Jan-Erik Antipin
Jani Luoto

Forecasting performance of the small-scale hybrid New Keynesian model

HELSINGIN KAUPPAKORKEAKOULU
HELSINKI SCHOOL OF ECONOMICS

W-451

Jan-Erik Antipin* – Jani Luoto**

Forecasting performance of the small-scale hybrid New Keynesian model

*Helsinki School of Economics,
Quantitative Methods of Economics and Management Science,
University of Tampere and
UNU-WIDER

**University of Jyväskylä,
School of Business and Economics

June
January
2008

HELSINGIN KAUPPAKORKEAKOULU
HELSINKI SCHOOL OF ECONOMICS
WORKING PAPERS
W-451

HELSINGIN KAUPPAKORKEAKOULU
HELSINKI SCHOOL OF ECONOMICS
PL 1210
FI-00101 HELSINKI
FINLAND

ISSN 1235-5674
(Electronic working paper)
ISBN 978-952-488-261-3

Helsinki School of Economics -
HSE Print 2008

Forecasting Performance of the Small-Scale Hybrid New Keynesian Model

Jan-Erik Antipin
University of Tampere,
Helsinki School of Economics
and UNU-WIDER

Jani Luoto
School of Business and Economics
University of Jyväskylä
PL 35
40014 University of Jyväskylä
FINLAND

Abstract

This paper uses quarterly ex post and real-time U.S. data to show that the very simple hybrid New Keynesian model of Clarida, Galí and Gertler [1999. The Science of Monetary Policy: A New Keynesian Perspective. Journal of Economic Literature 37, 1661-1707] can provide forecasts comparable to those based on Bayesian reduced-form vector autoregressive models. The issue is important, since several recent papers have suggested different ways to improve the forecast performance of New Keynesian models at the cost of increasing the complexity of model mechanisms, thus reducing the practicability of these approaches.

JEL Classification: C11; C32; E37; E47.

Keywords: New Keynesian model; Forecasting; Real-time data; Bayesian inference; Vector autoregressive models.

We thank Dr. Jesper Lindé (Sveriges Riksbanken) for providing the estimation data of his paper and supporting help in the first-stage of the model estimations.

1. Introduction

There is an increasing volume of literature focusing on developing New Keynesian (NK) models suitable for forecasting and quantitative policy analysis; see Sungbae and Schorfheide (2007) and references therein. Within this literature, Smets and Wouters (2003, 2005 and 2007), Christiano et al. (2005), Adolfson et al. (2007a), and Adolfson et al. (2005, 2008), construct large-scale NK models aiming to find a structural macroeconomic model which has a fit comparable to that of reduced-form Bayesian vector autoregressive (VAR) models. In these studies, additional shocks, frictions and measurement errors are introduced to the NK model mechanisms until the desired fit is achieved. This approach ignores model uncertainty, leading to inferences which are over-confident and decisions which are riskier than the policy-maker believes them to be. A promising alternative strategy is provided by Del Negro and Schorfheide (2004). In their approach an NK model is used to generate a prior distribution for the parameters of the VAR to improve the forecast and policy analysis performance of these models. Although this approach is promising, it is nonetheless complicated and the numerical methods required in estimation are time-consuming. The practicability of this approach can therefore be questioned; see also Del Negro et al. (2007). In the light of recent NK literature, it would be thus interesting to see whether a simple NK model, including only few shocks and the standard price rigidity, can have a fit comparable to other forecasting methods such as the Bayesian VARs commonly used as a benchmark.

This paper has two objectives. First, it provides a method for the Bayesian analysis of a simple hybrid NK model of Clarida et al. (1999). The method is very easy to implement and leads to savings in the CPU time required in posterior simulation, compared to the commonly used Kalman filter approach. Lindé (2005) estimates a version of the hybrid NK model with the full information maximum likelihood (FIML) method using U.S. data. We instead adopt a Bayesian full information framework, since the FIML estimates turned out to be very sensitive to starting values and since Bayesian methods allow incorporation of prior information which facilitates numerical maximization.

Our second objective is to compare the forecasting properties of the hybrid NK model against commonly used forecasting tools such as Bayesian VARs and naïve forecasts based on univariate random walks. Using quarterly U.S. data we show that the hybrid model can provide forecasts of key macroeconomic variables, inflation and short-term nominal interest rate, and a measure of the output gap comparable to forecasts based on reduced-form Bayesian VARs. Our results also indicate that the hybrid model predicts more accurately than naïve forecasts based on univariate random walks. In particular, these results hold for both ex post data and real-time data, which are available to policy-makers when forecasts are being made. Our results also confirm the finding of Smets and Wouters (2007) that the cross-equation restrictions implied by NK models work especially well in forecasting at medium-term horizons (from four to twelve quarters). For policy-makers, comparisons of forecasts at longer than one quarter horizon are of interest, since policy actions typically depend on expected future developments in the economy.

Finally, we find two major reasons for the good forecasting performance of the hybrid model. Firstly, the model allows both for the endogenous persistence in inflation and output and for the persistence of exogenous shock processes. This approach is commonly used in large-scale NK models, which forecast well. Secondly, our joint prior is well designed in allowing the parameters to be estimated fairly freely, while being sufficiently informative to keep the posterior distribution away from economically non-meaningful values.

The remainder of the paper is organized as follows. In section 2, we discuss the model, the prior and the data. We continue the analysis by reporting the posterior distributions of the parameters. In section 3, we explain the forecasting comparison methods, and present and discuss the results of a forecasting exercise. Section 4 concludes the paper.

2. Likelihood, Prior, Data, and Posterior

In this section we introduce a hybrid NK model. Its likelihood and the joint prior density function of the structural parameters are specified. We then describe the data and continue the analysis by reporting the posterior distributions of the parameters.

2.1. Model Likelihood

Let us consider the following hybrid NK model for period t inflation[1], π_t, and a measure of the output gap, x_t, respectively,

$$\pi_t = \alpha E_t \pi_{t+1} + (1-\alpha)\pi_{t-1} + \gamma x_t + \varepsilon_{\pi,t}, \tag{1}$$

$$x_t = \beta E_t x_{t+1} + (1-\beta)x_{t-1} - \beta_r (R_t - E_t \pi_{t+1}) + \varepsilon_{x,t}, \tag{2}$$

where parameters α and β satisfy the conditions $0 \le \alpha \le 1$ and $0 \le \beta \le 1$. Equation (1) is the hybrid New-Keynesian Phillips curve (NKPC), similar to that analyzed in Rudd and Whelan (2006), while Equation (2) is the aggregate demand equation. The model is very close to that carefully studied in Clarida et al. (1999).

The disturbance terms $\varepsilon_{\pi,t}$ and $\varepsilon_{x,t}$ in Equations (1) and (2) are assumed to follow univariate AR(1) processes:

$$\varepsilon_{\pi,t} = \rho_\pi \varepsilon_{\pi,t-1} + u_{\pi,t}, \tag{3}$$

$$\varepsilon_{x,t} = \rho_x \varepsilon_{x,t-1} + u_{x,t}, \tag{4}$$

where $\rho_\pi, \rho_x \in [-1,1]$, and $u_{\pi,t}$ and $u_{x,t}$ are independently and identically distributed (i.i.d.) random variables with zero means and variances σ_π^2 and σ_x^2, respectively.

[1]Price inflation is defined as the percent change in the price level from $t-1$ to t.

We close the model with the following Taylor rule for the nominal interest rate R_t,

$$R_t = (1-\rho)(\gamma_\pi \pi_t + \gamma_x x_t) + \rho R_{t-1} + \varepsilon_{R,t}, \tag{5}$$

where the parameter $\rho \in [0, 1]$ measures the degree of interest rate smoothing, the disturbance term $\varepsilon_{R,t}$ obeys $\varepsilon_{R,t} = \rho_R \varepsilon_{R,t\text{-}1} + u_{R,t}$, $\rho_R \in [-1,1]$, and $u_{R,t}$ is an i.i.d. random variable with zero mean and variance σ_R^2.

The model in Equations (1)-(5) can be solved analytically by using standard first-order log-linear methods. In particular, this paper follows Lindé (2005) in applying the solution algorithm of Söderlind (1999). The solution gives the equilibrium law of motion for the relevant state variables. Specifically, the state equation is given by $z_t = Cz_{t\text{-}1} + v_t$, where $z_t = (\varepsilon_{\pi,t}, \varepsilon_{x,t}, \varepsilon_{R,t}, \pi_{t\text{-}1}, x_{t\text{-}1}, R_{t\text{-}1})'$, $v_t = (u_{\pi,t}, u_{x,t}, u_{R,t}, 0, 0, 0)'$ and C is a nonlinear function of structural parameters. Given that the shocks are normally distributed and that the vector of observables $y_t = (\pi_t, x_t, R_t)'$ is a linear combination of the state variables, the common approach is to specify a recursive likelihood function for the model using the Kalman filter. The estimates of the model can then be obtained using standard non-linear optimization methods.

Alternatively, the analytical solution of the model can be written as a full information system of the vector of observables; see Lindé (2005). Specifically,

$$y_t = C_y y_{t-1} + C_\varepsilon \varepsilon_t, \tag{6}$$

where $\varepsilon_t = (\varepsilon_{\pi,t}, \varepsilon_{x,t}, \varepsilon_{R,t})'$ and C_y and C_ε are partitions of the solution matrix C conformably with y_t and ε_t, respectively.

Then denote $\varepsilon_t = \mathrm{P}\varepsilon_{t\text{-}1}$, where P is a diagonal matrix whose diagonal entries are given by ρ_π, ρ_x and ρ_R. The likelihood function for a sample of T observations can be written as

$$L(Y;\theta) \propto |C_\varepsilon|^{-T} |\Lambda|^{-T/2} \exp\left\{-0.5 \times tr\left(U\Lambda^{-1}U'\right)\right\}, \tag{7}$$

where $\theta = (\beta, \alpha, \beta_r, \gamma, \gamma_\pi, \gamma_x, \rho, \rho_\pi, \rho_x, \rho_R, \sigma_\pi, \sigma_x, \sigma_R)'$ is a vector comprising all model parameters and Λ a diagonal covariance matrix with diagonal entries σ_π^2, σ_x^2 and σ_R^2. Furthermore, the tth rows of $(T \times m)$ matrices Y and U are given by y_t' and u_t', respectively, where m is the number of observables and

$$u_t = C_\varepsilon^{-1}\left(y_t - \left(C_y + C_\varepsilon \mathrm{P} C_\varepsilon^{-1}\right) y_{t-1} + C_\varepsilon \mathrm{P} C_\varepsilon^{-1} C_y y_{t-2}\right). \tag{8}$$

In what follows, we adopt the full information approach of Equation (7), since the optimization algorithm based on it proved faster than the algorithm based on the recursive Kalman filter. Specifically, the Kalman filter approach requires roughly 4.5 times as much CPU time for posterior simulation as our approach (with a sample of 200 observations). Furthermore, both estimation methods were also found to produce similar results.

The model described in Equations (1)-(5) contains 13 parameters, collected in θ. It is fairly easy to see that all parameters are identifiable from the data. However, the maximum likelihood (ML) estimation of the model turned out to be a challenging task. In particular, the ML estimates of the parameters were very sensitive to the starting values of maximization due to a multimodal likelihood. This problem remained even when the parameter space was restricted to an economically feasible region. To illustrate this problem we give an example from the previous literature. Lindé (2005) estimates a version of the model in Equation (1)-(5) on U.S. data with full information maximum likelihood (FIML).[2] He finds positive and highly significant parameter estimates for the slope coefficients γ (≈ 0.05) and β_r (≈ 0.09). However, there exists a local equilibrium in which the likelihood is higher than that in Lindé's solution. At this equilibrium the slope coefficients γ and β_r are still positive, but rather close to zero. According to Lindé (2005),

[2] Lindé (2005) adds additional lags in the aggregate demand equation (2) and the monetary policy rule (3) to make disturbance terms $\varepsilon_{x,t}$ and $\varepsilon_{R,t}$ white noise.

the estimation can with different starting values converge to local equilibria with more or less plausible parameter values. To facilitate numerical maximization, we suggest using Bayesian methods, which allow incorporation of prior beliefs on parameters. While restricting, for example, the slope coefficients γ and β_r to be equal to some theoretical values gives an example of a very strong prior belief, other kinds of beliefs cannot easily be considered in the classical framework.

As seen in current literature, Bayesian methods have become a standard workhorse in analysing the NK models. Sungbae and Schorfheide (2007) provide an excellent review of the Bayesian methods developed in recent years to estimate and evaluate this class of models (see also Adolfson et al., 2007b). Rather than elaborating the details of Bayesian methods in analysing the NK models, which is already done in Sungbae and Schorfeide (2007), we discuss our choices of marginal prior distributions in the next subsection.

2.2 Marginal Priors

The starting-point in the Bayesian analysis is to determine the prior density function of the parameters, $p(\theta)$, which together with the likelihood function (7) yields the posterior density

$$q(\theta|Y) = \frac{p(\theta)L(Y;\theta)}{\int p(\theta)L(Y;\theta)d\theta}. \tag{9}$$

A typical informative prior reflects the researcher's subjective beliefs, summarizes information from the data not included in the estimation sample, or is based on both of them. Often the underlying economic theory provides a natural starting-point for the prior elicitation. We will use a very simple structural model as the basis of our prior knowledge. The model can be obtained by log-linearizing the aggregation of individual firms' pricing decisions and the consumption Euler equation without using ad hoc assumptions such as backward inflation indexation or habit formation in consumption. Specifically, the prior means of the parameters in θ are based on the following model,

$$\pi_t = bE_t\pi_{t+1} + \frac{(1-\kappa)(1-b\kappa)(1+\zeta)}{\kappa} x_t, \tag{10}$$

$$x_t = E_t x_{t+1} - (R_t - E_t\pi_{t+1}), \tag{11}$$

where b is the subjective discount factor, κ the frequency of price adjustment and ζ the elasticity of labor supply. Note that, for simplicity, a standard assumption on prior independence is used; see e.g. Zellner (1971). Del Negro and Schorfheide (2008) criticize this assumption as having the drawback that the resulting joint prior distribution may assign a non-negligible amount of probability mass to regions of the parameter space where the model is unreasonable. It is fairly easy to see that the undesirable property suggested by Del Negro and Schorfheide (2008) is not present in our joint prior.

Table 1 lists the marginal prior distributions of the parameters. The beta prior distributions of the parameters α and β are concentrated towards unity, but are nonetheless only weakly informative (see Equations 10 and 11 for motivation). The prior mean of the slope coefficient β_r is set at unity, while the prior mean of γ (1.00) can be obtained by setting the subjective discount factor, the elasticity of labor supply and the frequency of price adjustment at their standard calibrated values, e.g. 0.99, 2 and 0.57, respectively, in Equation (10). The prior variances of these parameters (γ, β_r) are set to be small enough to keep the posterior distribution away from economically non-meaningful values. The prior means of the policy parameters γ_π (1.50) and γ_x (0.50) are obtained from Taylor (1993).[3] However, some interest rate smoothing is also allowed a priori. That is, the prior mean of ρ is set at 0.50. With the given prior variances, the marginal prior distributions of these parameters (γ_π, γ_x, ρ) turned out to be practically noninformative.

[3] In Taylor (1993), the interest rate and the inflation rate are expressed on a yearly basis. Since we express them on a quarterly basis, the prior mean of γ_x should be set at 0.125 (0.5 divided by 4). However, the standard deviation of the measure of the output gap used in Taylor (1993) is markedly higher than the standard deviation of the measure of the output gap used in this paper. Thus, the prior mean of 0.5 can be seen to be justified in our case.

The standard deviations σ_π, σ_x, and σ_R are assumed to follow inverse-gamma distributions with shape and scale parameters yielding fairly loose priors. Finally, the normal prior distribution with zero mean and 0.75^2 variance is used for the transformed parameters

$$\phi_\pi = \frac{1}{2}\log\frac{1+\rho_\pi}{1-\rho_\pi},\ \phi_x = \frac{1}{2}\log\frac{1+\rho_x}{1-\rho_x} \text{ and } \phi_R = \frac{1}{2}\log\frac{1+\rho_R}{1-\rho_R}. \tag{12}$$

These marginal priors force the posterior distributions of the autoregressive parameters ρ_π, ρ_x and ρ_R to be located in the interval [–1, 1]. The marginal priors are also very loose, but nevertheless turned out to improve simulation efficiency.

2.3 Data and Results

Throughout this study the quarterly U.S. data from 1953:2 to 2004:4 are used. In addition to the entire sample, the models are estimated for the subsample periods 1953:2-1982:2 and 1982:3-2004:4, capturing the "Great Inflation" and "Great Moderation" periods, respectively. This serves as a convenient check for robustness and parameter constancy. We are aware that the nominal interest rate, as the instrument of monetary policy, provides a reasonable description of the Federal Reserve's operating procedures only after 1964; see Clarida et al. (1999). However, the first ten years of data are required to have a sufficiently long out-of-sample forecasting period. We form out-of-sample forecasts from 1976:4 to 2004:4 to have forecast series which covers a diverse spectrum of inflation volatility.

The output gap is measured as a logarithmic difference between the actual and the potential output level. Two measures of actual output are used: real gross domestic product (GDP) and non-farm business (NFB) sector output. The logarithm of the potential output is proxied by the one-sided Hodrick-Prescott (HP) trend estimate in the model

$$g_t = \tau_t + \eta_{1t}, \tag{13}$$

$$(1 - L)^2 \tau_t = \eta_{2t}, \tag{14}$$

where g_t is the logarithm of the measure of actual output, L is a lag operator and η_{1t} and η_{2t} are mutually uncorrelated white noise sequences with the relative variance $q = \text{var}(\eta_{1t})/\text{var}(\eta_{2t})$. The value of $q = 0.67 \times 10^{-3}$ is taken from Stock and Watson (1999). We use the previous approximation of potential output, since our focus is on forecasting and since it does not use the future values of the detrended variable, as the optimal two-sided trend extraction HP-filter for Equations (13) and (14) does.[4] Furthermore, Stock and Watson (1999) find, after experimenting with several methods suitable for forecasting, that this procedure produces plausible gap estimates which work fairly well in inflation forecasting.

The price inflation is measured as the log difference of the Implicit Price Deflator of GDP (NFB). All the series are seasonally adjusted. The source of the final vintage data is the FREDII databank of the Federal Reserve Bank of St. Louis, while that of the real-time data is the Federal Reserve Bank of Philadelphia. The Federal Funds rate (FFR) is used as the instrument of monetary policy. The nominal interest rate and inflation rate series are measured as quarterly changes corresponding to their appearance in the structural model. Finally, the data are demeaned prior to estimation.

[4] We also tested for detrending linear and quadratic trend methods which are suitable for forecasting, and found that the results presented are not sensitive to using these measures of potential output. Furthermore, we ran several regressions with the dataset used in Lindé (2005). The results of the regressions with our and his datasets were quite similar.

The estimation results[5] are presented in Table 1, in the topmost panel (A) for the entire sample and in the lower panels (B and C) for the two subsample periods. The data appear to be particularly informative in all these samples. That is, the variances of the posterior distributions are found to be systematically smaller than the prior variances. The posteriors are also relatively stable between the data sets and the subsamples with two exceptions. The variances of the stochastic error processes seem to have fallen in the second subsample period. Sims and Zha (2006) and Smets and Wouters (2007) find similar evidence in U.S. data concerning the variance of monetary policy shocks. Our results also indicate that the Federal Reserve seemed to respond to the output gap and inflation more strongly during the second subsample period. The latter result is in accordance with that of Boivin and Giannoni (2006) and Smets and Wouters (2007), while Bernanke and Mihov (1998), Leeper and Zha (2003) and Canova (2006) find a relatively stable interest rate rule for the post WWII sample.

The Taylor principle is fulfilled in all the samples. This contradicts Clarida et al. (2000), who report that the Federal Reserve responded less than one-to-one to inflation during the period 1960-1979 (pre-Volcker period), thus violating the Taylor principle. In line with our result, for example, Smets and Wouters (2007) and Rabanal and Rubio-Ramírez (2005) find the inflation coefficient to be greater than one.

[5] To generate a Monte Carlo sample from the posterior of θ we used a version of the random walk Metropolis algorithm for Markov Chain Monte Carlo (MMCMC). The algorithm uses a multivariate normal distribution for the jump distribution on changes in θ. Our simulation procedure was as follows: we first simulated 10,000 draws using a diagonal covariance matrix with diagonal entries 0.00001 in the jump distribution. We then used these draws to estimate the posterior covariance matrix of θ and scale it by the factor $2.4^2/13$, to obtain an optimal covariance matrix for the jump distribution; see e.g. Gelman et al. (2004). We continue by a simulating 10,000 draws and calculated a more accurate covariance matrix for θ. We repeated this roughly 2 times. We then added noise to the posterior median to obtain overdispersed starting values and simulated three chains of length 30,000. We excluded the first 5000 simulations as a burn-in period in each chain and picked out every 25th draw from the Markov chain, yielding a sample of 3000 draws, which economizes on storage space and reduces autocorrelation across draws. The convergence of the chains was checked using Gelman and Rubin's convergence diagnostic R (also called 'potential scale reduction factor'); see Gelman and Rubin (1992). The diagnostic values close to 1 indicate approximate convergence and values smaller than 1.1 are acceptable in most cases. In our case the diagnostic was estimated to be between 1.01 and 1.03 for all parameters and all models. The multivariate version of Gelman and Rubin's diagnostic proposed by Brooks and Gelman (1998) was between 1.01 and 1.02 for each model; the convergence was thus fairly good. The frequencies of accepted jumps were roughly 0.21. Finally, the previous adaptive Metropolis algorithm is used because the covariance matrix estimate based on the local behaviour of the posterior at its highest peak turned out to give too optimistic a view of precision, and thus failed to yield an efficient covariance matrix for the normal jump distribution.

The point estimate of α (0.08) indicates a very insignificant role for the forward-looking behavior in the Phillips curve. This result is in accordance with those of Fuhrer (1997), Lindé (2005) and Rudd and Whelan (2006), but at odds with the results of Smets and Wouters (2003, 2005 and 2007), Adolfson et al. (2005) and Galí et al. (2005). The latter group of authors obtain relatively low parameter estimates for the degree of price indexation. Our estimates were obtained using a statistical measure of the output gap. Galí and Gertler (1999) and Galí et al. (2005) have suggested that the key reason for the lack of success of the forward-looking NKPC is that the detrended output is not a good proxy for real marginal costs. Contrary to their finding, Rudd and Whelan (2006), who used both the output gap and labor's share as a proxy for real marginal cost, found that the evidence for the forward-looking behavior in the NKPC was very weak.

The point estimates of β are high, supporting the traditional forward-looking intertemporal Euler equation. In contrast, previous studies have typically observed a high degree of habit persistence; see e.g. Christiano et al. (2005), and Smets and Wouters (2007). However, there seems to be a trade-off between the forward-looking behavior of the demand equation and the persistence of autoregressive demand shocks. In our paper, the high autoregressive parameter of the exogenous shock process (ρ_x = 0.79) takes into account the degree of persistence observed in the data. In Smets and Wouters (2007), the habit formation of consumption takes into account the high persistence, while the autoregressive parameter of exogenous shocks is estimated to be relatively small (0.36). Smets and Wouters (2007), however, assume a high habit parameter 0.7 (with 0.01 prior variance), a priori. Finally, the persistence of monetary policy shocks (ρ_R) is relatively low and equal to that estimated by Smets and Wouters (2007).

3. Forecast Comparison

In this section we first discuss the forecasting methods. We then provide some details for the forecasting comparison methods. Finally, we report the results of a forecasting exercise.

3.1 Measuring the Prediction Performance of Competitive Models

It is fairly easy to see that Equation (6) can be treated as a reduced-form VAR with lag-length 2 and normally distributed errors with covariance matrix $\Sigma = C_\varepsilon \Lambda C_\varepsilon'$. Thus, the conditional predictive distribution of Equation (6) for the joint lead time 1 through H, $p(y_{t+1},\ldots,y_{t+H}|Y,\theta)$, is multivariate normal; see Lütkepohl (1993). This facilitates straightforward simulations from $p(y_{t+1},\ldots,y_{t+H}|Y,\theta)$, given the posterior p.d.f. of θ. The method for obtaining the posterior p.d.f. of θ was explained in the previous section.[6]

The predictive performance of the hybrid NK model is compared to two Bayesian VARs and to naïve forecasts based on univariate random walks. The VAR systems consist of the same three variables, $y_t = (\pi_t, x_t, R_t)'$, as the hybrid NK model. The data are not however demeaned prior to estimation. Diffuse and Normal-Diffuse priors are used for the parameters of the VAR models; see Kadiyala and Karlsson (1997) for discussion. Parameterization of the Normal-Diffuse prior is based on the assumption that the variables behave as if they had random walk components; see Litterman (1980). That is, the prior means are set at zero except for the elements corresponding to the first own lag of each variable. The prior variances of the parameters in the ith equation of a p-lag VAR ($k = 1,\ldots, p$)[7] are given by π_1/k, $\pi_2 s_i^2/s_j^2 k$ ($i \neq j$) and $\pi_3 s_i^2$, for the parameters on own lags, foreign lags and a constant, respectively; see Litterman (1986) and Kadiyala and Karlsson (1997) for the motivation of this prior variance specification. A scale factor accounting for the different scales of the variables, s_i^2, is set at the residual standard error of equation i. The relative tightness of the prior is set at the commonly used values of hyper-parameters, $\pi_1 = 0.05$ and $\pi_2 = 0.005$; see e.g. Kadiyala and Karlsson (1997) and

[6] In a rolling forecast exercise, a total of 113 chains were simulated from each model. The posterior estimates of θ are based on 30,000 draws. The first 6,000 draws were discarded as a burn-in period. To reduce the size of output files, every 12th draw was saved. The predictive likelihoods are thus computed on the basis of 2000 draws from the Markov chain. Geweke (1992) proposed a convergence diagnostic for Markov chains based on a test for the equality of means of the first and last parts of the chain (in this paper the first 10% and the last 50% of observations were used). The test statistic is a standard Z-score; the difference between the two sample means divided by its estimated standard error. The standard error is estimated from the spectral density at zero and so takes into account any autocorrelation. The hypothesis of the equality of means was not rejected for most parameters at the 5 % significance level.

[7] In our paper, p is set at 4. The fractional marginal likelihoods (FML) of Villani (2001), which were used in preliminary data analysis, supported this choice in over 99% of the estimated regressions.

Litterman (1986). The tightness of the constant terms is set at $\pi_3 = 0.05$, which shrinks the processes towards driftless univariate random walk. This prior specification provides a suitable description for the processes of inflation, nominal interest rate and detrended output. The posterior distributions were simulated using the Gibbs sampling algorithm[8] of Kadiyala and Karlsson (1997) for the Normal-Diffuse prior specification and the matricvariate Student's *t* distribution for the Diffuse prior specification. The predictive likelihoods were computed on the basis of 2,000 draws from the posteriors.

The forecasting performance of the models is examined using the standard rolling forecast procedure, which entails making forecasts using data dated before the forecasting period. The forecasting procedure is as follows: using data up to a given time point T all the parameters in the model are estimated and the predictive distribution over $y_{T+1},\ldots,y_{T+H}$ is computed.[9] Moving forward one period, all the parameters are re-estimated and the forecast distribution of $y_{T+2},\ldots,y_{T+H+1}$ is computed. This is continued until no more data are available to compute the one-step-ahead forecast errors. The period over which the dynamic forecast distributions are computed in this manner is 1976:4 through 2004:4. In addition to the entire forecasts sample, the forecasts are also compared for the subsample period 1990:1-2004:4 (the sample period of Smets and Wouters, 2007). This serves as a check of robustness of the results and increases the comparability of our results to those in previous literature; especially in the paper of Smets and Wouters (2007).

Adolfson et al. (2007a) recommend use of several univariate and multivariate measures to determine the accuracy of the point forecasts. The two commonly used univariate measures of accuracy, the root mean squared forecast error (RMSE) and the mean absolute forecast error (MAE) are computed as

[8] 2,200 draws were simulated and the first 200 draws from the Markov chain were neglected as a burning period.
[9] Note also that when the forecasts are evaluated the data is demeaned and the gap estimates are computed using the data up to time *T*. Furthermore, when the analysis is based on demeaned data, the posterior median forecasts are computed and the means are added to the median forecasts.

$$RMSE_i(h) = \sqrt{N_h^{-1} \sum_{t=T}^{T+N_h-1} e_{i,t}^2(h)}, \tag{15}$$

$$MAE_i(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} \left| e_{i,t}(h) \right|, \tag{16}$$

respectively, where $e_{i,t}(h) = y_{i,t+h} - \hat{y}_{i,t+h|t}$ is the ith element of the h-step-ahead forecast error, $\hat{y}_{t+h|t}$ the h-step-ahead posterior median forecast of y_{t+h} and N_h the number of the h-step-ahead forecasts ($h = 1,\ldots,H$). However, only the RMSEs are reported, since these two measures turned out to produce equal results.

Two multivariate accuracy measures of point forecast, the log determinant statistic and the trace statistic, are also used in addition to the univariate measures. The multivariate statistics are based on the scaled h-step-ahead mean squared error (MSE) matrix

$$T_M(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} \bar{e}_t(h) \cdot \bar{e}_t'(h) \,, \tag{17}$$

where $\bar{e}_t(h) = M^{-1} e_t(h)$ and M is a scaling matrix accounting for the different scales of the variables being forecasted.[10] As discussed in Adolfson et al. (2007a), the forecasting performance of the least predictable dimensions, that is, those corresponding to the highest eigenvalues of the square matrix $T_M(h)$, mainly determine the trace statistic $\text{tr}[T_M(h)] = \lambda_1 + \ldots + \lambda_m$, while the log determinant statistic $\log |T_M(h)| = \log \lambda_1 + \ldots + \log \lambda_m$ also takes into account the forecasting performance of the most predictable dimensions (the lowest eigenvalues). It is also obvious that when the lowest eigenvalue of $T_M(h)$ approaches zero, the most predictable dimension determines the log determinant statistic.

[10] We follow Adolfson et al. (2007a) and set M equal to the diagonal of the sample covariance matrix of the y_t from 1976:4 to 2004:4 (1990:1 to 2004:4).

Finally, in view of the increasing interest for forecast uncertainty, we also compare the prediction performance of the competitive models using the log predictive density score (LPDS), which is a measure of the accuracy of multivariate density forecasts; see Adolfson et al. (2007a). To be more concrete, let $\hat{y}_{t+h|t}$ and $\Omega_{t+h|t}$ denote the posterior mean and covariance matrix of the h-step-ahead forecast distribution $p_t(y_{t+h})$. Then, under the normality assumption of $p_t(y_{t+h})$, the LPDS of the h-step-ahead predictive density at time t is defined as

$$\begin{aligned} S_t(y_{t+h}) &= -2\log p_t(y_{t+h}) \\ &= m\log(2\pi) + \log\left|\Omega_{t+h|t}\right| + \left(y_{t+h} - \hat{y}_{t+h|t}\right)'\Omega_{t+h|t}^{-1}\left(y_{t+h} - \hat{y}_{t+h|t}\right). \end{aligned} \tag{18}$$

We report the averages of the LPDSs over the evaluated h-step-ahead forecasts,

$$S(h) = N_h^{-1}\sum_{t=T}^{T+N_h-1} S_t(y_{t+h}). \tag{19}$$

This measure takes into account the forecasting performance of the predictive density as a whole.

3.2 Results

Figures 1-3 summarize the forecasting performance of the competitive models. Specifically, Figure 1 reports the RMSEs in quarterly percentage terms, Figure 2 the log determinant and the trace statistics, and Figure 3 the averages of the LPDS statistic. Figures 4-6 gives the corresponding statistics for the forecasts based on real-time data. The results based on the NFB data were similar to those based on the GDP data and in order to save space we report only the latter. All the statistics are reported at the 1 to 12 quarters horizons. [11] In the figures, a small value favors the model.

[11] We do not report the marginal likelihood, since it captures only the one-step-ahead predictive performance of the full model and is therefore too restricted for forecasting comparison.

A few key findings emerge from the figures. Firstly, although the models are very simple they seem to forecast particularly well. According to the RMSEs, the small-scale models appear to produce more accurate point forecasts, on both inflation and the Federal Funds rate,[12] than the large-scale Bayesian VAR of Smets and Wouters (2007). In addition, the models turned out to produce real-time inflation forecasts which outperformed the naïve forecasts up to six quarters in the 1990:1-2004:4 subsample (see Figure 4). This result gives some perspective to the forecast accuracy of the hybrid model, when we take into account the finding of Atkeson and Ohanian (2001) that the one-year-ahead Federal Reserve's Greenbook inflation forecast has not been better on average than the naïve forecast since 1984.

Secondly, all the forecast comparison methods appear to yield similar conclusions. In the entire sample the forecasts of the hybrid model outperform those of the Bayesian VARs, while in the low inflation subsample (1990:1-2004:4) all the multivariate forecasting methods seem to produce equally accurate forecasts. Thus, the restrictions (stationary and cross-equation) implied by the hybrid model appear to help in forecasting especially well during high inflation periods. According to the univariate and multivariate measures of forecast accuracy, this result is most obvious at medium-term horizons. One exception is the nominal interest rate. The hybrid model forecasts this series very well in all samples and forecasting horizons (see Figure 1 and 4). In particular, all these results hold for both ex post data and real-time data.

Taking a closer look at the figures we see that the hybrid model is superior to the naïve forecasts at all samples and horizons, except for the longer horizon inflation forecasts in the low inflation subsample. In this subsample, the Bayesian VARs also give slightly better inflation and output gap forecasts than the hybrid model, according to the RMSEs. However, the improvement in the predictability of the variables is clearly negligible.

[12] The GDP forecasts are not directly comparable to results of Smets Wouters (2007), since they use the log difference of GPD series, while we use the GDP gap series.

It also seems that the shrinking prior does not improve the forecasting performance of VARs in terms of point forecasting accuracy. This is not surprising, since the VAR systems are particularly parsimonious and, hence, do not suffer from the over-parameterization problem. However, the LPDS statistics (see Figures 3 and 6) support a slightly better forecasting density for the Normal-Diffuse prior specification in the low inflation subsample. Over the entire sample the LPDSs support Bayesian VARs at the shorter forecasting horizons (1 to 4 quarter); however, the hybrid model again outperforms the VARs at the longer horizons.

In sum, it seems fair to say that the simple hybrid NK model captures the predictable behavior of the three U.S. key macroeconomic variables very well. The reason for its good forecasting performance may be that the model allows both for the endogenous persistence in inflation and output and for the persistence of the exogenous shock processes. This approach is commonly used in large-scale NK models, which forecast well. Our joint prior is also well designed in allowing the parameters to be estimated fairly freely, while being informative enough to keep the posterior distribution away from economically non-meaningful values.

4. Conclusion

Several recent papers have suggested different ways to improve the forecast performance of New Keynesian models. Unfortunately, improvement in fit is achieved at the cost of increasing the complexity of model mechanisms, which reduces the practicability of these approaches. This paper, in contrast, has shown that the very simple hybrid New Keynesian model of Clarida et al. (1999) can provide forecasts comparable to those based on commonly used benchmark models such as reduced-form Bayesian VARs and univariate random walks.

Our forecasting evidence indicates that the restrictions implied by the hybrid model work especially well in high inflation regimes. According to several univariate and multivariate measures of forecast accuracy, the forecasts of the hybrid model outperform those of the

Bayesian VARs when high inflation periods are forecast. In the low inflation forecast subsample, the methods produce equally accurate forecasts. One exception was the nominal interest rate. The hybrid model seems to forecast this series very well in all samples and horizons. The hybrid model also predicts more accurately than the naïve forecasts based on univariate random walks. Finally, we note that all these findings hold for both ex post and real-time data.

Table 1. Priors and Posteriors

	Prior Distr.			Posterior Distr. (BKT)			Posterior Dist. (NFB)		
Panel A: Sample 1954:2 – 2004:4									
Par.	Distr.	Mean	St.Dev.	Median	5%	95%	Median	5%	95%
α	Beta	0.67	0.24	0.08	0.02	0.19	0.08	0.02	0.18
γ	Gamma	1.00	0.32	0.03	0.02	0.05	0.03	0.02	0.05
β	Beta	0.67	0.24	0.75	0.65	0.84	0.74	0.65	0.83
β_r	Gamma	1.00	0.32	0.10	0.05	0.16	0.12	0.07	0.20
γ_π	Gamma	1.5	0.61	1.82	1.50	2.32	1.82	1.46	2.34
γ_x	Gamma	0.5	0.35	0.59	0.40	0.87	0.49	0.33	0.76
ρ	Beta	0.5	0.22	0.87	0.83	0.91	0.89	0.85	0.92
ρ_π	Normal	0	0.54	-0.38	-0.46	-0.28	-0.42	-0.50	-0.33
ρ_x	Normal	0	0.54	0.79	0.67	0.87	0.78	0.66	0.86
ρ_R	Normal	0	0.54	0.12	-0.00	0.24	0.11	-0.00	0.24
σ_π	Invgam.	0.40	3.96	0.29	0.26	0.32	0.34	0.31	0.37
σ_x	Invgam.	0.40	3.96	0.16	0.12	0.20	0.21	0.16	0.28
σ_R	Invgam.	0.40	3.96	0.22	0.20	0.24	0.22	0.20	0.24
Panel B: Sample 1954:2 – 1982:2									
α	Beta	0.67	0.24	0.08	0.02	0.20	0.08	0.02	0.21
γ	Gamma	1.00	0.32	0.05	0.03	0.07	0.05	0.03	0.07
β	Beta	0.67	0.24	0.79	0.66	0.94	0.77	0.66	0.92
β_r	Gamma	1.00	0.32	0.19	0.11	0.32	0.21	0.12	0.35
γ_π	Gamma	1.5	0.61	1.86	1.46	2.46	1.81	1.41	2.47
γ_x	Gamma	0.5	0.35	0.52	0.29	0.86	0.47	0.25	0.77
ρ	Beta	0.5	0.22	0.84	0.78	0.90	0.87	0.80	0.92
ρ_π	Normal	0	0.54	-0.35	-0.48	-0.21	-0.41	-0.52	-0.28
ρ_x	Normal	0	0.54	0.77	0.60	0.87	0.76	0.60	0.87
ρ_R	Normal	0	0.54	0.11	-0.06	0.29	0.10	-0.06	0.27
σ_π	Invgam.	0.40	3.96	0.34	0.30	0.39	0.41	0.36	0.46
σ_x	Invgam.	0.40	3.96	0.24	0.17	0.34	0.32	0.22	0.43
σ_R	Invgam.	0.40	3.96	0.28	0.25	0.31	0.28	0.25	0.32
Panel C: Sample 1982:3 – 2004:4									
α	Beta	0.67	0.24	0.08	0.02	0.20	0.08	0.02	0.20
γ	Gamma	1.00	0.32	0.05	0.03	0.08	0.04	0.03	0.06
β	Beta	0.67	0.24	0.83	0.70	0.97	0.86	0.73	0.98
β_r	Gamma	1.00	0.32	0.19	0.11	0.32	0.23	0.13	0.37
γ_π	Gamma	1.5	0.61	2.65	1.93	3.64	2.63	1.88	3.75
γ_x	Gamma	0.5	0.35	0.89	0.57	1.35	0.69	0.42	1.08
ρ	Beta	0.5	0.22	0.90	0.86	0.93	0.91	0.87	0.94
ρ_π	Normal	0	0.54	-0.35	-0.50	-0.17	-0.37	-0.50	-0.22
ρ_x	Normal	0	0.54	0.88	0.79	0.94	0.88	0.80	0.94
ρ_R	Normal	0	0.54	0.29	0.11	0.47	0.34	0.16	0.52
σ_π	Invgam.	0.40	3.96	0.22	0.19	0.25	0.24	0.20	0.28
σ_x	Invgam.	0.40	3.96	0.09	0.06	0.12	0.11	0.08	0.16
σ_R	Invgam.	0.40	3.96	0.12	0.10	0.13	0.11	0.10	0.13

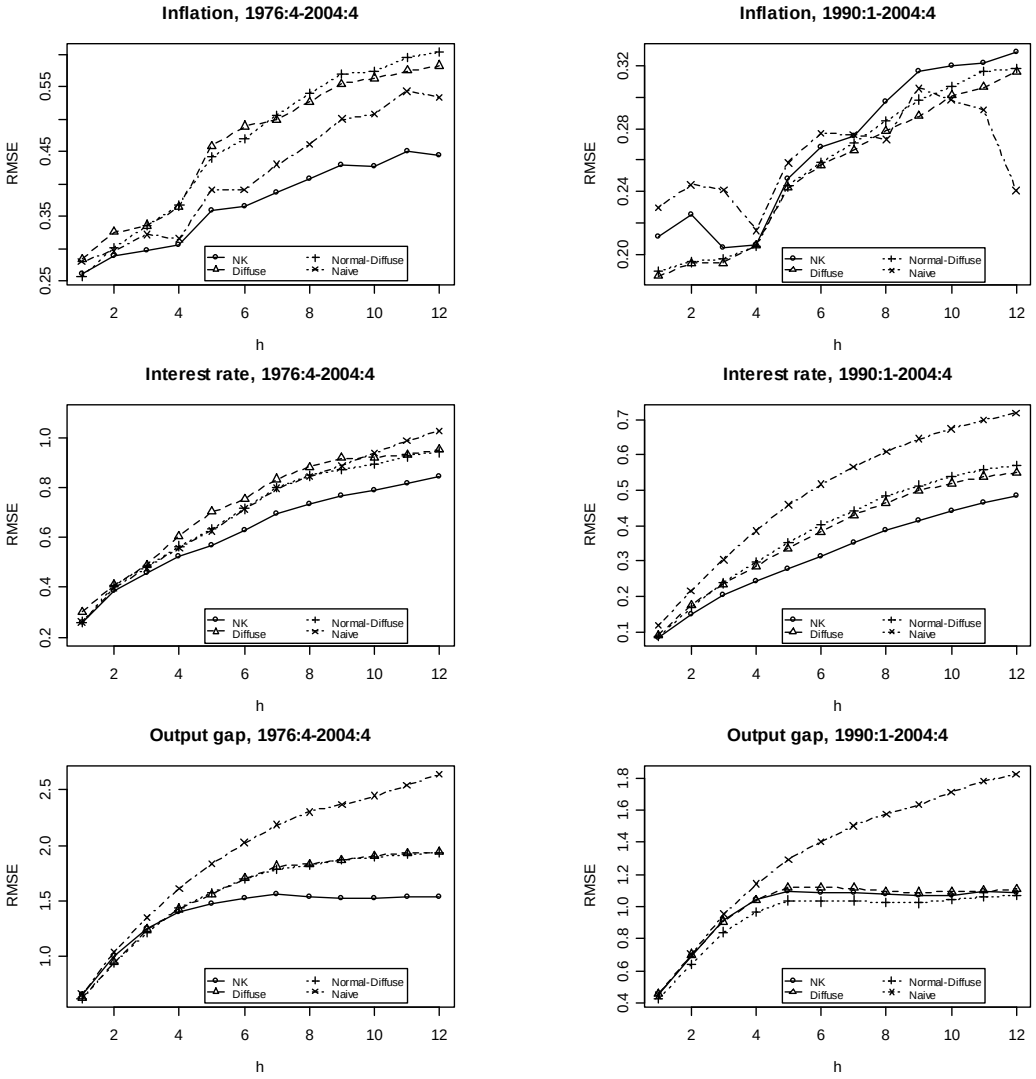

Figure 1. The Root mean squared forecast errors for the competitive models

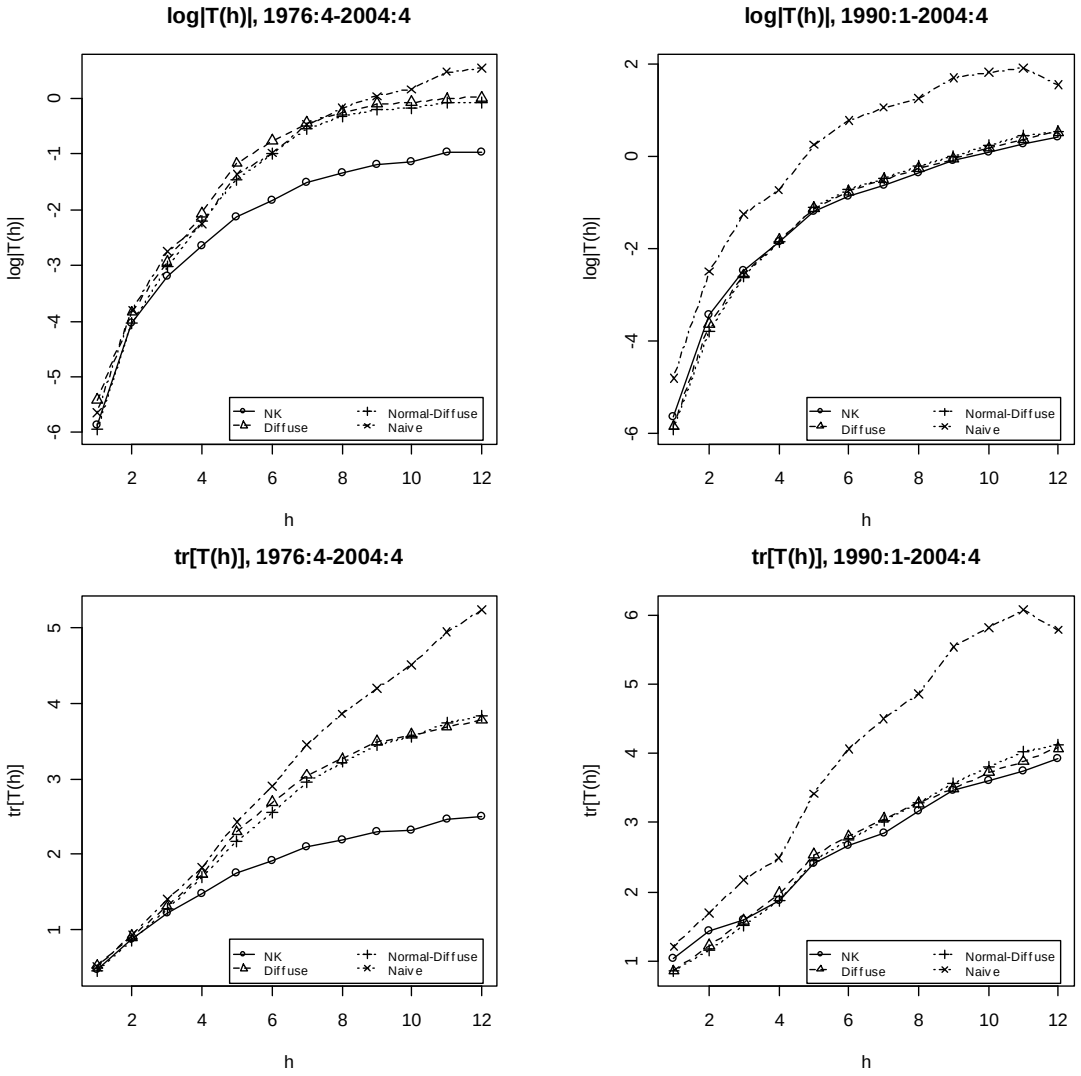

Figure 2. The log determinant statistics and the trace statistics for the competitive models

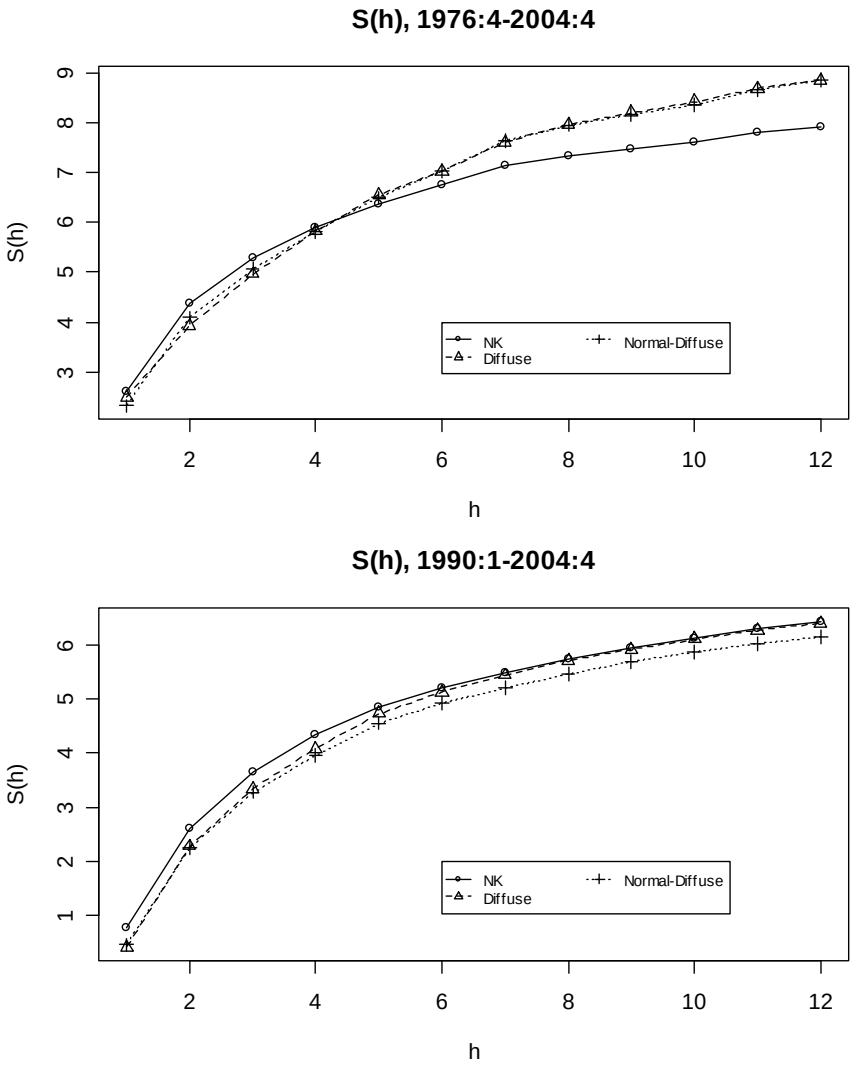

Figure 3. The average log predictive density scores for the competitive models

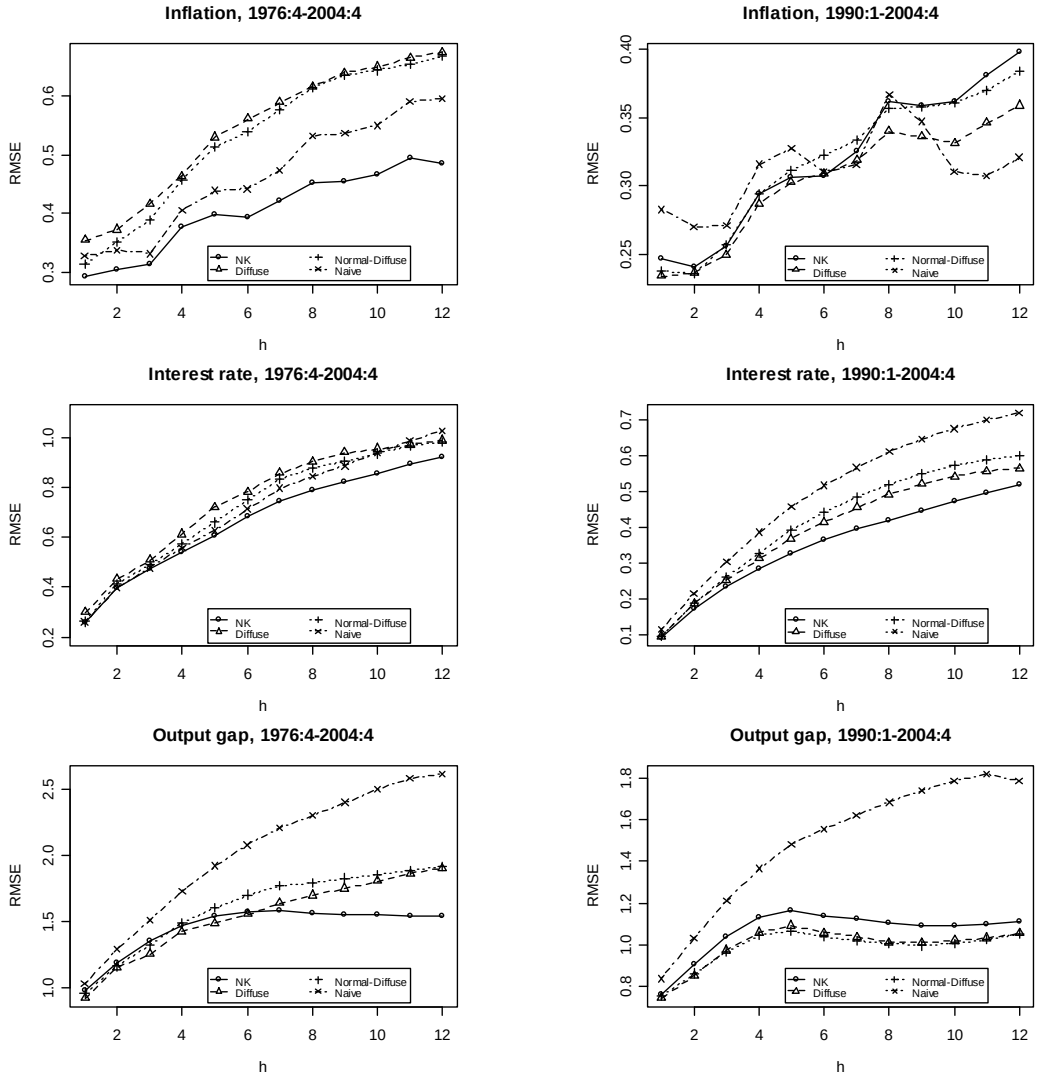

Figure 4. The Root mean squared forecast errors for the competitive models (real-time data)

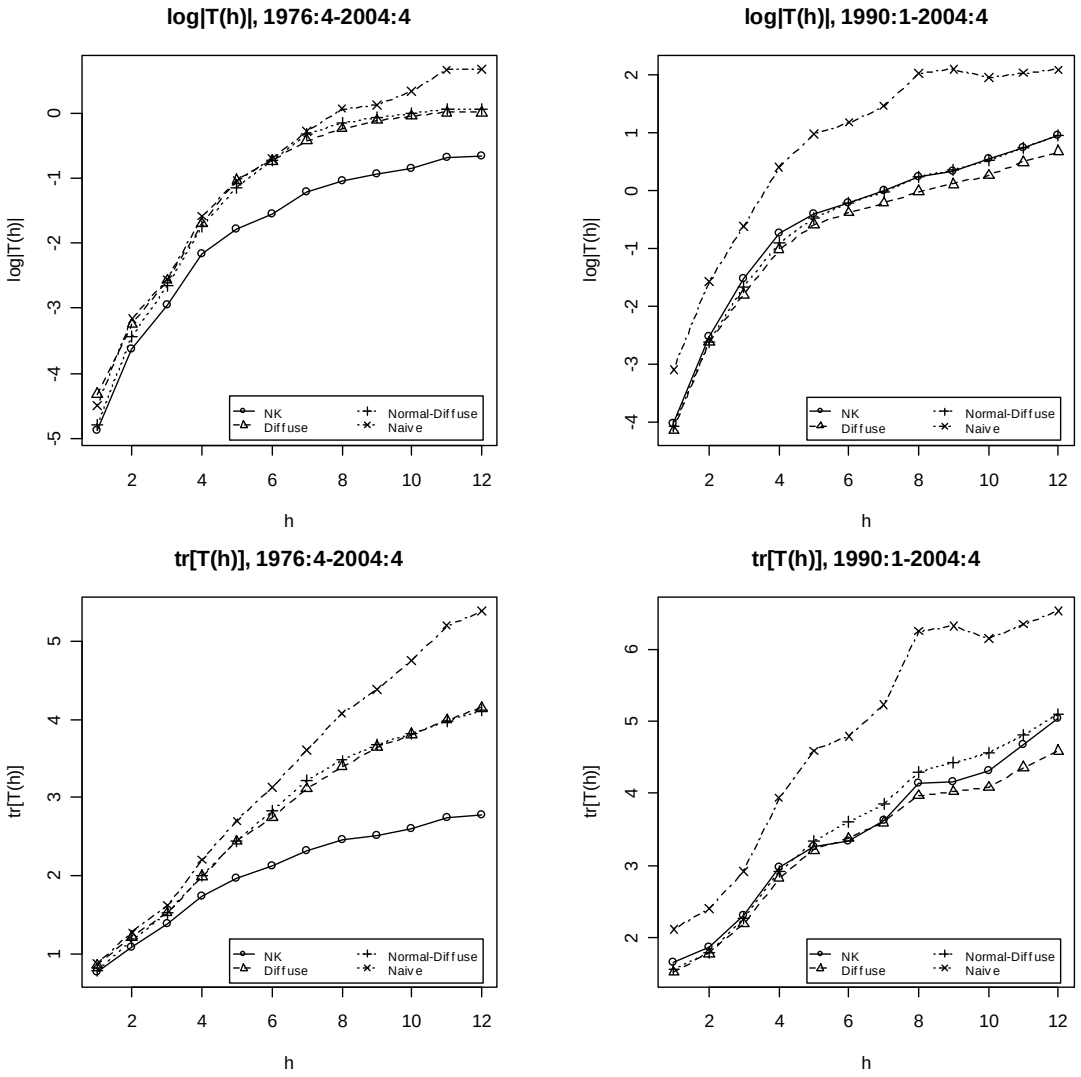

Figure 5. The log determinant statistics and the trace statistics for the competitive models l (real-time data)

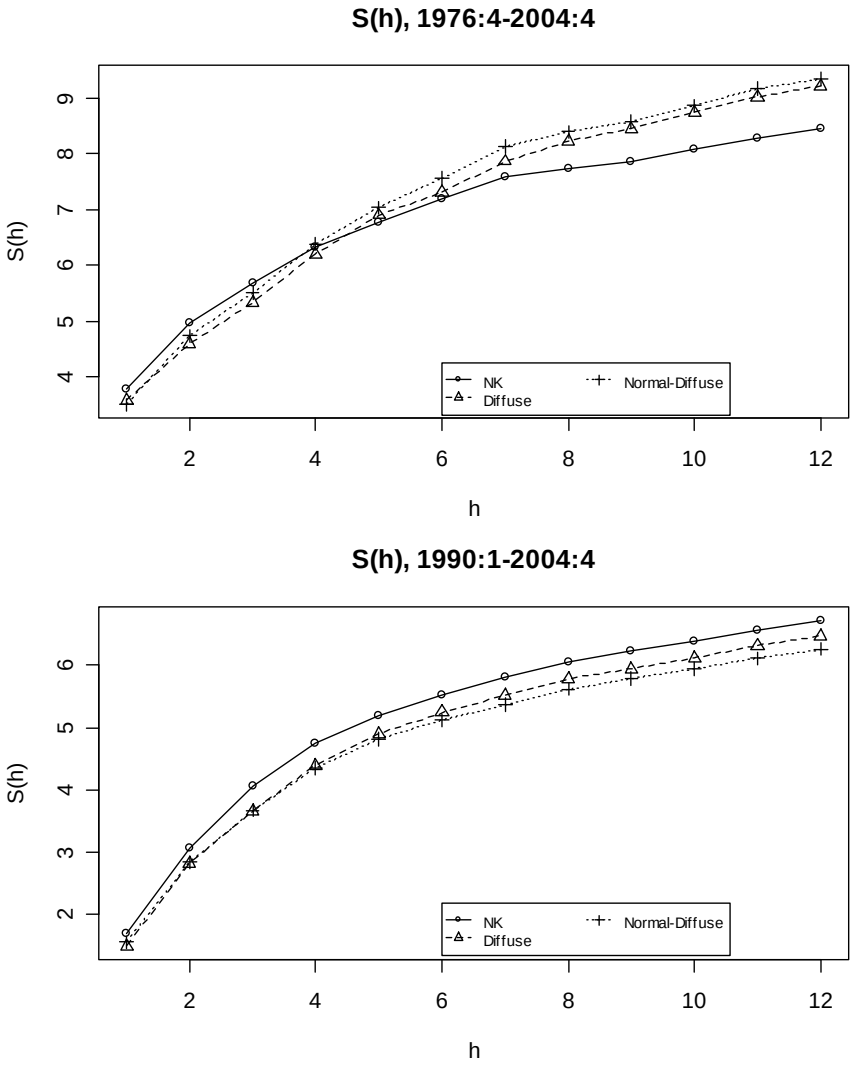

Figure 6. The average log predictive density scores for the competitive models (real-time data)

References:

Adolfson, Malin, Stefan Laséen, Jesper Lindé, and Mattias Villani, "Evaluating an Estimated New Keynesian Small Open Economy Model," *Journal of Economic Dynamics and Control* forthcoming (2008).

Adolfson, Malin, Jesper Lindé, and Mattias Villani, "Forecasting Performance of an Open Economy DSGE Model," *Econometric Reviews* 26:2-4 (2007a), 289–328.

Adolfson, Malin, Stefan Laséen, Jesper Lindé, and Mattias Villani, "The Role of Sticky Prices in an Open Economy DSGE Model: A Bayesian Investigation," *Journal of the European Economic Association* 3:2-3 (2005), 444–457.

Adolfson, Malin, Jesper Lindé, and Mattias Villani, "Bayesian Inference in DSGE Models - Some Comments," *Econometric Reviews* 26:2-4 (2007b), 173–185.

Atkeson, Andrew, and Lee, E. Ohanian, "Are Phillips Curves Useful for Forecasting Inflation?," *Federal Reserve Bank of Minneapolis Quarterly Review* 25:1 (2001), 2–11.

Bernanke, Ben, and Ilian Mihov, "Measuring Monetary Policy," *The Quarterly Journal of Economics* 113:3 (1998), 869–902.

Boivin, Jean, and Marc P. Giannoni, "Has Monetary Policy Become More Effective?," *The Review of Economics and Statistics* 88:3 (2006), 445–462.

Brooks, Stephen P., and Andrew Gelman, "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics* 7:4 (1998), 434–455.

Canova, Fabio, "Monetary Policy and the Evolution of the US Economy," CEPR Discussion Papers No. 5467 (January 2006).

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy* 113:1 (2005), 1–46.

Clarida, Richard, Jordi Galí, and Mark Gertler, "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature* 37 (December 1999), 1661–1707.

Clarida, Richard, Jordi Galí, and Mark Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics* 115:1 (2000), 147–180.

Del Negro, Marco, and Frank Schorfheide, "Priors from General Equilibrium Models for VARs," *International Economic Review* 45:2 (2004), 643–673.

Del Negro, Marco, Frank Schorfheide, Frank Smets, and Ralf Wouters, "On the fit of New-Keynesian Models," *Journal of Business & Economic Statistics* 25:2 (2007), 143–162.

Del Negro, Marco, and Frank Schorfheide, "Forming Priors for DSGE Models," Manuscript (2008).

Fuhrer, Jeffrey C., "The (Un)Importance of Forward-looking Behavior in Price Specifications," *Journal of Money, Credit and Banking* 29:3 (1997), 338–350.

Galí, Jordi, and Mark Gertler, "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics* 44:2 (1999), 195–222.

Galí, Jordi, Mark Gertler, and J. David López-Salido, "Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve," *Journal of Monetary Economics* 52:6 (2005), 1107–1118.

Gelman, Andrew, and Donald B. Rubin, "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science* 7:4 (1992), 457–472.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, *Bayesian Data Analysis 2nd edition* (Boca Raton: Chapman & Hall/CRC, 2004).

Geweke, John, "Evaluating the Accuracy of Sampling-based Approaches to Calculating Posterior Moments" (pp. 169–193), in José M. Bernardo, James O. Berger, A. Philip David and Adrian F.M. Smith (Eds.), *Bayesian Statistics*, Vol. 4 (Oxford Oxford University Press, UK, 1992).

Kadiyala, K. Rao and Sune Karlsson, "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," *Journal of Applied Econometrics* 12:2 (1997), 99–132.

Leeper, Eric, and Tao Zha, "Modest Policy Interventions," *Journal of Monetary Economics* 50:8 (2003), 1673–1700.

Lindé, Jesper, "Estimating New-Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach," *Journal of Monetary Economics* 52:6 (2005), 1135–1149.

Litterman, Robert B., "A Bayesian Procedure for Forecasting with Vector Autoregressions," Mimeo, Massachusetts Institute of Technology (1980).

Litterman, Robert B., "Forecasting with Bayesian vector autoregressions - Five years of experience," *Journal of Business and Economic Statistics* 4:1 (1986), 25–38.

Lütkepohl, Helmut, *Introduction to Multiple Time Series Analysis, 2nd edition* (New York: Springer-Verlag, 1993).

Rabanal, Pau, and Juan F. Rubio-Ramírez, "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics* 52:6 (2005), 1151–1166.

Rudd, Jeremy, and Karl Whelan, "Can Rational Expectations Sticky-Price Models Explain Inflation Dynamics?," *American Economic Review* 96:1 (2006), 303–320.

Sims, Christopher, and Tao Zha, "Were There Regime Switches in U.S. Monetary Policy?," *American Economic Review* 96:1 (2006), 54–81.

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

Smets, Frank, and Ralf Wouters, "Comparing Shocks and Frictions in US and Euro Area Business Cycles: a Bayesian DSGE Approach," *Journal of Applied Econometrics* 20:2 (2005), 161–183.

Smets, Frank, and Ralf Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review* 97:3 (2007), 586–606.

Stock, James, and Mark Watson, "Forecasting Inflation," *Journal of Monetary Economics* 44:2 (1999), 293–335.

Sungbae, An, and Frank Schorfheide, "Bayesian Analysis of DSGE Models," *Econometric Reviews* 26:2-4 (2007), 113–172.

Söderlind, Paul, "Solution and Estimation of RE Macromodels with Optimal Policy," *European Economic Review* 43:4-6 (1999), 813–823.

Taylor, John B., "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* 39 (December 1993), 195–214.

Villani, Mattias, "Fractional Bayesian Lag Length Inference in Multivariate Autoregressive Processes," *Journal of Time Series Analysis* 22:1 (2001), 67–86.

Zellner, Arnold, "*An Introduction to Bayesian Inference in Econometrics*" (New York John Wiley and Sons, Inc., 1971).