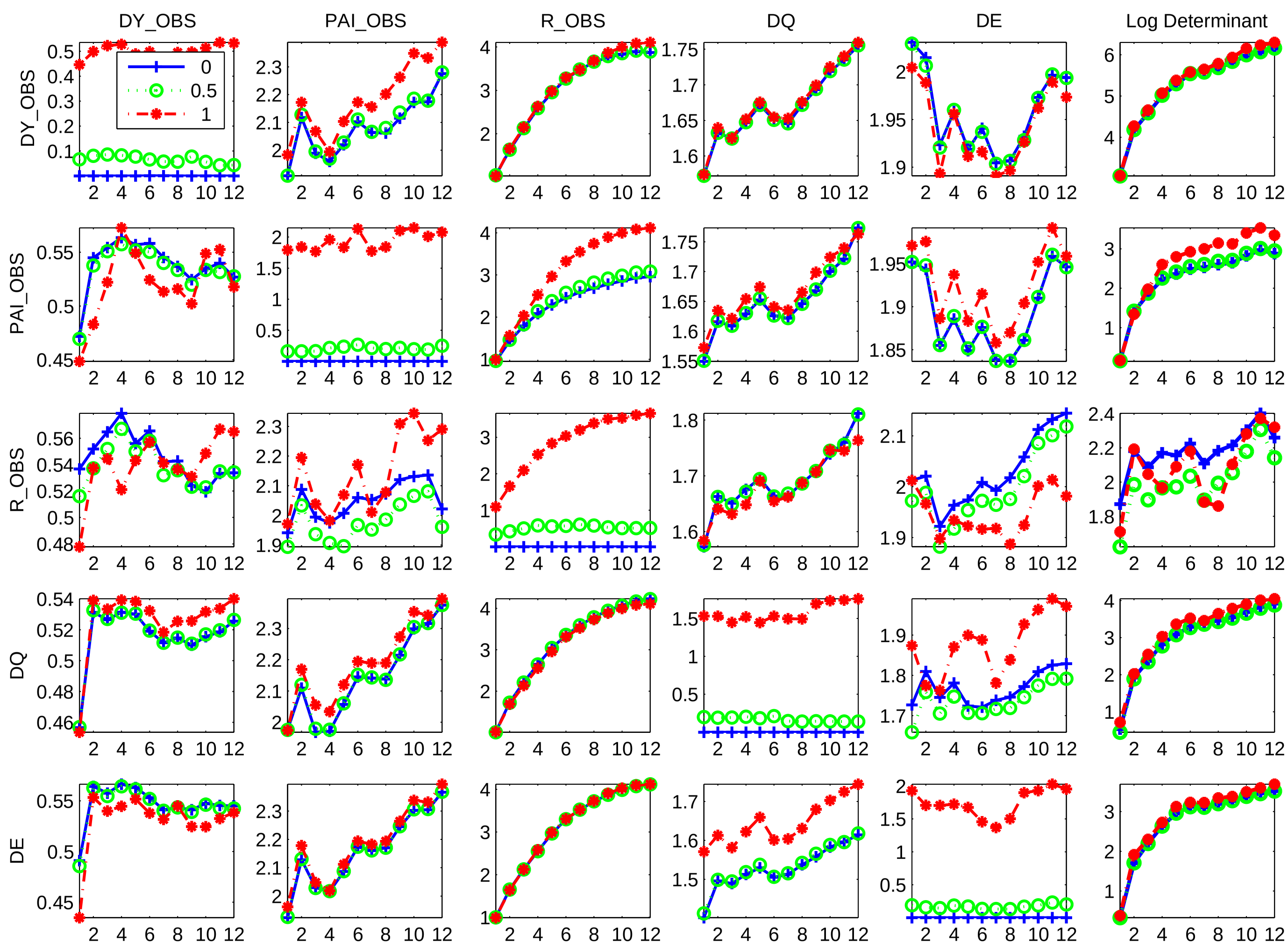

Figure 2: RMSE and Log Determinant statistics for conditional forecasts with 1-step-ahead anticipation of the conditioning variable to the left. Hard conditions(0), soft conditions(0.5) and No condition (1)

The second row in figure 2 shows a strong link between inflation and interest rate as conditioning is superior to no conditioning. And here the gains seem to be increasing with the forecast horizon. Inflation also helps in the forecasts of both terms of trade changes and exchange rate changes and again in sync with the results from table 1. Not surprisingly then, the verdict from the log determinant statistics in the last column is in favor of conditioning.

Although information on output growth helped predict inflation in the first row, in the second row information on inflation does not help predict output growth despite the fact that the model captures the correct sign of the correlation between the two variables. This warns us that even in cases where the model captures the correct sign of the correlation, we cannot say a priori in which direction the causality goes.

Information on interest rate may help predict inflation as shown in the third row. However, in this case, soft conditions are better than hard conditions, which are too tight. Conditioning on terms of trade changes (see row number four), helps predict output growth as well as inflation and exchange rate changes and here again, soft conditions dominate hard conditions. The link between terms of trade changes and exchange rate changes remains strong when the conditioning variable is exchange rate changes, which also helps predict inflation albeit not significantly. Two other strong links are noticeable: the one between inflation and interest rate on the one hand and the one between terms of trade changes and inflation on the other.

In all of the other plots in figure 2, unconditional forecasts are not uniformly dominated by conditional forecasts which in some cases even worsen the forecasts. Put differently, in the presence of a misspecified model hard conditions do not necessarily dominate soft conditions even if the conditioning information is accurate. This is seen most starkly for instance in the RMSEs of output growth conditional on inflation or on exchange rate changes, and terms of trade and exchange rate changes conditional on the interest rate. 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 to the variables whose specification should be improved.

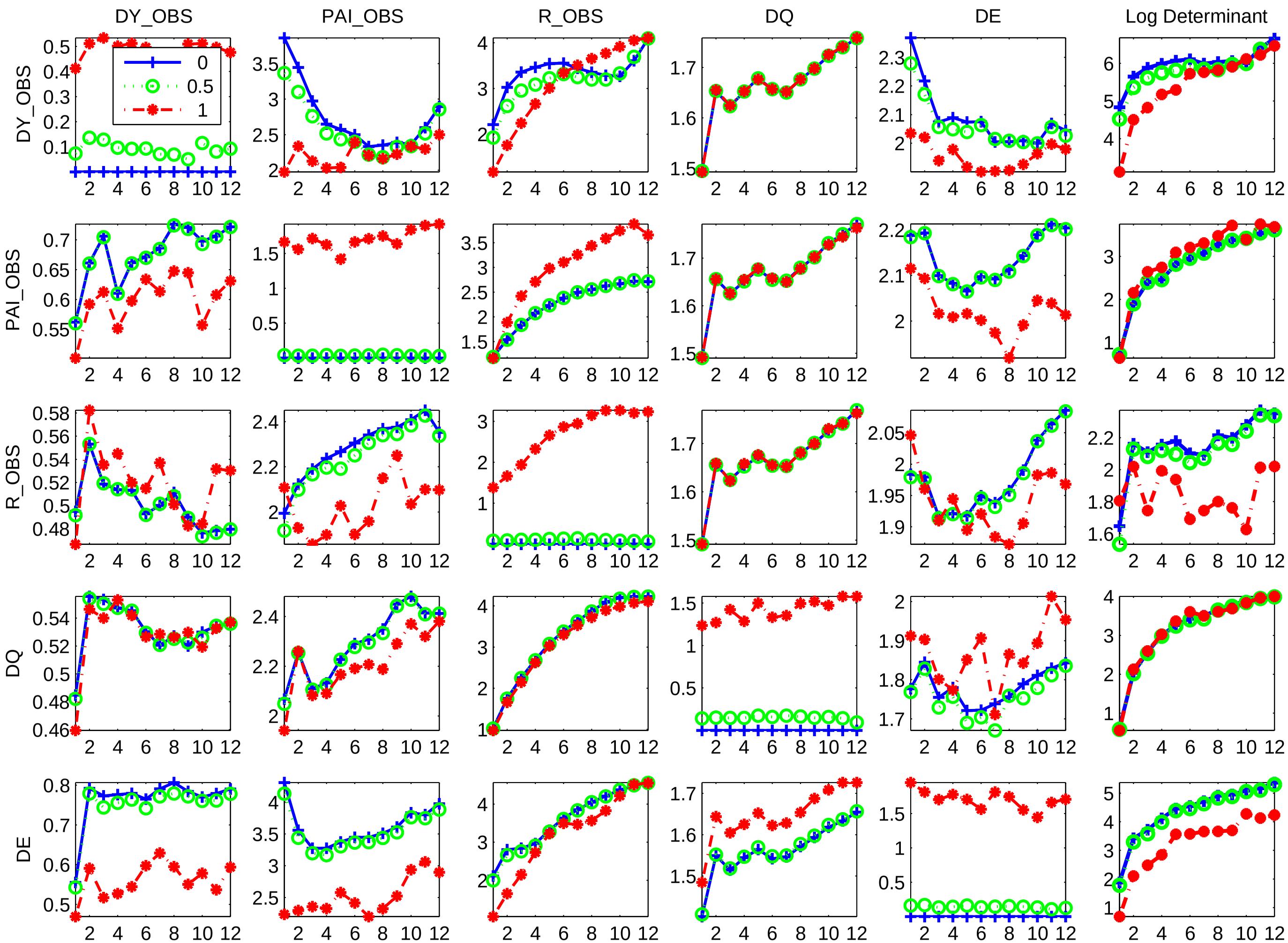

Figure 3: RMSE and Log Determinant statistics for conditional forecasts with 4-step-ahead anticipation of the conditioning variable to the left. Hard conditions(0), soft conditions(0.5) and No condition (1)

So far we have assumed that the agents did not observe in advance the sequence of shocks beyond one period ahead. But as discussed in the introduction, if conditioning information is available with the potential of improving forecasts, rational agents will exploit that information. The natural next step then is to extend the number of periods for which the agents in the model can observe shocks. The first extension considered is 4 periods ahead. The results of the conditional forecast exercise with 4 anticipated periods ahead are displayed in figure 3. The forecasting performance deteriorates substantially. Output growth in the first row, continues to help predict the interest rate towards the end of the forecast horizon. Inflation in the second row helps predict interest rates too. The two-way relationship between terms of trade and exchange rate changes is maintained and in particular, the RMSEs for exchange rate changes conditional on terms of trade changes remind us that even if the information is accurate, hard conditioning may be too tight.

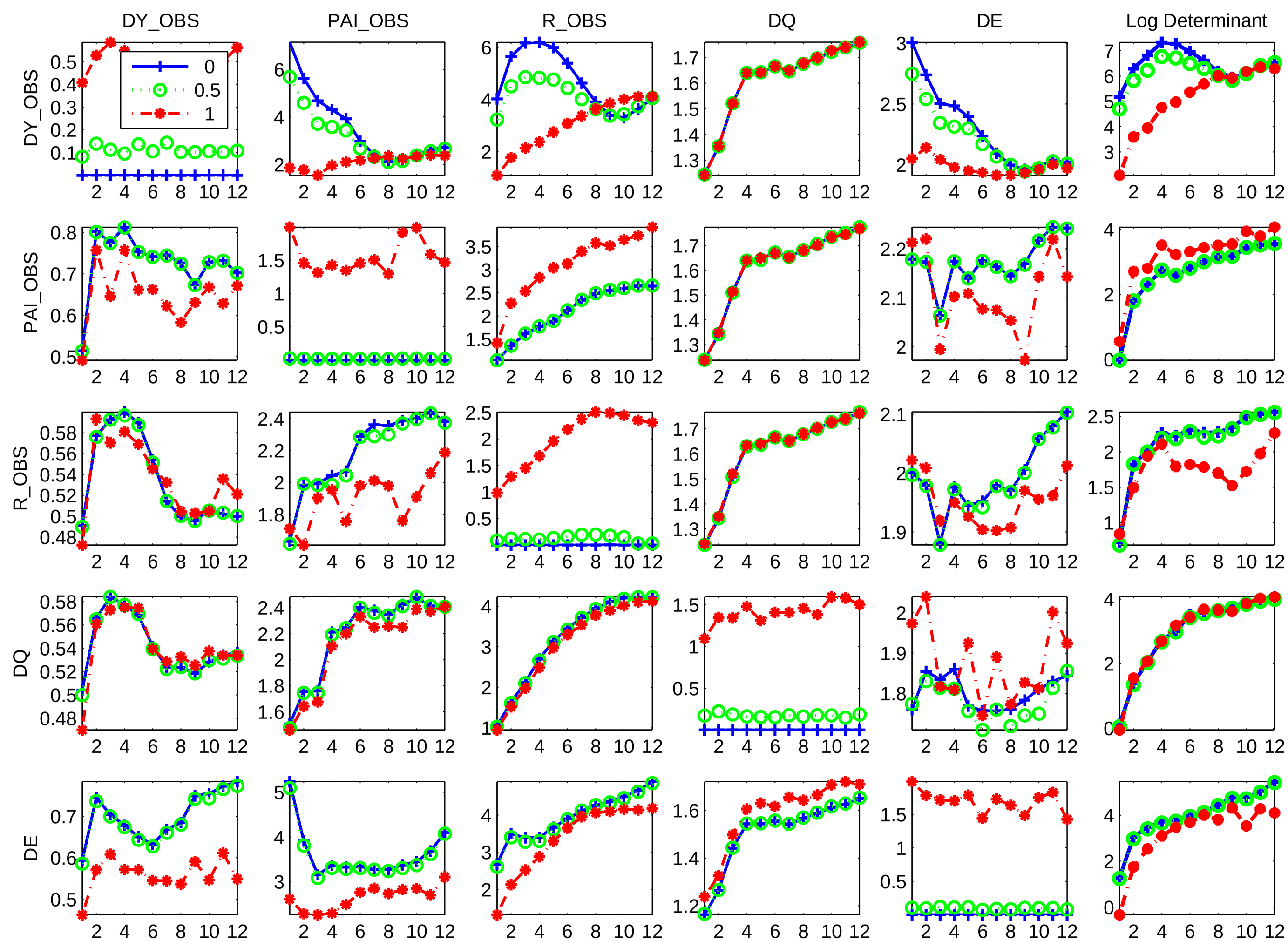

Figure 4: RMSE and Log Determinant statistics for conditional forecasts with 8-step-ahead anticipation of the conditioning variable to the left. Hard conditions(0), soft conditions(0.5) and No condition (1)

Allowing agents to anticipate shocks 8 periods ahead, the results are somewhat similar to those for 4-step ahead anticipated shocks (see figure 4). Output growth still helps to predict interest rate changes, while the explanatory power of inflation on the interest rate continues to be strong, just as the link between exchange rate changes and terms of trade changes. The explanatory power of interest rate on output growth becomes more and more important as agents are allowed to observe future shocks beyond one step ahead and this is confirmed also for 12-step ahead anticipated shocks (figure 5).

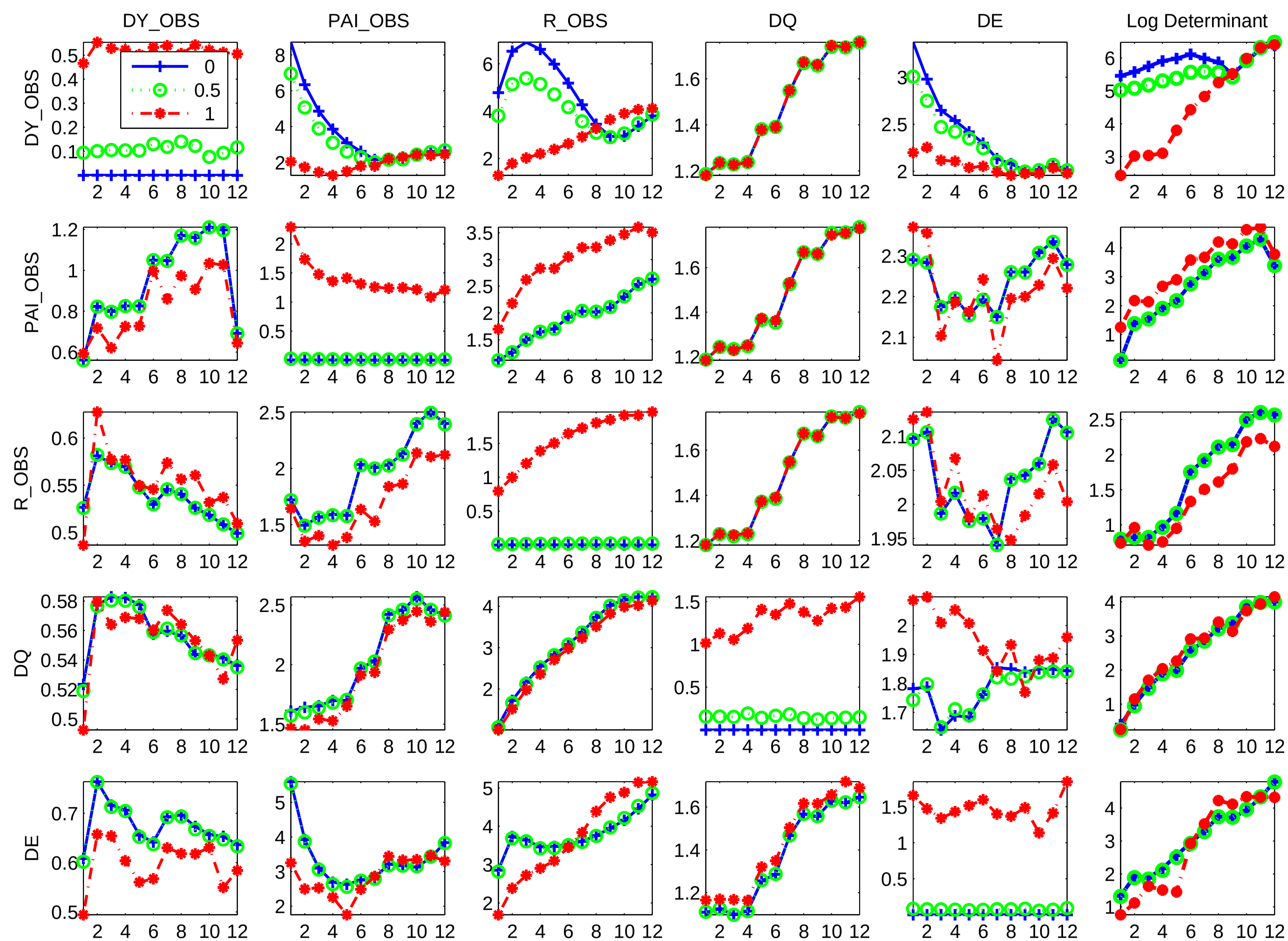

Figure 5: RMSE and Log Determinant statistics for conditional forecasts with 12-step-ahead anticipation of the conditioning variable to the left. Hard conditions(0), soft conditions(0.5) and No condition (1)

All in all, looking at the results of the log determinant statistics from figures 2 to 5, conditioning on output growth in the LS07 model will tend to worsen the forecasts for other variables in the first periods of the forecast horizon, with the exception of the interest rate but only when agents do not anticipate shocks beyond one period. Information one inflation helps improve the forecasts of interest rates and those forecasts seem to improve as the horizon for which agents can anticipate future shocks increases. Forecasting on the interest rate does not seem to improve the forecasts for other variables in this model. Information on terms of trade has a positive impact on the forecasts for the exchange rate and vice-versa. And so, although the model is misspecified, it does capture some of the correlations between some of the conditioning variables (like inflation) and the other variables (like interest rate), such that information on the conditioning variable implies improved forecasts for some of the other variables in the model. This suggests that even if a model is misspecified along one dimension, it may still be useful in explaining the behavior of some other variables.

It is difficult to trace the exact reasons why, except for inflation conditional interest rate, conditioning information does not significantly improves forecasts as agents are allowed to know more about shocks beyond one step ahead. It might be the case that the observed data are not generated with the agents having information about the future. In any case, it is well possible to imagine environments in which the anticipated shocks assumption would be more relevant than the unanticipated shocks assumption. 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.

7 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 Generalized solution

Consider the model

$$E\left[\Theta_{-1}y_{t-1} + \Theta_0 y_t + \Theta_1 y_{t+1} + \Psi\varepsilon_t\right] = 0$$

We use an undetermined coefficients method to guess a solution of the form

$$y_t = Ay_{t-1} + \sum_{j=1}^{n} B_j \varepsilon_{t+j} \tag{10}$$

This guess implies that A solves matrix polynomial

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

this could be solved in various ways using either eigensystem methods as in Anderson and Moore (1983), King and Watson (1998), Schur decomposition methods as in Klein (1999), Sims (1996) or undetermined coefficient methods as in Binder and Pesaran (1994) and Uhlig (1999).

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

$$B_1 = -\left[\Theta_1 A + \Theta_0\right]^{-1}\Psi$$

$$B_j = \left(-\left[\Theta_1 A + \Theta_0\right]^{-1}\Theta_1\right)^{j-1} B_1 \qquad j = 2, ..., n$$

The solution for the B_j depends not only on the reduced form solution through A, but also directly on the structural form with matrices Θ_1 and Θ_0. When $n = 1$, the solution is that of any rational expectation model as in equation (2). When $n > 1$, 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. This solution is general in the sense that it allows for the possibility of some shocks being unanticipated and some others not.

B Proof of the proposition

Proof. We first show that $R'(RR')^{-1}r^*$ has the smallest variance among all linear estimators. If $\hat{\varepsilon}$ is linear in r^*, then it can be written as $\hat{\varepsilon} = Ar^* = AR\varepsilon$. We then have that $\hat{\varepsilon} - \varepsilon = (AR - I)\varepsilon$ and the variance of this expression is given by: $var(\hat{\varepsilon} - \varepsilon) = (AR - I)\Sigma_\varepsilon (AR - I)'$. Minimizing this expression for A, we find $A = \Sigma_\varepsilon R'[R\Sigma_\varepsilon R']^{-1}$. Since the covariance matrix Σ_ε of ε is the identity matrix, the result follows.

We now turn to showing that $R'(RR')^{-1}r^* = M_2(RM_2)^{-1}r^*$. Using the singular value decomposition of R, we have $R = USM'$, where U and M are unitary matrices satisfying $U'U = I$ and $M'M = I$, and S is a diagonal matrix with the singular values of R on its diagonal. $M = [M_2, M_1]$ where M_2 is an orthonormal basis for the column space of R and M_1 is the null space of R. This implies that M_2 is associated with nonzero singular values of R, while M_1 is associated with zero singular values. It follows that $R = USM' = U[D,0][M_2, M_1]' = UDM_2'$.

Now

$$
\begin{aligned}
R'(RR')^{-1}r^* &= [UDM_2']'\left(R[UDM_2']'\right)^{-1}r^* \\
&= M_2D'U'(RM_2D'U')^{-1}r^* \\
&= M_2D'U'(D'U')^{-1}(RM_2)^{-1}r^* \\
&= M_2(RM_2)^{-1}r^*
\end{aligned}
$$

. ■

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.

Anderson, G. and G. Moore (1983). An efficient procedure for solving linear perfect foresight models. Unpublished manuscript.

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.

Binder, M. and M. H. Pesaran (1994). Multivariate rational expectations models and macroeconometric modelling: A review and some new results. Seminar paper.

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.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student t-distributions subject to linear constraints. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 571–578.

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

King, R. G. and M. W. Watson (1998). The solution of singular linear difference systems under rational expectations. *International Economic Review 39(4)*, 1015–1026.

Klein, P. (1999). Using the generalized schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control*.

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.

Rodriguez-Yam, G., R. A. Davis, and L. L. Scharf (2004). Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. Technical report, Department of Statistics, Colorado State University.

Sims, C. A. (1996). Solving linear rational expectations models. Seminar paper.

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).

Uhlig, H. (1999). A toolkit for analyzing nonlinear dynamic stochastic models easily. User guide.

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