Professional Forecasters and Real-Time Forecasting with a DSGE Model

Frank Smets,[†] Anders Warne[‡] and Rafael Wouters[§]

January 21, 2014

Abstract: This paper analyses the real-time forecasting performance of the New Keynesian DSGE model of Galí, Smets, and Wouters (2012) estimated on euro area data. It investigates to what extent forecasts of inflation, GDP growth and unemployment by professional forecasters improve the forecasting performance. We consider two approaches for conditioning on such information. Under the "noise" approach, the mean professional forecasts are assumed to be noisy indicators of the rational expectations forecasts implied by the DSGE model. Under the "news" approach, it is assumed that the forecasts reveal the presence of expected future structural shocks in line with those estimated over the past. The forecasts of the DSGE model are compared with those from a Bayesian VAR model, an AR(1) model, a sample mean and a random walk.

Keywords: Bayesian methods, estimated New Keynesian model, real-time database, Survey of Professional Forecasters, Macroeconomic forecasting, euro area.

JEL Classification Numbers: E24, E31, E32.

1. Introduction

Following the seminal work of Croushore and Stark (2001) on constructing a real-time data set for the US economy, it has become standard to use real-time data when analysing the out-of-sample forecast performance of alternative empirical macromodels.[1] With a few exceptions much less real-time data analysis has been done on the euro area, partly because a comprehensive real-time euro area data set has only recently become available.[2] This paper uses the European Central Bank (ECB) real-time data base (RTDB)—described in Giannone, Henry, Lalik, and Modugno (2012) and available on the ECB's website—to perform two types of analysis.

In this paper we investigate the forecasting performance of the Galí, Smets, and Wouters (2012, GSW) model in real time over the EMU period and compare it with four alternative non-structural linear models. The GSW model is a version of the model by Smets and Wouters (2003, 2007) which has been shown to forecast reasonably well. It is therefore of interest to see to what extent these results are robust to the real-time nature of the underlying data in the euro area. Recently, a similar exercise on US data was performed by Edge and Gürkaynak (2010).

[†] European Central Bank and KU Leuven.

[‡] European Central Bank.

[§] National Bank of Belgium.

[1] See, for example, Croushore (2011) and the literature review on real-time data analysis compiled by Dean Croushore at `https://facultystaff.richmond.edu/~dcrousho/docs/realtime_lit.pdf`. For an early real-time forecasting exercise, see Diebold and Rudebusch (1991).

[2] Two exceptions are Coenen, Levin, and Wieland (2005) and Coenen and Warne (2013).

Moreover, we analyse to what extent the forecasts of euro area GDP growth, inflation and unemployment by professional forecasters (from the ECB's Survey of Professional Forecasters, SPF) help improving the forecast performance of the DSGE model. We consider two interpretations. Under the "noise" interpretation, the mean professional forecasts are assumed to be noisy indicators of the rational expectations forecasts implied by the DSGE model. Under the "news" interpretation, it is assumed that the forecasts reveal the presence of expected future structural shocks in line with those estimated over the past. This exercise is similar to the one performed by Del Negro and Schorfheide (2013) for the United States.

Two sets of results are worth highlighting. First, we find support that the point forecasts of the benchmark GSW model can be improved in a mean squared sense via a BVAR, in particular for consumption and real wages, where the GSW model systematically overpredicts real wage growth and underpredicts consumption. These variables are also poorly predicted under the noise and news versions of the GSW model when compared to the BVAR model. At the same time, the inflation forecasts from the models using the SPF data are much improved relative to the BVAR, making the overall picture more difficult to assess, also when comparing the point forecasts of the DSGE models to common univariate non-structural models.

Second, when comparing the root mean squared errors from the point forecasts of the news version of the GSW model to the benchmark GSW model, also using formal tests, we find that inflation and real wage forecasts are improved over the one- to four-quarter-ahead horizon, while the one- and two-quarter-ahead point forecasts of real GDP growth are also ameliorated when using the 1- and 2-year-ahead SPF data as pure conditioning information. The main forecasting cost is the relative deterioration in the interest rate forecasts. For the noise version, however, it is mainly the inflation forecasts that are improved relative to the benchmark GSW model, while the interest rate forecasts are aggravated. Overall, we therefore find that the GSW model forecasts for the euro area can be improved by using the SPF data, provided that these data are included in line with the news interpretation.

The rest of the paper is structured as follows. Section 2 presents the GSW model. Section 3 presents the real-time data base including the Survey of Professional Forecasts. Section 4 discusses the full-sample estimation results of the benchmark GSW model and provides a brief comparison with the findings for the United States reported in Galí et al. (2012). Section 5 contains the findings of the real-time forecast comparison exercise. Finally, Section 6 summarises the main findings and concludes.

2. The Galí-Smets-Wouters Model

This section describes the log-linearized equilibrium conditions of the GSW model. It is a standard medium-sized DSGE model with sticky prices and wages that can explain the main macroeconomic time series, such as output and inflation, and is very similar to Smets and Wouters (2007, SW). One main difference is that it models the labor supply decision on the extensive margin (whether to work or not), rather than on the intensive margin (how many hours to work), which allows us to include unemployment as an observable variable.

The model includes eight exogenous shocks: a neutral, factor-augmenting productivity shock ($\widehat{\varepsilon}_t^a$), a labor supply shock ($\widehat{\varepsilon}_t^s$), a price markup shock ($\widehat{\varepsilon}_t^p$), a wage markup shock ($\widehat{\varepsilon}_t^w$), a risk premium shock ($\widehat{\varepsilon}_t^b$), an exogenous spending shock ($\widehat{\varepsilon}_t^g$), an investment-specific technology shock ($\widehat{\varepsilon}_t^q$), and a monetary policy shock ($\widehat{\varepsilon}_t^r$). In addition, eight observable variables are used to estimate the model. Next, we describe the main structural equations with E_t denoting the rational expectations operator conditional on the information at time t and $\widehat{}$ denoting deviation of the variable from its steady state growth path.

- Consumption Euler equation. Consumption, $\widehat{c}_t$, depends on lagged consumption because of habit formation and expected future consumption, as well as the expected short-term real interest rate, $(\widehat{r}_t - E_t\widehat{\pi}_{t+1} - \widehat{\varepsilon}_t^b)$:

$$\widehat{c}_t = c_1 E_t\left[\widehat{c}_{t+1}\right] + \left(1 - c_1\right)\widehat{c}_{t-1} - c_2\left(\widehat{r}_t - E_t\widehat{\pi}_{t+1} - \widehat{\varepsilon}_t^b\right),$$

  where $\widehat{r}_t$ is the short-term nominal interest rate and $\widehat{\pi}_t$ is the inflation rate, with $c_1 = 1/(1 + (h/\tau))$, $c_2 = (1 - (h/\tau))/(1 + (h/\tau))$, where h is the external habit parameter, τ is the trend growth rate, and $\widehat{\varepsilon}_t^b$ is the exogenous AR(1) risk premium process.

- Investment Euler equation. Investment, $\widehat{i}_t$, also depends on past and expected future investment, as well as the value of capital, $\widehat{q}_t^k$:

$$\widehat{i}_t = i_1\widehat{i}_{t-1} + \left(1 - i_1\right)E_t\widehat{i}_{t+1} + i_2\widehat{q}_t^k + \widehat{\varepsilon}_t^q,$$

  with $i_1 = 1/(1 + \beta)$, $i_2 = i_1/(\tau^2\varphi)$ where β is the discount factor, φ is the elasticity of the capital adjustment cost function, and $\widehat{\varepsilon}_t^q$ is the exogenous AR(1) process for the investment-specific technology.

- Value of the capital stock. The value of the capital stock is determined by an arbitrage equation which equalizes the expected return on holding capital to the expected real interest rate:

$$\widehat{q}_t^k = -\left(\widehat{r}_t - E_t\widehat{\pi}_{t+1} - \widehat{\varepsilon}_t^b\right) + q_1 E_t\widehat{r}_{t+1}^k + \left(1 - q_1\right)E_t\widehat{q}_{t+1}^k,$$

  with $\widehat{r}_t^k$ the rental rate on capital, $q_1 = r^k/(r^k + (1 - \delta))$, where r^k is the steady-state rental rate on capital, and δ the depreciation rate.

- Goods market equilibrium. In equilibrium, aggregate demand—which consists of consumption, investment, the resources spent on adjusting capital utilization ($\widehat{v}_t$), and an exogenous demand component (spending shock)—has to equal aggregate supply. The latter is determined by a standard Cobb-Douglas production function in effective capital services, $\widehat{k}_t$, and hours worked, $\widehat{n}_t$:

$$\begin{aligned}\widehat{y}_t &= c_y\widehat{c}_t + i_y\widehat{i}_t + v_y\widehat{v}_t + \widehat{\varepsilon}_t^g,\\ &= \phi_p\left(\alpha\widehat{k}_t + \left(1 - \alpha\right)\widehat{n}_t + \widehat{\varepsilon}_t^a\right),\end{aligned}$$

where $\widehat{y}_t$ is output, $c_y = 1 - i_y - g_y$ is the steady-state consumption-output ratio, g_y the steady-state exogenous spending to output ratio, $i_y = (\tau + \delta - 1)k_y$ is the steady-state investment-output ratio, k_y the steady-state capital-output ratio, and $v_y = r^k k_y$. The parameter ϕ_p reflects the fixed costs in production, which is assumed to correspond to the price markup in steady state, while $\widehat{\varepsilon}_t^g$ and $\widehat{\varepsilon}_t^a$ are the AR(1) processes representing exogenous demand components and the TFP process.[3]

- Price-setting under the Calvo model with indexation. Inflation is sticky and depends on past and expected future inflation, as well as on the difference between the average ($\widehat{\mu}_{p,t}$) and the natural ($\widehat{\mu}_{p,t}^n$) price markup:

$$\widehat{\pi}_t - \gamma_p \widehat{\pi}_{t-1} = \pi_1 \big(E_t \widehat{\pi}_{t+1} - \gamma_p \widehat{\pi}_t \big) - \pi_2 \left(\widehat{\mu}_{p,t} - \widehat{\mu}_{p,t}^n \right),$$

with $\pi_1 = \beta$, $\pi_2 = (1 - \theta_p \beta)(1 - \theta_p)/[\theta_p(1 + (\phi_p - 1)\varepsilon_p)]$, with θ_p and γ_p respectively the probability of price changes and price indexation of the Calvo model, and ε_p the curvature of the aggregator function. The average price markup is equal to the inverse of the real marginal cost $\widehat{mc}_t = (1 - \alpha)(\widehat{w}_t - \widehat{p}_t) + \alpha \widehat{r}_t^k + \widehat{\varepsilon}_t^a$, which is determined by the real wage ($\widehat{w}_t - \widehat{p}_t$) and the rental rate on capital. The natural price markup is equal to $100\widehat{\varepsilon}_t^p$, i.e. it is proportional to the price markup shocks. These shocks are assumed to follow an exogenous ARMA(1,1) process.

- Wage setting under the Calvo model with indexation. Wage inflation depends on expected future wage inflation, due to partial wage indexation it also depends on past and current inflation, while nominal wage stickiness implies that it reacts gradually to the difference between the average ($\widehat{\mu}_{w,t}$) and the natural ($\widehat{\mu}_{w,t}^n$) wage markup:

$$\Delta \widehat{w}_t = \gamma_w \widehat{\pi}_{t-1} + \beta E_t \big(\Delta \widehat{w}_{t+1} - \gamma_w \widehat{\pi}_t \big) - w_1 \big(\widehat{\mu}_{w,t} - \widehat{\mu}_{w,t}^n \big),$$

with Δ being the first difference operator, $w_1 = (1 - \beta\theta_w)(1 - \theta_w)/\theta_w(1 + \epsilon_w \omega)$, θ_w and γ_w respectively the probability of wage changes and wage indexation of the Calvo model, ω the inverse elasticity of labor supply, and ϵ_w the curvature of the aggregator function.

- Average and natural wage markups and unemployment. The wage markup is defined as the difference between the real wage and the marginal rate of substitution, which is a function of the smoothed trend of consumption, $\widehat{z}_t$, employment, $\widehat{e}_t$, and the labor supply

[3] The innovation of the TFP process enters the process describing exogenous spending with the parameter ρ_{ga}; see Table 2 in Section 4.

shock:

$$\begin{aligned}
\widehat{\mu}_{w,t} &= \widehat{w}_t - \widehat{p}_t + \widehat{z}_t + \widehat{\varepsilon}_t^s + \omega\widehat{e}_t, \\
&= \omega\widehat{u}_t. \\
\widehat{\mu}_{w,t}^n &= 100\widehat{\varepsilon}_t^w, \\
&= \omega\widehat{u}_t^n. \\
\widehat{z}_t &= \big(1-\upsilon\big)\widehat{z}_{t-1} + \frac{\upsilon}{1-(h/\tau)}\left[\widehat{c}_t - \frac{h}{\tau}\,\widehat{c}_{t-1}\right],
\end{aligned}$$

where $\widehat{u}_t$ is the unemployment rate, $\widehat{u}_t^n$ is the natural rate of unemployment (the unemployment rate that would prevail in the absence of nominal wage rigidities), $\widehat{\varepsilon}_t^w$ is assumed to be an exogenous ARMA(1,1) process, while $\widehat{\varepsilon}_t^s$ is an AR(1) process representing an exogenous labor supply shock and υ is a parameter capturing the short-run wealth effects on labor supply. The labor force is given by $\widehat{l}_t = \widehat{e}_t + \widehat{u}_t$.

- Capital accumulation equation. The capital stock, $\widehat{\bar{k}}_t$, is determined by its lagged value, investment, and the investment-specific technology shock:

$$\widehat{\bar{k}}_t = \kappa_1\widehat{\bar{k}}_{t-1} + \big(1-\kappa_1\big)\widehat{i}_t + \kappa_2\widehat{\varepsilon}_t^q,$$

with $\kappa_1 = (1-\delta)/\tau$, and $\kappa_2 = (\tau+\delta-1)(1+\beta)\tau\varphi$. Capital services used in production is defined as: $\widehat{k}_t = \widehat{v}_t + \widehat{\bar{k}}_{t-1}$, where $\widehat{v}_t$ is capital utilization.

- Optimal capital utilization condition. The degree of capital utlization depends positively on the rental rate on capital:

$$\widehat{v}_t = \frac{1-\psi}{\psi}\,\widehat{r}_t^k,$$

where ψ is the elasticity of the capital utilization cost function.

- Optimal capital/labor input condition:

$$\widehat{k}_t = \widehat{w}_t - \widehat{p}_t - \widehat{r}_t^k + \widehat{n}_t.$$

- Monetary policy rule:

$$\widehat{r}_t = \rho_R\widehat{r}_{t-1} + \big(1-\rho_R\big)\left(r_\pi\widehat{\pi}_t + r_y\widehat{y}_t^{\mathrm{gap}} + r_{\Delta y}\Delta\widehat{y}_t^{\mathrm{gap}}\right) + \widehat{\varepsilon}_t^r,$$

with $y_t^{\mathrm{gap}} = \widehat{y}_t - \widehat{y}_t^{\mathrm{flex}}$, the difference between actual output and the output in the flexible price and wage economy, i.e. in the absence of distorting price and wage markup shocks.

As productivity is written in terms of hours worked, we also introduce an auxiliary equation to link from observed total employment $(\widehat{e}_t)$ to unobserved hours worked as in SW (2003):

$$\widehat{e}_t - \widehat{e}_{t-1} = E_t\widehat{e}_{t+1} - \widehat{e}_t + \frac{\big(1-\beta\theta_e\big)\big(1-\theta_e\big)}{\theta_e}\big(\widehat{n}_t - \widehat{e}_t\big),$$

where θ_e is the fraction of firms that are able to adjust employment to its desired total labor input.

The model is consistent with a balanced steady-state growth path, driven by deterministic labor augmenting trend growth. The observed variables for the euro area are given by quarterly data on the log of real GDP (y_t), the log of real private consumption (c_t), the log of real total investment (i_t), the log of the GDP deflator ($p_{y,t}$), the log of real wages $w_t - p_{y,t}$, the log of total employment (e_t), the unemployment rate (u_t), and the short-term nominal interest rate (r_t). With all variables except the unemployment rate and the interest rate being measured in first differences, the measurement equations for the euro area are given by:

$$\begin{bmatrix} \Delta y_t \\ \Delta c_t \\ \Delta i_t \\ \pi_{y,t} \\ \Delta w_t - \pi_{y,t} \\ \Delta e_t \\ u_t \\ r_t \end{bmatrix} = \begin{bmatrix} \bar{\tau} + \bar{e} \\ \bar{\tau} + \bar{e} \\ \bar{\tau} + \bar{e} \\ \bar{\pi} \\ \bar{\tau} \\ \bar{e} \\ \bar{u} \\ 4\bar{r} \end{bmatrix} + \begin{bmatrix} \Delta \widehat{y}_t \\ \Delta \widehat{c}_t \\ \Delta \widehat{i}_t \\ \widehat{\pi}_t \\ \Delta \widehat{w}_t - \Delta \widehat{\pi}_t \\ \Delta \widehat{e}_t \\ \widehat{u}_t \\ 4\widehat{r}_t \end{bmatrix}, \tag{1}$$

where $\widehat{u}_t = \widehat{l}_t - \widehat{e}_t$. The steady-state parameters are determined as

$$\bar{\tau} = 100(\tau - 1), \qquad \bar{\pi} = 100(\pi - 1), \qquad \bar{r} = 100\left(\frac{\pi\tau}{\beta} - 1\right), \qquad \bar{u} = 100\left(\frac{\phi_w - 1}{\omega}\right),$$

where $(\phi_w - 1)$ is the steady-state labor market markup, π is steady-state inflation, while $\bar{e}$ reflects steady-state labour force growth and is added to the real variables that are not measured in per capita terms.

The following parameters are not identified by the estimation procedure and therefore calibrated: $g_y = 0.18$, $\delta = 0.025$, and $\varepsilon_p = 10$.

3. The Euro Area RTDB and the SPF

Following GSW, we estimate the DSGE model using eight macroeconomic time series for the euro area: real GDP, consumption, investment, employment, unit labor costs, GDP deflator inflation, the Euribor rate and the unemployment rate, with the first five log differenced. Real-time vintages of these data are available for downloading from the ECB's Statistical Data Warehouse and described in Giannone et al. (2012).[4] The frequency of the vintages is monthly corresponding to their publication in the ECB's Monthly Bulletin and the first vintage starts in January 2001. The latest available vintage we use in this paper is March 2011.

[4] See also the detailed information about the RTDB in Giannone, Henry, Lalik, and Modugno (2010a).

Table 1 presents the time flow of data releases available for the euro area Real-Time Data Base (RTDB) and the Survey of Professional Forecasters (SPF).[5] We take the vintage of the last month of the quarter, in order to convert the monthly vintages into a quarterly vintage. As is clear from the Table, this implies that monthly unemployment and HICP inflation are available for the first month of the quarter, whereas the monthly interest rate is available for the first and second month of the quarter. As we need the full quarter of monthly observations to construct the quarterly observation, we ignore the partial information available during the quarter. This implies that quarterly unemployment, HICP inflation and the interest rate are observed with a one quarter lag. Using the vintage of the last month in the quarter implies that the quarterly series are also typically available with one lag, with the exception of employment and wage compensation which are only available with a two quarter lag. In the forecasting exercises of Section 5, we will use the method of Waggoner and Zha (1999) to "nowcast" employment and wages based on information during the same quarter on real GDP and the other variables.[6]

Each monthly data vintage from the RTDB typically only covers data starting in the mid 1990s. To extend the real-time data backwards, we make use of updates of the quarterly database constructed for estimating the Area-Wide Model (AWM). Since 2000 the AWM database is updated annually; see Fagan, Henry, and Mestre (2005).

Figure 1 plots the first release and the first annual revision of real GDP growth, GDP deflator inflation and the unemployment rate (left panel), as well as the difference between the first release and the first annual revision (right panel). The standard deviation of the annual revision in real GDP growth lies between 0.1 and 0.2 and is quite persistent. In the most recent recession, the downward revision was particularly large. The variability of the annual revision in inflation is of the same size but much less persistent. Finally, revisions in unemployment are the most persistent.

One source of revision in the euro area data set is the increasing number of EU countries being a member of the euro area. Over the estimation sample the euro area developed from 12 to 16 members: Updates 4, 5, and 6 of the AWM database cover the euro area 12 data and are taken from 2003, 2004, and 2006, respectively. The euro area 13 composition is available in update 7 from 2007, while the euro area 15 composition is available in update 8, dated 2008. The last two updates that we make use of, 9 and 10, both cover the euro area 16 composition and were frozen in 2009 and 2010. The available files prior to update 7 are dated in September although the time they were frozen is unknown; as of update 7 the AWM data is frozen at the beginning of August.

[5] See, e.g., Garcia (2003) and Bowles, Friz, Genre, Kenny, Meyler, and Rautanen (2007) for information on the ECB's SPF. For a recent study using SPF data, see Genre, Kenny, Meyler, and Timmermann (2013).

[6] Relative to the vintage date, employment and wages are actually backcasted, while the remaining variables are nowcasted.

Table 1 also shows that the SPF forecasts for HICP inflation, real GDP growth and unemployment typically become available in the first month of the quarter.[7] We associate this forecast with the quarter. The SPF data set contains average one-year and two-year ahead forecasts covering the period 1999Q1–2010Q4. Due to the different frequency and lags in the release of HICP inflation, real GDP and unemployment, the end date of the one-year and two-year ahead forecasts differs across the variables. For HICP inflation, the Q1-released one-year ahead forecasts refers to annual inflation in December in the same year, the Q2-release refers to March in the following year, etc. For real GDP growth, the "one-year ahead forecast" in the Q1-release refers to annual growth in the third quarter of the same year, etc. Finally, for the unemployment rate the "one-year ahead" in the Q1-release refers to the unemployment rate in November the same year, the Q2-release to the rate in February next year, etc. If we take the release-quarters as the current date for these forecasts, then for HICP inflation and unemployement we may think of this as having three and seven-quarters ahead forecasts and for real GDP growth two and six-quarters ahead forecasts.

The information set available to the professional forecasters is smaller than the RTDB available in the last month of the quarter, as last quarter's national account data are not available early in the quarter. On the other hand, it is clear that the professional forecasters have a lot more information available to nowcast the last quarter than the data we use from the RTDB. As a result, it is not clear whether the net information advantage is positive or negative.

4. Full-Sample Estimation Results

In this section we first discuss the estimation results using the latest-vintage full sample data set and make some comparisons with those reported for the United States in GSW (2011). We estimate the model over the period 1985Q1–2010Q4 using Bayesian full-system estimation techniques as in SW (2003) and (2007). The period from 1980Q1 till 1984Q4 is used as training period.[8]

Table 2 reports the parameter estimates as well as the prior distributions that we have used.[9] A few striking differences with the US results are worth mentioning. First, the average unemployment rate over the 1985–2010 period is quite a bit higher in the euro area (about nine percent) than in the United States (five percent). In steady state, the unemployment rate is

[7] The inflation forecasts in the SPF only covers HICP inflation and not the GDP deflator. We therefore use the HICP inflation forecasts. In the estimation under the noise interpretation, the difference is picked up by the measurement error term. The model under the news interpretation is estimated from the RTDB data only, and SPF forecasts are only used as conditioning information when forecasting.

[8] Provided that the log-linearized GSW model has a unique and convergent solution for a given value of the parameters it can be written on state-space form. The Kalman filtering and smoothing algorithms that take missing observations into account can then be used for estimation and (conditional) forecasting with state-space models; see, e.g., Durbin and Koopman (2012). The same algorithms are also used when the GSW model is extended to take the SPF into account via either the news or the noise interpretation.

[9] Most of the priors we have used are standard in the literature and, overall, quite uninformative. With many parameters to estimate, there are not many good reasons for picking different priors for the euro area and the US. For those where one may argue that the parameters should be different, such as the stickiness parameters, we have opted to use quite uninformative priors.

proportional to the wage markup and the labor supply elasticity. For the euro area, the wage markup is estimated to be quite a bit higher[10] (around 50 percent) and the labor supply elasticity somewhat lower. In other words, labor supply responds less to changes in real wages in the euro area.

Second, the parameter, ν, governing the short-run wealth effects on labor supply, is quite small (0.08) as in the United States. Roughly speaking this amounts to a preference specification closer to that in Greenwood, Hercowitz, and Huffman (1988), in which the wealth effects are close to zero in the short run. As discussed at length in GSW, this helps ensure that not only employment, but also the labor force moves procyclically in response to most shocks.

Third, turning to some of the other parameters that enter the price and wage Phillips curve, the euro area economy appears to be much more sticky than the US economy. The estimated degree of price and wage indexation is relatively small (around 0.25) in both areas, but the estimated Calvo probability of unchanged wages and prices are quite a bit higher. The average wage contract duration is a bit higher than 3 quarters, whereas the average duration of unchanged prices is higher than six quarters. This is consistent with some of the micro evidence on price and wage adjustment.[11]

Fourth, it is worth pointing out that the monetary policy reaction coefficient to the output gap (defined as the deviation relative to the constant markup output) is quite high (0.19), whereas the coefficient on inflation is quite a bit lower (though higher than one).

Finally, focusing on the volatility and persistence of the eight structural shocks, the striking difference is that the risk premium shock is much more persistent in the euro area, whereas the investment-specific technology shock is much less persistent.

Overall, the estimation results for the euro area point to a less flexible economy with more persistence in the effects of various shocks on economic activity, prices and unemployment.

Before turning to the real-time forecasting results, it is also worth discussing briefly the forecast error variance decomposition at the 10 and 40 quarter horizon (Table 3). At the business cycle frequency about half of the fluctuations in output are driven by demand shocks and particularly the risk premium shock. The risk premium shock explains almost two thirds of the movement in unemployment at the 2.5 year horizon. The monetary policy shock another 12 percent. The most important shock driving output is the productivity shock. Price inflation is mostly driven by the price markup shock (61 percent) and the wage markup shock (17 percent).

In the longer run (after ten years), the role of wage markup shocks becomes more important in driving both unemployment and inflation. This is, however, much less so than in the United States where those shocks account for between 60 and 80 percent of the movements. The role of demand shocks in explaining real output and unemployment falls somewhat in the longer run,

[10] This may reflect a higher degree of unionization and collective wage bargaining in the euro area than in the US; see WDN (2009).

[11] See, for instance, Altissimo, Ehrmann, and Smets (2006) and WDN (2009).

but remains much more important than in the US. Productivity shocks become relatively more important. In the longer run, inflation is mostly driven by price and wage markup shocks.

These full-sample estimation results are very similar when we re-estimate the model using the SPF forecasts as noisy indicators of the model-consistent expectations (see Section 5). We find that the estimates of the standard deviation of the iid normal measurement error are relatively large: 0.76 for expected annual real GDP growth, 0.32 for expected GDP deflator inflation and 0.60 for the expected unemployment rate.

5. Real-Time Forecasting Performance

In this section we evaluate the real-time forecasting performance of the GSW model over the EMU period and compare it with seven alternative models. With the exception of three simple non-structural models, each of these models is re-estimated on an annual basis from the first RTDB vintage in 2001Q1 onwards; i.e. the second based on the 2002Q1 vintage and so on. The forecasts are conditional on the data observed in the last historical period, where the available information in that period is used to backcast the variables that are missing in that period (typically employment and wage compensation). For example, the RTDB vintage 2001Q1 forecasts are computed for 2000Q4–2001Q4 with conditioning assumptions for 2000Q4 based on the historical data available for that quarter.

One question in real-time forecast evaluation exercises is which actual data to use to evaluate the forecast against and to calculate the forecast errors. As is common in the literature, we use the first annual revision of the data (as in Figure 1). We have checked the robustness of our findings against two possible alternatives for the actual data: (1) the first release data and (2) latest vintage data. Overall, the results are very similar.

We compare the point forecasts of the GSW model with seven alternative models. The four competing non-structural models are the random-walk, the sample mean, an AR(1) model for deviations around the sample mean, and a BVAR model using the same eight observed variables as in the GSW model. The sample mean model is re-estimated in each quarter and uses the data from the last 40 quarters. The AR(1) model takes the deviations around the sample mean for the previous model and adds an autoregressive lag to this and is therefore similar to the AR(1)-gap model used for inflation forecasting in Faust and Wright (2013). The autoregressive parameter for each variable is re-estimated for each RTDB vintage using the available data for that variable.[12]

[12] We have also tested replacing the sample mean for annual real GDP growth, annual GDP deflator inflation, and the unemployment rate with the "five-year-ahead" mean point forecasts of annual real GDP growth, HICP inflation, and the unemployment rate from the SPF vintages. Such data are available for each vintage from 2001Q1 and onwards, and for Q1 in 1999 and 2000. The use of the SPF for replacing means was suggested to us by one of the referees. However, the MSE's from these variants of the sample mean and AR(1)-gap models worsened the forecasting performance of these two models for both real GDP growth and GDP deflator inflation, while the forecasts of the unemployment rate improved somewhat. We therefore decided against using the "five-year-ahead" mean point forecasts from the SPF. Note also that the quotation marks reflect the fact that the five-year-ahead forecasts concerns calender years and are therefore for most vintages not strictly five-year-ahead.

The BVAR estimation follows Villani (2009). It is estimated using a prior on the steady-state mean and standard deviation of the variables which is the same as the prior steady-state mean and standard deviation used in estimating the DSGE model (with the exception of the standard deviation of unemployment). In addition, a fairly standard Minnesota-type prior with a diffuse prior on the covariance matrix is used.

The benchmark GSW model is also compared with three alternative estimated GSW models in which the mean forecasts of real GDP growth, HICP inflation, and unemployment from the SPF are used as additional information. We consider two interpretations of those professional forecasts.[13] Under the "noise" interpretation, the mean professional forecasts are assumed to be noisy indicators of the rational expectations forecasts implied by the DSGE model. Specifically, mean SPF forecasts of annual inflation is added to the set of measurement equations in (1):

$$\pi^{\mathrm{a}}_{t+3|t} = 4\bar{\pi} + E_t\big[\widehat{\pi}_{t+3} + \widehat{\pi}_{t+2} + \widehat{\pi}_{t+1} + \widehat{\pi}_t\big] + \eta_{\pi,t},$$

where $\pi^{\mathrm{a}}_{t+3|t}$ is the mean SPF forecast of annual inflation between $t+3$ and $t-1$ in period t, and $E_t[\widehat{\pi}_{t+i}]$ is the rationally expected quarterly inflation rate (in deviation from steady-state) in period $t+i$ using information available until period t in the GSW model. The sum of the first and second term (in brackets) on the right hand side is therefore equal to the rational expectations forecast of annual inflation three quarters ahead, while $\eta_{\pi,t}$ is an iid normal measurement error with mean zero and standard deviation σ_π. Measurement equations are similarly added to (1) for the mean SPF forecast of unemployment three quarters ahead and for the mean SPF forecast of annual real GDP growth two quarters ahead, both with individual measurement errors. The additional randomness makes it possible to estimate the parameters of the GSW model extended with the SPF data.[14] As discussed in Section 4, the standard deviations of the errors in the measurement equations are quite large. Conditional forecasts are computed using the Waggoner and Zha (1999) approach with hard conditions; see Warne (2013) for details on the implementation in linear state-space models.

Under the "news" interpretation, it is assumed that the forecasts reveal the presence of expected future structural shocks in line with those estimated over the past. This exercise is similar to the one performed by Del Negro and Schorfheide (2013) for the United States. In this case, the corresponding DSGE model forecast of annual real GDP growth two quarters ahead, annual GDP deflator inflation three quarters ahead, and the unemployment rate three quarters ahead will be identical to the mean SPF forecast. The Waggoner and Zha methodology is again used to compute the conditional forecasts, and we report results for two cases: one in which we

[13] While the SPF is conducted quarterly and the data we have collected from the RTDB are also quarterly, it would be possible to augment the GSW model with monthly conjunctural data using a mixed frequency approach as suggested by Giannone, Monti, and Reichlin (2010b). An important aspect of their methodology is that the extra information provided by the monthly conjunctural data is that it is valuable (relative to the augmented model) only because it is more timely. For an application to short-term forecasting of Austrian real GDP see, e.g., Cervená and Schneider (2010).

[14] For inflation expectations, our approach is similar to Del Negro and Eusepi (2011), where we add measurement noise instead of modifying the policy rule.

only use the one-year ahead forecasts and another one in which we use in addition the two-year ahead SPF forecasts.[15]

The forecasting performance exercise below addresses two main questions. First, are the benchmark GSW model forecasts improved upon by utilizing the SPF data? Second, can the GSW models with and without conditioning on the SPF data compete with the non-structural models?

Figure 2 displays the RMSEs for the three cases of the GSW model where the SPF data are utilized as conditioning information *relative* to the RMSE of the benchmark GSW model. This means that for values below (above) unity the SPF-based GSW model has a lower (higher) RMSE than the benchmark GSW model. A few finding are worth highlighting. First, the relative RMSEs are typically lower than unity for real GDP growth, inflation, and unemployment suggesting that the SPF data is useful for improving the forecasts of these variables. Second, the relative RMSEs for the two cases of the GSW model subject to the news interpretation (dashed and dotted lines in Figure 2) are close and may reflect that the information in the 2-year-ahead SPF at best leads to a marginal improvement in forecasting performance of the GSW model. Third, the GSW model subject to the noise interpretation (dash-dotted lines) appear to worsen the real wage forecasts and marginally improve the short-run employment forecasts. For the news cases, the opposite result is supported. Fourth, all models using SPF data seem to worsen the interest rate forecasts. Fifth, consumption and investment forecasts may be marginally improved with a tendence for the news models to fare better than the noise model.

To examine the findings in more detail we turn to Table 4, where percentile values from the approximating distribution of the modified Diebold-Mariano test statistic are displayed; see Harvey, Leybourne, and Newbold (1997) for computational details. We have opted to follow the suggestion of Harvey et al. and compare the modified statistic to the Student's t-distribution with $N_h - 1$ degrees of freedom, with N_h being the number of h-step-ahead forecast errors, rather than to its asymptotic normal distribution. A low percentile value indicates that the SPF-based forecasts (noise or news) are better, while a high percentile value favors the forecasts of the benchmark model. A value below or equal to 5 percent is shown in bold in the Table, while a value above or equal to 95 percent is displayed in italics. With respect to the five observations listed in the previous paragraph, the results in Table 4 supports the finding that the inflation forecasts are improved from a RMSE perspective when the SPF is taken into account for both

[15] An alternative to the noise and news implementations of the SPF data would be to employ the methodology developed by Monti (2010). While her approach has several attractive properties we have opted not to use it. First, the monthly vintage that we have selected to represent the quarter (third month of each quarter) is not consistent with the condition that judgmental forecasts are based on an information set which comprises the information available in the RTDB vintage. For the euro area RTDB this condition could only be satisfied if the first month of each quarter would have been used, while our choice of using the third month is based on using a vintage that basically includes information available when the ECB/Eurosystem staff macroeconomic projections exercises are conducted. Second, her approach involves replacing unknown population moments of the Kalman filter with sample moments calculated using the judgmental forecasts. The SPF data has a very short historical sample for the first vintages in the forecasting study, namely eight observations on the judgmental forecasts for the 2000Q4 vintage. It therefore seems unlikely that the unknown population moments required by her approach can be meaningfully estimated for this vintage and those that immediately follow.

the news models and the noise model. Furthermore, the news models seem to help improving the short-run real GDP growth forecasts, while the evidence on unemployment suggests that the forecasts are neither improved when using the SPF data nor are they exacerbated.

The results in Table 4 also supports the third point above, i.e. that the SPF information is useful for improving real wage forecasts for the news models and that this data aggrevates the real wage forecasts for the noise model. Furthermore, the interest rate forecasts are generally exacerbated when the SPF data are included, especially for the shorter forecast horizons. Overall, it would appear from the modified Diebold-Mariano tests that the SPF data is useful for improving the benchmark GSW model point forecasts under the news interpretation, while the evidence for the noise case is less convincing. It should be kept in mind that each test is based on $N_h = 36 - h$ observations for the h-step-ahead forecasts and should therefore be interpreted with caution as the reference distribution need not provide a good approximation to the unknown small sample distribution of the test statistic.

Turning to the second question about how well the GSW model can compete with the non-structural models, we first consider the RMSEs reported in Figure 3, which are all relative to the RMSEs of the BVAR model. In other words, values above unity favor the BVAR and below unity the specified model. Concerning the four GSW model based cases (denoted by DSGE, noise, 1-year, 1&2-year in the graph) it can be seen that they generally have lower RMSEs for inflation than the BVAR, especially when taking the SPF forecasts on board, and for the interest rate. On the other hand, the BVAR has lower RMSEs for consumption, employment and real wages, although the news model based on the 1 and 2-year-ahead SPF comes close for wages. Compared with the nonstructural models, the BVAR has lower RMSEs for output, employment, and unemployment, while the non-structural models have better point forecasts for inflation and real wages.

Table 5 displays the percentile values of the modified Diebold-Mariano statistics for the four GSW model based cases with and without the SPF data relative to the BVAR in the upper half, and for the three univariate non-structural models versus the BVAR in the lower half. Our findings based on the RMSEs are to some extent confirmed, especially for the comparisons between the GSW model based cases and the BVAR. The inflation forecasts of the GSW model cases seem to outperform the BVAR when the SPF conditioning information is utilized, while consumption in particular and real wages are generally better forecasted using the BVAR. As noted above, the news case with 1 and 2-year-ahead SPF data seems to improve real wage forecast sufficiently relative to the benchmark GSW and the news case using only the 1-year-ahead SPF data to match the point forecasts of the BVAR from a MSE perspective. An inspection of the forecast errors reveals that the GSW model systematically overpredicts real wage growth, while it underpredicts consumption. A similar result was found in Christoffel, Coenen, and Warne (2011) which evaluated the forecast performance of the NAWM for the euro area; see also Warne, Coenen, and Christoffel (2013). The New Keynesian model, which assumes a constant steady-state labor share and consumption to output ratio, has a difficult time explaining the falling labor share and the rising consumption to GDP ratio over this period. The non-structural models (except for the sample mean) do better in this respect, especially for the one and two-step-ahead point forecasts. The noise versus news interpretation does matter for the predictive performance regarding wage growth. In the news model the higher inflation HICP forecasts are rationalised by higher expected markup shocks, which at the same time tend to reduce expected wage growth and thereby alleviate part of the upward bias of the benchmark DSGE model. In the noise model, the overprediction of real wage growth is instead magnified.

The graphs in Figure 4 plot the log-determinant and the trace statistic of the MSE matrix for the four GSW model based cases with and without the SPF data, the AR(1) model, the sample mean, and the random walk model relative to the values obtained for the BVAR model. Concerning the log-determinant statistic it should be noted that it is negative for all models and we have therefore opted to compute the relative statistic as minus unity times the ratio of log-determinants. Hence, a low value suggests better point forecasts from a multivariate perspective with minus one being the log-determinant of the BVAR model. Similarly, low values of the trace statistic is also an indicator of better multivariate point forecasts, with values below unity reflecting that the trace of the MSE matrix of the given model is lower than the corresponding trace statistic of the BVAR model.

From the trace statistics on the right hand side in Figure 4 it would seem that the point forecasts of the GSW model based cases under the news interpretation do (marginally) better than all other models for all horizons, while the point forecasts of the benchmark GSW model and noise broadly matches those of the BVAR. The three simple non-structural models all have trace statistics greater than the BVAR, especially the sample mean for the shorter forecast horizons. Turning to the log-determinant on the left hand side of Figure 4 the picture is far more complex. For one-step-ahead forecasts the BVAR does better, while the AR(1) and the noise case of the GSW model have statistics close to the BVAR for two-step-ahead forecasts. For the three and four-step-ahead forecases, the AR(1) model seems to perform best according to this metric.

To summarize the findings on the relative forecasting performance of the models, it sppears that the SPF data is useful for improving the point forecasts of the GSW model, in particular for the cases subject to the news interpretation. The main cost concerns the forecasts of the short-term nominal interest rate at the one to three-step-ahead horizon. Once the point forecasts of the GSW model are compared with those of non-structural models the results are mixed. Compared with a BVAR model which does not take the SPF data into account, the news models improve the inflation forecasts and the one-step-ahead real GDP growth forecasts. However, these results seem to be mainly driven by utilizing the SPF rather than having a structural model. Hence, a BVAR which takes the SPF data on board when forecasting may very well do at least as good the news models when forecasting these variables. At the same time, the real wage point forecasts of the BVAR can be improved on, in particular when comparing them to forecasts from the three univariate non-structural models.

The final exercise we shall conduct concerns how good the forecasts of these models are from an absolute perspective. To study this issue we follow, e.g., Mincer and Zarnowitz (1969) and Edge and Gürkaynak (2010) and conduct so called Mincer-Zarnowitz regressions. Let x_t be the (actual) value at time t of the forecasted variable while the h-step-ahead forecast of this variable is denoted by $x_{t|t-h}^{(m)}$ for model m. The regression equation is now

$$x_t = \alpha_h^{(m)} + \beta_h^{(m)} x_{t|t-h}^{(m)} + \varepsilon_{h,t}^{(m)}, \quad t = 1, \ldots, N_h,$$

where $\varepsilon_{h,t}^{(m)}$ is a mean zero error term. If the forecasts of model m are efficient, then the intercept is zero while the slope is unity, and the variance of the error term is small compared with the variance of x_t (high R^2). An intercept different from zero indicates that the forecasts have on average been biased (relative to the selected actual), and if the slope coefficient is greater (less) than unity then the forecasts have consistently underpredicted (overpredicted) the variable. A low R^2 indicates that little of the variation of x_t is captured by the variation of the forecast.

Table 6 summarizes the evidence from the forecast accuracy regressions, limiting the presentation to three models (the benchmark GSW model, the news model with 1-year-ahead SPF data, and the BVAR model), and four variables (output, inflation, unemployment, and the interest rate).[16] The results suggest that forecasts of inflation, unemployment, and the interest rate have been poor by all methods. For output, however, the regression outcomes are broadly consistent with efficient forecasts, especially for the BVAR and the longer forecast horizons.

6. Conclusion

In this paper we have evaluated the real-time forecasting performance of the New Keynesian model of Galí, Smets, and Wouters (2012) estimated on euro area data. First, we find that the benchmark GSW model forecasts tend to be improved when adding one to two-year-ahead professional forecasts of real GDP growth, inflation, and the unemployment rate to the conditioning data without otherwise changing the DSGE model. This is consistent with the news interpretation where it is assumed that the forecasts reveal the presence of expected future structural shocks in line with those estimated over the past. The consequence of utilizing the professional forecasts within the context of the GSW model is that inflation and real wage forecasts improve considerably, as well as one- and two-quarter-ahead real GDP forecasts. The cost appears to be that the short-term nominal interest rate forecasts deteriorate. The evidence under the noise

[16] The standard errors within parenthesis have been computed with the Newey and West (1987) method under the assumption that the error term follows an MA(h) process, except for unemplyment and the interest rate where it is assumed to follow an MA($h-1$) process. This is due to the fact that, excluding the unemployment rate and the interest rate, all other variables need to be nowcasted for some vintages. There are 3 such vintages for output, 7 vintages for consumption and investment, 14 for inflation, and all 36 vintages for employment and real wages. We have also made use of this information when computing the modified Diebold-Mariano test.

interpretation, where the mean professional forecasts are assumed to be noisy indicators of the rational expectations forecasts implied by the DSGE model, is less convincing. Although inflation forecasts seem to be improved under the noise model, real wage forecasts seem to deteriorate as well as the short-term nominal interest rate forecasts.

Second, a BVAR model is also able to improve the benchmark GSW model forecasts, in particular for consumption and real wages where the DSGE model systematically overpredicts real wage growth and underpredicts consumption. These variables are also poorly predicted under the noise and news interpretations of the GSW model relative to the BVAR. At the same time, the inflation forecasts from these models are typically much improved relative to the BVAR.

Third, the point forecasts of the variables are generally not efficient when viewed through the lens of Mincer-Zarnowitz regressions. One exception concerns output where the regression evidence for most of the models is consistent with a zero intercept, unit slope coefficient, and a relatively high adjusted R^2. It should be kept in mind that the forecast performance study covers a sample with 36 real-time vintages from the ECB's Statistical Data Warehouse. In view of this rather small sample, the evidence needs to be interpreted with caution. Furthermore, we only study point forecasts and do not take other moments of the predictive distributions into account.

Acknowledgements

We are grateful to two anonymous referees and the handling editor for useful comments. We have also benefitted from comments by Jonathan Wright and Francesca Monti, as well as participants at the American Economic Association Annual Meeting in Chicago, January 2012, the 7th ECB Workshop on Forecasting Techniques, May 2012, and the SPF Workshop at the European Central Bank, November 2013. The views expressed are our own and should not be attributed to the ECB, the NBB or the Eurosystem.

References

Altissimo, F., Ehrmann, M., and Smets, F. (2006), "Inflation Persistence and Price-Setting Behaviour in the Euro Area: A Summary of the IPN Evidence," ECB Occasional Paper Series No. 46.

Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A., and Rautanen, T. (2007), "The ECB Survey of Professional Forecasters (SPF): A Review After Eight Years' Experience," ECB Occasional Paper Series No. 59.

Cervená, M. and Schneider, M. (2010), "Short-Term Forecasting GDP with a DSGE Model Augmented by Monthly Indicators," Oestereichische Nationalbank Working Paper No. 163.

Christoffel, K., Coenen, G., and Warne, A. (2011), "Forecasting with DSGE Models," in M. P. Clements and D. F. Hendry (Editors), *The Oxford Handbook of Economic Forecasting*, 89–127, Oxford University Press, New York.

Coenen, G., Levin, A., and Wieland, V. (2005), "Data Uncertainty and the Role of Money as Information Variable for Monetary Policy," *European Economic Review*, 49, 975–1006.

Coenen, G. and Warne, A. (2013), "Risks to Price Stability, the Zero Lower Bound and Forward Guidance: A Real-Time Assessment," ECB Working Paper Series No. 1582.

Croushore, D. (2011), "Frontiers of Real-Time Data Analysis," *Journal of Economic Literature*, 49, 72–110.

Croushore, D. and Stark, T. (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, 105, 111–130.

Del Negro, M. and Eusepi, S. (2011), "Fitting Observed Inflation Expectations," *Journal of Economic Dynamics and Control*, 35, 2105–2131.

Del Negro, M. and Schorfheide, F. (2013), "DSGE Model-Based Forecasting," in G. Elliott and A. Timmermann (Editors), *Handbook of Economic Forecasting*, volume 2, 57–140, North Holland, Amsterdam.

Diebold, F. X. and Rudebusch, G. D. (1991), "Forecasting Output with the Composite Leading Index: A Real-Time Analysis," *Journal of the American Statistical Association*, 86, 603–610.

Durbin, J. and Koopman, S. J. (2012), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, 2nd edition.

Edge, R. M. and Gürkaynak, R. S. (2010), "How Useful Are Estimated DSGE Model Forecasts for Central Bankers?" *Brookings Papers on Economic Activity*, Fall, 209–259.

Fagan, G., Henry, J., and Mestre, R. (2005), "An Area-Wide Model for the Euro Area," *Economic Modelling*, 22, 39–59.

Faust, J. and Wright, J. H. (2013), "Forecasting Inflation," in G. Elliott and A. Timmermann (Editors), *Handbook of Economic Forecasting*, volume 2, 2–56, North Holland, Amsterdam.

Galí, J., Smets, F., and Wouters, R. (2012), "Unemployment in an Estimated New Keynesian Model," in D. Acemoglu and M. Woodford (Editors), *NBER Macroeconomics Annual 2011*, 329–360, University of Chicago Press.

Garcia, J. A. (2003), "An Introduction to the ECB's Survey of Professional Forecasters," ECB Occasional Paper Series No. 8.

Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2013), "Combining Expert Forecasts: Can Anything Beat the Simple Average?" *International Journal of Forecasting*, 29, 108–121.

Giannone, D., Henry, J., Lalik, M., and Modugno, M. (2010a), "An Area-Wide Real-Time Database for the Euro Area," ECB Working Paper Series No. 1145.

Giannone, D., Henry, J., Lalik, M., and Modugno, M. (2012), "An Area-Wide Real-Time Database for the Euro Area," *Review of Economics and Statistics*, 94, 1000–1013.

Giannone, D., Monti, F., and Reichlin, L. (2010b), "Incorporating Conjunctural Analysis in Structural Models," in V. Wieland (Editor), *The Science and Practice of Monetary Policy Today*, 41–57, Springer Verlag.

Greenwood, J., Hercowitz, Z., and Huffman, G. (1988), "Investment, Capacity Utilization and the Real Business Cycle," *American Economic Review*, 78, 402–417.

Harvey, D., Leybourne, S., and Newbold, P. (1997), "testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13, 281–291.

Mincer, J. A. and Zarnowitz, V. (1969), "The Evaluation of Economic Forecasts," in J. A. Mincer (Editor), *Economic Forecasts and Expectagtions: Analysis of Forecasting Behavior and Performance*, 3–46, NBER.

Monti, F. (2010), "Combining Judgment and Models," *Journal of Money, Credit and Banking*, 42, 1641–1662.

Newey, W. K. and West, K. D. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

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

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

Villani, M. (2009), "Steady-State Priors for Vector Autoregressions," *Journal of Applied Econometrics*, 24, 630–650.

Waggoner, D. F. and Zha, T. (1999), "Conditional Forecasts in Dynamic Multivariate Models," *Review of Economics and Statistics*, 81, 639–651.

Warne, A. (2013), "YADA Manual — Computational Details," Manuscript, European Central Bank. Available with the YADA distribution.

Warne, A., Coenen, G., and Christoffel, K. (2013), "Predictive Likelihood Comparisons with DSGE and DSGE-VAR Models," ECB Working Paper Series No. 1536.

WDN (2009), "Wage Dynamics in Europe: Final Report of the Wage Dynamics Network (WDN)," Mimeo, http://www.ecb.europa.eu/home/pdf/wdn_finalreport_dec2009.pdf.

TABLE 1: Time flow of data releases available for the RTDB and the SPF over a quarter.

	Month 1		Month 2	Month 3
			Quarter	
	⇑ RTDB M1	⇑ SPF	⇑ RTDB M2	⇑ RTDB M3
Monthly series	u_{m-2}	u_{m-2}	u_{m-2}	u_{m-2}
	π_{m-2}	$\boldsymbol{\pi_{m-1}}$	π_{m-2}	π_{m-2}
	r_{m-1}	r_{m-1}	r_{m-1}	r_{m-1}
Quarterly series	y_{q-2}	y_{q-2}	y_{q-2}	y_{q-1}
	c_{q-2}	c_{q-2}	c_{q-2}	c_{q-1}
	i_{q-2}	i_{q-2}	i_{q-2}	i_{q-1}
	$p_{y,q-2}$	$p_{y,q-2}$	$p_{y,q-2}$	$p_{y,q-1}$
	e_{q-2}	e_{q-2}	e_{q-2}	e_{q-2}
	w_{q-2}	w_{q-2}	w_{q-2}	w_{q-2}
	u_{q-2}	u_{q-2}	u_{q-1}	u_{q-1}
	r_{q-1}	r_{q-1}	r_{q-1}	r_{q-1}

Note: Unemployment is denoted by u, HICP by π, the average quarterly 3-month nominal interest rate by r, real GDP by y, real private consumption by c, the GDP deflator by p_y, total employment by e, and wages by w.

TABLE 2: Prior distributions and posterior estimates for the US and euro area models.

	Prior			Posterior							
				United States (1966:1–2007:4)				Euro area (1985:1–2009:4)			
parameter	type	mean	st.dev	mode	mean	5%	95%	mode	mean	5%	95%
structural parameters											
φ	N	4.0	1.0	4.09	3.96	2.34	5.58	4.65	4.77	3.34	6.31
h	B	0.7	0.1	0.78	0.75	0.65	0.85	0.65	0.64	0.54	0.72
ω	N	2.0	1.0	3.99	4.35	3.37	5.32	5.66	5.56	4.49	6.63
υ	B	0.5	0.2	0.02	0.02	0.01	0.04	0.06	0.12	0.03	0.34
θ_p	B	0.5	0.15	0.58	0.62	0.53	0.71	0.85	0.85	0.79	0.90
θ_w	B	0.5	0.15	0.47	0.55	0.44	0.66	0.74	0.72	0.60	0.89
γ_p	B	0.5	0.15	0.26	0.49	0.20	0.78	0.22	0.27	0.11	0.49
γ_w	B	0.5	0.15	0.16	0.18	0.07	0.29	0.22	0.25	0.12	0.42
ψ	B	0.5	0.15	0.57	0.56	0.36	0.75	0.46	0.48	0.29	0.69
ϕ_p	N	1.25	0.12	1.74	1.74	1.61	1.88	1.48	1.48	1.31	1.65
ϕ_w	N	1.25	0.12	1.18	1.22	1.15	1.29	1.53	1.51	1.41	1.62
α	N	0.3	0.05	0.17	0.17	0.14	0.20	0.22	0.22	0.19	0.26
θ_e	B	0.5	0.15	–	–	–	–	0.71	0.71	0.65	0.76
ρ_R	B	0.75	0.1	0.85	0.86	0.82	0.89	0.86	0.86	0.81	0.89
r_π	N	1.5	0.25	1.91	1.89	1.62	2.16	1.25	1.27	1.02	1.57
r_y	N	0.12	0.05	0.15	0.16	0.11	0.22	0.19	0.19	0.14	0.25
$r_{\Delta y}$	N	0.12	0.05	0.24	0.25	0.20	0.30	0.02	0.02	−0.00	0.06
$\bar{\pi}$	G	0.62	0.1	0.62	0.66	0.49	0.83	0.55	0.56	0.44	0.70
$\bar{\beta}$	G	0.25	0.1	0.31	0.31	0.17	0.43	0.24	0.27	0.13	0.43
$\bar{l}$	N	0.0	2.0	−1.65	−1.52	−3.83	0.77	–	–	–	–
$\bar{e}$	N	0.2	0.5	–	–	–	–	0.22	0.22	0.20	0.25
τ	N	0.4	0.1	0.34	0.34	0.30	0.37	0.14	0.14	0.08	0.20
τ_{wE}	N	0.2	0.1	0.07	0.08	0.03	0.12	–	–	–	–

Note: The prior distribution types are normal (N), standardized beta (B), gamma (G), and uniform (U). The parameter $\bar{\beta} = 100(\beta^{-1} - 1)$. The parameter ϕ_w has prior mean 1.5 and standard deviation 0.25 for the euro area, while the parameter τ has prior mean 0.3 and standard deviation 0.1 for the vintages prior to 2008 and standard deviation 0.05 thereafter. The US results are taken from Galí, Smets, and Wouters (2012).

Table 2: Continued.

	Prior			Posterior							
				United States (1966:1–2007:4)				Euro area (1985:1–2009:4)			
parameter	type	mean	st.dev	mode	mean	5%	95%	mode	mean	5%	95%
st.dev. of the innovations											
σ_a	U	2.5	1.44	0.41	0.42	0.37	0.46	0.58	0.60	0.46	0.78
σ_b	U	2.5	1.44	1.73	1.60	0.56	2.50	0.24	0.28	0.16	0.44
σ_g	U	2.5	1.44	0.47	0.48	0.43	0.52	0.30	0.31	0.28	0.35
σ_q	U	2.5	1.44	0.42	0.42	0.34	0.49	0.49	0.49	0.39	0.60
σ_r	U	2.5	1.44	0.21	0.22	0.19	0.24	0.11	0.11	0.10	0.13
σ_p	U	2.5	1.44	0.05	0.11	0.03	0.18	0.35	0.49	0.21	1.02
σ_w	U	2.5	1.44	0.04	0.06	0.01	0.13	0.30	0.76	0.16	3.66
σ_s	U	2.5	1.44	1.07	1.17	0.89	1.45	1.02	1.07	0.85	1.33
persistence of the exogenous processes: ρ = AR(1), μ = MA(1)											
ρ_a	B	0.5	0.2	0.98	0.98	0.97	0.99	0.98	0.98	0.97	0.99
ρ_b	B	0.5	0.2	0.36	0.42	0.19	0.67	0.91	0.91	0.84	0.96
ρ_g	B	0.5	0.2	0.97	0.97	0.96	0.99	0.99	0.99	0.98	1.00
ρ_{ga}	N	0.5	0.25	0.69	0.69	0.55	0.83	0.18	0.19	0.09	0.30
ρ_q	B	0.5	0.2	0.72	0.75	0.62	0.88	0.36	0.35	0.18	0.53
ρ_r	B	0.5	0.2	0.09	0.10	0.02	0.17	0.30	0.30	0.16	0.44
ρ_p	B	0.5	0.2	0.76	0.43	0.07	0.79	0.56	0.53	0.27	0.76
μ_p	B	0.5	0.2	0.59	0.57	0.24	0.96	0.44	0.47	0.25	0.71
ρ_w	B	0.5	0.2	0.99	0.98	0.97	1.00	0.91	0.89	0.81	0.95
μ_w	B	0.5	0.2	0.67	0.63	0.35	0.91	0.85	0.80	0.65	0.90

Note: The uniform priors all have lower bound 0 and upper bound 5. The parameter ρ_{ga} measures the effect of TFP innovations on exogenous spending. The persistence parameter for the labor supply process $\widehat{\varepsilon}_t^s$ is calibrated and given by $\rho_s = 0.999$.

TABLE 3: Variance decompositions in percent for the US and the euro area models.

variance decomposition	output	inflation	employment	unemployment
10 quarter horizon				
demand shocks				
risk premium	6 / 32	2 / 12	16 / 67	20 / 64
exogenous spending	3 / 0	1 / 0	7 / 1	8 / 0
investment specific	9 / 2	3 / 0	12 / 2	10 / 1
monetary policy	5 / 6	8 / 0	11 / 11	11 / 11
supply shocks				
productivity	59 / 54	6 / 8	5 / 1	4 / 2
price markup	2 / 0	27 / 61	3 / 0	0 / 0
labor market shocks				
wage markup	6 / 0	53 / 17	18 / 2	41 / 15
labor supply	11 / 3	0 / 0	29 / 12	5 / 4
40 quarter horizon				
demand shocks				
risk premium	2 / 14	1 / 12	6 / 43	7 / 54
exogenous spending	1 / 0	1 / 0	3 / 4	3 / 0
investment specific	5 / 1	2 / 0	4 / 1	3 / 1
monetary policy	2 / 2	5 / 0	4 / 7	4 / 9
supply shocks				
productivity	56 / 75	4 / 12	3 / 0	1 / 0
price markup	1 / 0	18 / 53	1 / 2	0 / 0
labor market shocks				
wage markup	17 / 0	67 / 19	39 / 4	80 / 27
labor supply	17 / 5	0 / 0	40 / 0	2 / 3

Note: The first entry gives the variance decompositions for the US (1966:1–2007:4) from GSW (2012); the second entry for the euro area (1985:1–2009:4).

Table 4: Modified Diebold-Mariano tests of GSW model based cases with SPF conditioning information versus the benchmark GSW model (RTDB vintages 2001Q1–2010Q4).

	Noise				News 1-year				News 1 & 2-year			
Variable	1	2	3	4	1	2	3	4	1	2	3	4
Output	0.22	0.21	0.42	0.48	**0.01**	**0.03**	0.20	0.39	**0.01**	**0.04**	0.20	0.41
Consumption	0.07	0.24	0.42	0.46	**0.02**	0.33	0.56	0.66	**0.01**	0.31	0.58	0.71
Investment	0.62	0.41	0.41	0.44	0.48	0.10	**0.04**	0.08	0.63	0.23	0.07	0.07
Inflation	**0.01**	**0.01**	**0.02**	**0.03**	**0.03**	**0.00**	**0.02**	**0.05**	**0.01**	**0.00**	**0.02**	**0.05**
Employment	0.16	0.19	0.45	0.49	0.91	0.77	0.36	0.54	0.92	0.80	0.37	0.52
Real wages	*0.99*	*1.00*	*0.99*	*1.00*	**0.03**	**0.00**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**
Unemployment	0.25	0.20	0.35	0.44	0.37	0.42	0.54	0.63	0.27	0.36	0.51	0.63
Interest rate	*0.95*	0.89	0.88	0.88	*0.98*	*0.98*	0.94	0.88	*0.99*	*0.98*	*0.96*	0.91

Note: The modified Diebold-Mariano test has been calculated as in Harvey, Leybourne, and Newbold (1997, equation 9) for the squared forecast errors of a GSW model based case where the SPF is used as conditioning information (noise model with 1-year-ahead SPF; news model with 1-year-ahead SPF; news model with 1 and 2-year-ahead SPF) versus the squared forecast errors of the benchmark GSW model. Percentile values taken from the Student's t-distribution with $N_h - 1$ degrees of freedom are shown above, with N_h being the number of h-step-ahead forecast errors, $N_h = 36 - h$. Small percentile values favor models that include the SPF as conditioning data, and large percentile values favour the DSGE model without this data. Bold-faced numbers refer the percentile values less than or equal to 5 percent, and numbers in italics to percentile values greater than or equal to 95 percent.

TABLE 5: Modified Diebold-Mariano tests of structural and non-structural models versus the BVAR model (RTDB vintages 2001Q1–2010Q4).

	Benchmark GSW				Noise				News 1-year				News 1 & 2-year			
Variable	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Output	0.23	0.36	0.49	0.65	0.13	0.17	0.39	0.73	**0.04**	0.09	0.18	0.67	**0.04**	0.09	0.18	0.72
Consumption	*1.00*	*1.00*	*0.99*	*0.99*	*0.98*	*1.00*	*1.00*	*1.00*	*0.99*	*1.00*	*0.99*	*0.99*	*0.99*	*1.00*	*0.99*	*0.98*
Investment	0.21	0.27	0.32	0.39	0.21	0.17	0.15	0.16	0.19	0.20	0.21	0.27	0.22	0.22	0.22	0.26
Inflation	0.68	0.26	0.12	0.10	**0.03**	**0.05**	**0.02**	**0.04**	0.07	0.06	**0.02**	**0.02**	**0.04**	**0.04**	**0.02**	**0.02**
Employment	0.89	0.88	0.89	0.94	0.86	0.86	0.85	0.86	0.93	0.88	0.86	0.86	0.94	0.88	0.86	0.85
Real wages	*0.98*	*1.00*	*1.00*	*1.00*	*0.99*	*1.00*	*1.00*	*1.00*	0.76	0.94	*0.98*	*1.00*	0.59	0.71	0.65	0.80
Unemployment	0.34	0.48	0.60	0.59	0.06	0.20	0.52	0.59	0.20	0.33	0.87	0.91	0.21	0.29	0.81	0.91
Interest rate	0.10	0.15	0.21	0.16	0.24	0.21	0.27	0.20	0.64	0.36	0.33	0.21	0.63	0.36	0.33	0.20
	AR(1)				RW				Mean							
Output	0.89	0.87	0.83	0.81	0.92	0.92	0.89	0.89	0.90	0.85	0.81	0.76				
Consumption	*0.97*	0.84	0.72	0.75	0.94	0.77	0.49	0.58	*0.96*	0.90	0.85	0.83				
Investment	0.81	0.77	0.75	0.73	0.83	0.80	0.79	0.79	0.91	0.84	0.78	0.72				
Inflation	0.20	0.21	0.08	0.06	0.24	0.24	0.11	0.08	0.28	0.10	**0.04**	**0.04**				
Employment	0.94	0.92	0.90	0.90	*0.95*	0.93	0.92	0.92	0.93	0.88	0.84	0.81				
Real wages	**0.01**	**0.05**	0.09	0.16	**0.03**	0.09	0.12	0.19	**0.01**	**0.03**	0.08	0.15				
Unemployment	0.82	0.84	0.86	0.92	0.81	0.84	0.87	0.93	*1.00*	*0.99*	*0.98*	*0.97*				
Interest rate	0.65	0.39	0.32	0.16	0.61	0.38	0.32	0.17	*1.00*	*1.00*	*0.97*	0.77				

Note: The modified Diebold-Mariano test has been calculated as in Harvey, Leybourne, and Newbold (1997, equation 9) for the squared forecast errors of the model displayed in the header of each column group versus the squared forecast errors of the BVAR model. Percentile values taken from the Student's t-distribution with $N_h - 1$ degrees of freedom are shown above, with N_h being the number of h-step-ahead forecast errors, $N_h = 36 - h$. Small percentile values favor the model displayed in the header, and large percentile values favour the BVAR model. Bold-faced numbers refer the percentile values less than or equal to 5 percent, and numbers in italics to percentile values greater than or equal to 95 percent.

TABLE 6: Forecast accuracy regressions for selected variables: GSW model with and without SPF and BVAR (RTDB vintages 2001Q1–2010Q4).

	Benchmark GSW				News 1-year				BVAR			
Forecast	1	2	3	4	1	2	3	4	1	2	3	4
Output												
Slope	1.17	1.26	1.25	0.67	1.11	1.30	1.90	1.57	1.05	1.07	1.59	1.26
	(0.09)	(0.24)	(0.54)	(0.71)	(0.08)	(0.20)	(0.50)	(1.19)	(0.10)	(0.24)	(0.67)	(1.56)
Intercept	−0.35	−0.83	−1.26	−0.42	−0.27	−0.88	−2.36	−2.26	−0.23	−0.53	−1.68	−1.28
	(0.18)	(0.45)	(1.06)	(1.52)	(0.17)	(0.38)	(0.94)	(2.47)	(0.21)	(0.47)	(1.19)	(2.77)
Adj. R^2	0.08	0.30	0.56	0.77	0.07	0.23	0.39	0.71	0.11	0.36	0.57	0.79
Inflation												
Slope	0.54	0.43	−0.27	−0.70	0.64	0.56	0.01	−0.60	0.48	0.24	−0.11	−0.13
	(0.16)	(0.24)	(0.35)	(0.51)	(0.17)	(0.27)	(0.40)	(0.57)	(0.14)	(0.34)	(0.86)	(1.27)
Intercept	1.07	1.31	2.45	3.15	0.87	1.04	2.01	3.13	1.15	1.59	2.22	2.26
	(0.28)	(0.41)	(0.57)	(0.83)	(0.31)	(0.49)	(0.72)	(1.06)	(0.28)	(0.36)	(0.33)	(0.31)
Adj. R^2	−0.01	0.00	0.01	0.00	−0.01	−0.00	0.01	0.00	−0.00	0.00	0.01	0.01
Unemployment												
Slope	0.84	0.83	0.74	0.53	0.83	0.78	0.68	0.50	0.79	0.75	0.69	0.59
	(0.06)	(0.13)	(0.23)	(0.34)	(0.06)	(0.10)	(0.17)	(0.24)	(0.05)	(0.09)	(0.15)	(0.21)
Intercept	1.31	1.40	2.19	3.95	1.44	1.90	2.78	4.26	1.77	2.20	2.66	3.56
	(0.53)	(1.07)	(1.94)	(2.83)	(0.49)	(0.85)	(1.41)	(1.97)	(0.45)	(0.75)	(1.21)	(1.70)
Adj. R^2	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03				
Interest rate												
Slope	0.93	0.80	0.56	0.24	0.98	0.80	0.50	0.22	0.81	0.64	0.51	0.28
	(0.06)	(0.15)	(0.29)	(0.38)	(0.08)	(0.19)	(0.30)	(0.37)	(0.05)	(0.09)	(0.17)	(0.21)
Intercept	0.13	0.40	1.07	2.13	−0.10	0.32	1.19	2.13	0.42	0.80	1.04	1.86
	(0.19)	(0.53)	(1.04)	(1.40)	(0.26)	(0.67)	(1.12)	(1.44)	(0.16)	(0.35)	(0.67)	(0.90)
Adj. R^2	−0.02	0.02	0.06	0.08	−0.01	0.03	0.06	0.08	−0.02	0.00	0.04	0.07

FIGURE 1: First release and annual revision data for real GDP growth (Δy_t), GDP deflator inflation ($\pi_{y,t}$), and the unemployment rate (u_t), 2000Q4–2010Q4.

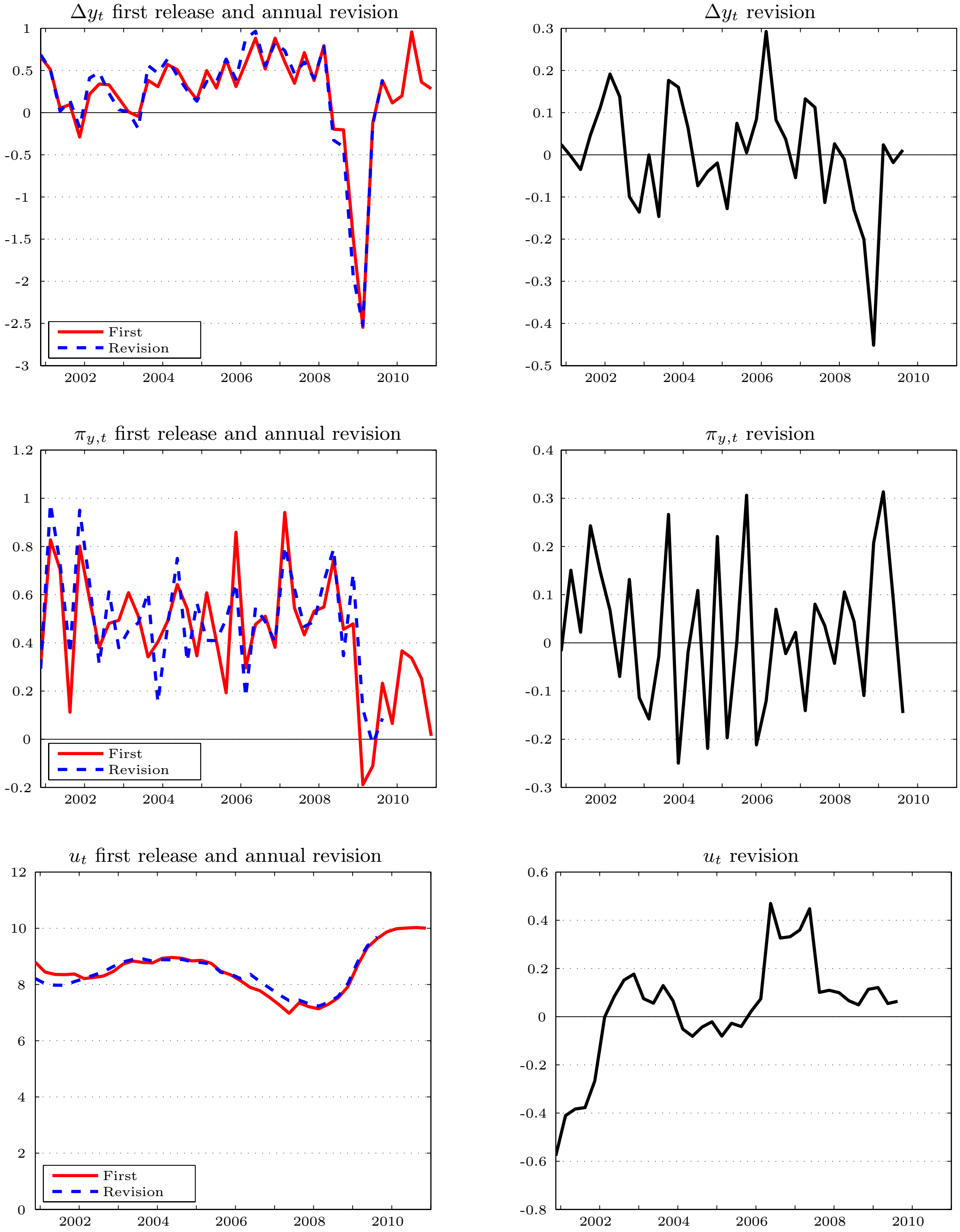

FIGURE 2: Relative RMSEs when conditioning on SPF data for DSGE models compared with RMSEs for the DSGE model without the SPF. The calculations are based on the RTDB vintages 2001Q1–2010Q4.

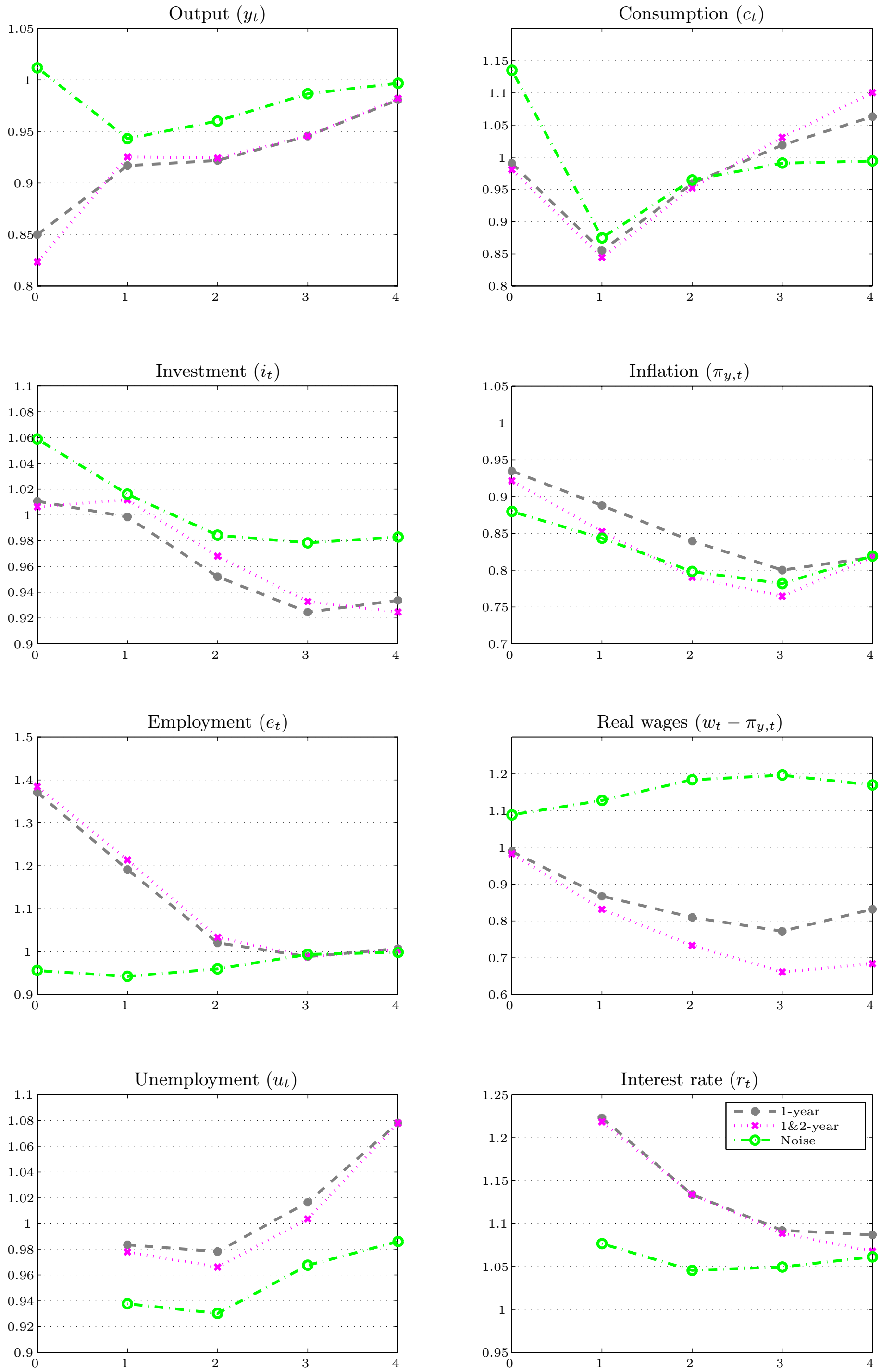

FIGURE 3: Relative RMSEs of structural and non-structural model compared with RMSEs for the BVAR model. The calculations are based on the RTDB vintages 2001Q1–2010Q4.

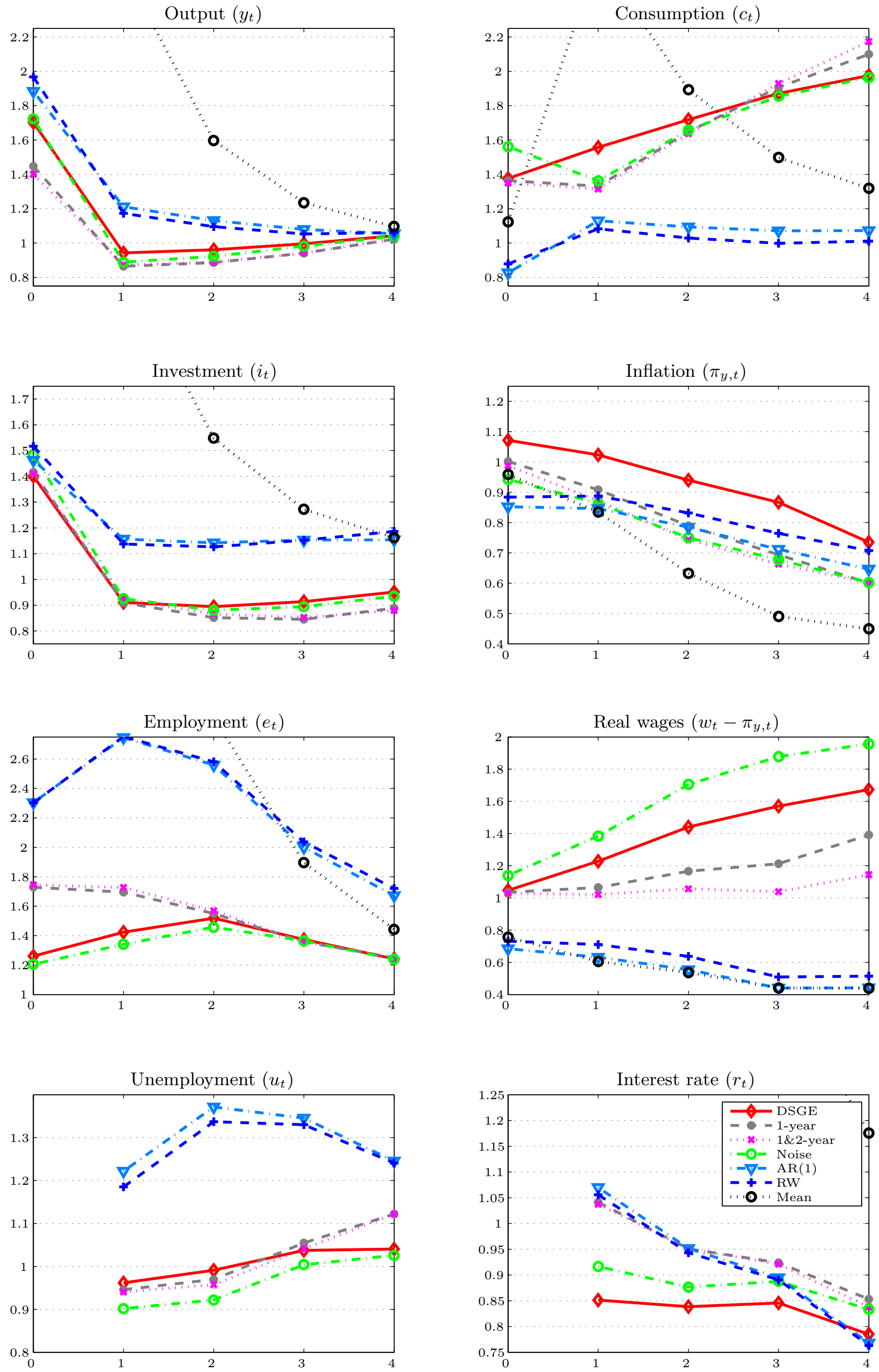

FIGURE 4: Multivariate MSE statistics for the RTDB vintages 2001Q1–2010Q4.