Estimating a medium–scale DSGE model with expectations based on small forecasting models

Sergey Slobodyan (CERGE–EI) and Raf Wouters (NBB)

February 28, 2009 [Preliminary version]

Abstract

In this paper we evaluate the empirical performance of a medium–scale DSGE model (Smets and Wouters 2007) when agents form expectations about forward variables by using small forecasting models. Agents learn about these simple AR and VAR forecasting models through Kalman filter estimation and they combine them either using a prediction based weighting scheme or fixed weights. The results indicate that a model, in which agents use a mixture of simple forecasting models to form expectations, does fit the data better than the full rational expectations model. Adaptive learning leads to substantial time variation in the coefficients of the forecasting models. Especially the beliefs about the dynamics of the inflation process turn out to be very important for the overall performance of the model. Agents' beliefs about the persistence of inflation display a peak the late seventies, and follow a clear downward trend starting during the Volcker disinflation period. This pattern in beliefs, which is in line with other recent evidence in the literature on inflation persistence, implies that the response of inflation to the various shocks declined significantly over the last 25 years. In this way, adaptive learning about inflation persistence also explains the observed decline in both the mean and the volatility of inflation as well as the flattening of the Phillips curve. Allowing for learning about inflation dynamics also results in lower estimates for the persistence of the exogenous processes that drive price and wage dynamics in the Rational Expectation version. We also find that the implicit beliefs of agents based on small forecasting models are more closely related to the survey evidence on inflation expectations than the beliefs under rational expectations.

JEL codes: C11, D84, E30, E52

1 Introduction

Most empirical DSGE models retain the hypothesis of Rational Expectations in the sense that expectations of agents are model consistent. Smets and Wouters (2003-2007) have shown that these models, when equipped with a rich set of frictions and a general stochastic structure, explain the data relatively well. It remains however somewhat problematic that these models require highly persistent exogenous shocks to explain the observed persistence in the data. Milani (2004) and Orphanides & Williams (2003) claim that learning can significantly influence the macroeconomic dynamics and increase the persistence in the model. For instance, Milani estimates a small scale model both under RE and learning and shows that the learning reduces the scale of structural frictions and results in an improved marginal likelihood relative to the RE model. Orphanides and Williams (2005) illustrate how adaptive learning can lead to inflation persistence. Slobodyan & Wouters (2007) analyse the learning dynamics in the SW model and found that learning hardly influences model dynamics if the information set used in the learning process is close to the rational expectations. Restricting information available to the agents may improve the model fit and better match the IRFs with those from the best–fitting DSGE–VAR models. We explore this issue further in this paper by assuming that agents form expectations about forward variables by using small forecasting models.

We follow Evans & Honkapohja (2001), Milani and Orphanides & Williams by assuming that economic agents do not have perfect knowledge of the reduced form parameters of the model when forming expectations about the future. Expecting the agents to take too many variables into account is unrealistic. Therefore, we study what happens if forecasts are based on small models, much smaller than those implied by the RE solution. One can never be sure that a particular model is the best. Therefore, we allow the agents to run a set of forecasting models and create combined forecasts taking past performance into account, using Bayesian Model Averaging techniques. Agents forecast future values of the lead variables with a linear function in the endogenous model variables. Agents learn about these simple AR and VAR forecasting models through Kalman filter estimation and they combine them either using a prediction based weighting scheme or fixed weights. Sargent and Williams (2005) showed that even if Kalman filter and constant gain learning are asymptotically equivalent on average, their transitory behavior may differ a lot. In particular, Kalman filter tends to result in much faster adjustment of agents' beliefs. With faster adjustment of beliefs, we are able to understand better whether the initial beliefs or time–varying coefficients matter more for the improved model fit.

The results indicate that a model, in which agents use a mixture of simple forecasting models to form expectations, does fit the data better than the full rational expectations model. Equal weight model averaging tends to generate an aggregate forecasting model that is on par or better than the best individual forecasting model. Marginal likelihood of the DSGE model is better as well when model weights are independent of past forecasting performance.

Relative to the DSGE model under rational expectations, models with learning are estimated to have lower persistence in the exogenous shocks, especially in the price and wage markup shocks, somewhat lower indexation for wages and lower investment adjustment cost, but more habits and a higher interest rate smoothing parameter in the policy rule.

Adaptive learning leads to substantial time variation in the coefficients of the forecasting models. Especially the beliefs about the dynamics of the inflation process turn out to be very important for the overall performance of the model. Agents' beliefs about the mean and the persistence of inflation display a peak in the late seventies, and follow a clear downward trend starting during the Volcker disinflation period. This pattern in beliefs, which is in line with other recent evidence in the literature on inflation persistence, implies that the response of inflation to the various shocks declined significantly over time. In this way, adaptive learning about inflation persistence explains the observed rise and decline in both the mean and the volatility of inflation over the last forty years. We also find that the implicit beliefs of agents based on small forecasting models are more closely related to the survey evidence on inflation expectations than the beliefs under rational expectations.

In the next section, we present the medium-scale DSGE model which is similar to Smets and Wouters (2007) except for the definition of the output gap. In section 3, we discuss the setup of the learning process: private agent form their expectations based on a combination of small forecasting models which are updated over time using the Kalman filter, and combined with Bayesian or simple averaging methods. We also discuss how initial beliefs are selected. The estimation outcomes for the model with learning are discussed in section 4. The time variation that is introduced by the learning dynamics is analysed more in detail in section 5. We illustrate the dynamics in the beliefs, and their consequences for the impulse responses and the overall variance of the model.

2 Model

In this paper, we evaluate the potential role of adaptive learning dynamics in an estimated medium-scale DSGE model. The model that we consider in this application is the one estimated in Smets and Wouters (2007) applied to the US economy over the period 1966-2005. This DSGE model contains many frictions that affect both nominal and real decisions of households and firms. The model is based on CEE (2005) and Smets and Wouters (2003). As in Smets and Wouters (2005), we extend the model so that it is consistent with a balanced steady state growth path driven by deterministic labour-augmenting technological progress. Households maximise a non-separable utility function with two arguments (goods and labour effort) over an infinite life horizon. Consumption appears in the utility function relative to a time-varying external habit variable. Labour is differentiated by a union, so that there is some monopoly power over wages, which results in an explicit wage equation and allows for the introduction of sticky nominal wages à la Calvo. Households rent capital services to firms and decide how much capital to accumulate given the capital adjustment costs they face. As the rental price of capital changes, the utilisation of the capital stock can be adjusted at increasing cost. Firms produce differentiated goods, decide on labour and capital inputs, and set prices, again according to the Calvo model. The Calvo model in both wage and price setting is augmented by the assumption that prices that are not re-optimised are partially indexed to past inflation rates. Prices are therefore set in function of current and expected marginal costs, but are also determined by the past inflation rate. The marginal costs depend on wages and the rental rate of capital. Similarly, wages depend on past and expected future wages and inflation. In both goods and labour markets we replace the standard Dixit-Stiglitz aggregator with an aggregator which allows for a time-varying demand elasticity which depends on the relative price as in Kimball (1995). As shown by Eichenbaum and Fischer (2007), the introduction of this real rigidity allows us to estimate a more reasonable degree of price and wage stickiness. The model also contains seven stochastic shocks to technology, preferences and policy behaviour. The number of structural shocks matches with the number of observables that are used in estimation.

Contrary to Smets and Wouters (2007), we assume in this paper that monetary policy does not react to the natural output level, which is defined as the output that would prevail in the flexible price and wage economy without distortionary price and wage shocks. Instead, we assume that monetary policy reacts to output relative to the underlying productivity process. By doing so, we do not need to model the flexible economy, which reduces considerably the number of forward variables appearing in the model. It is shown later on that the estimation results for this model under the rational expectations hypothesis are very similar to the original results in Smets and Wouters (2008). While at the same time, agents have to forecast fewer variables, which makes the learning process much more robust

2.1 The decision problems of firms and households and the equilibrium conditions

2.1.1 Final goods producers

The final good Y_t is a composite made of a continuum of intermediate goods $Y_t(i)$ as in Kimball (1995). The final good producers buy intermediate goods, package them into Y_t, and sell the final good to consumers, investors and the government in a perfectly competitive market. They maximize profits:

$$\max_{Y_t, Y_t(i)} P_t Y_t - \int_0^1 P_t(i) Y_t(i) di$$
$$s.t. \left[\int_0^1 G\left(\frac{Y_t(i)}{Y_t}; \epsilon_t^p \right) di \right] = 1$$

where P_t and $P_t(i)$ are the price of the final and intermediate goods respectively, and G is a strictly concave and increasing function characterised by $G(1) = 1$. ϵ_t^p is an exogenous process that reflects shocks to the aggregator function that result in changes in the elasticity of demand and therefore in the mark up. We will constrain $\epsilon_t^p \in (0, \infty)$.

Combining the first–order conditions with respect to $Y_t(i)$ and Y_t results in:

$$Y_t(i) = Y_t G'^{-1} \left[\frac{P_t(i)}{P_t} \int_0^1 G'\left(\frac{Y_t(i)}{Y_t} \right) \frac{Y_t(i)}{Y_t} di \right]$$

As in Kimball (1995), the assumptions on G imply that the demand for input $Y_t(i)$ is decreasing in its relative price, while the elasticity of demand is a positive function of the relative price (or a negative function of the relative output).

2.1.2 Intermediate goods producers

Intermediate good producer i uses the following technology:

$$Y_t(i) = \varepsilon_t^a \overline{K}_t(i)^\alpha \left[\gamma^t L_t(i) \right]^{1-\alpha} - \gamma^t \Phi \tag{1}$$

where $\overline{K}_t(i)$ is capital services used in production, $L_t(i)$ is a composite labour input and Φ is a fixed cost. γ^t represents the labour–augmenting deterministic growth rate in the economy and ε_t^a is total factor productivity.

The firm's profit is given by:

$$P_t(i) Y_t(i) - W_t L_t(i) - R_t^k \overline{K}_t(i).$$

where W_t is the aggregate nominal wage rate and R_t^k is the rental rate on capital.

Cost minimization yields the following first–order conditions:

$$(\partial L_t(i)) : \quad \Theta_t(i) \gamma^{(1-\alpha)t} (1-\alpha) \varepsilon_t^a \overline{K}_t(i)^\alpha L_t(i)^{-\alpha} = W_t \tag{2}$$
$$(\partial \overline{K}_t(i)) : \quad \Theta_t(i) \gamma^{(1-\alpha)t} \alpha \varepsilon_t^a \overline{K}_t(i)^{\alpha-1} L_t(i)^{1-\alpha} = R_t^k \tag{3}$$

where $\Theta_t(i)$ is the Lagrange multiplier associated with the production function and equals marginal cost MC_t.

Combining these FOCs and noting that the capital–labour ratio is equal across firms implies:

$$\overline{K}_t = \frac{\alpha}{1-\alpha} \frac{W_t}{R_t^k} L_t \tag{4}$$

The marginal cost MC_t is the same for all firms and equal to:

$$MC_t = \alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)} W_t^{1-\alpha} R_t^{k\,\alpha} \gamma^{-(1-\alpha)t} \left(\varepsilon_t^a\right)^{-1} \tag{5}$$

Under Calvo pricing with partial indexation to lagged inflation, the optimal price set by the firm that is allowed to re–optimise results from the following optimisation problem:

$$\max_{\widetilde{P}_t(i)} E_t \sum_{s=0}^{\infty} \xi_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \left[\widetilde{P}_t(i)(\Pi_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p}) - MC_{t+s}\right] Y_{t+s}(i)$$

$$s.t.\ Y_{t+s}(i) = Y_{t+s} G'^{-1}\left(\frac{P_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s}\right)$$

where $\widetilde{P}_t(i)$ is the newly set price, ξ_p is the Calvo probability of being allowed to optimise one's price, ι_p is the degree of indexation to lagged inflation, π_t is inflation defined as $\pi_t = P_t/P_{t-1}$, $[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}}]$ is the nominal discount factor for firms (which equals the discount factor for the households that are the final owners of the firms), $\tau_t = \int_0^1 G'\left(\frac{Y_t(i)}{Y_t}\right) \frac{Y_t(i)}{Y_t} di$ and

$$X_{t,s} = \left\{ \begin{array}{c} 1 \ for\ s = 0 \\ (\Pi_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p})\ for\ s = 1, ..., \infty \end{array} \right\}$$

The first-order condition is given by:

$$E_t \sum_{s=0}^{\infty} \xi_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} Y_{t+s}(i) \left[X_{t,s} \widetilde{P}_t(i) + \left(\widetilde{P}_t(i) X_{t,s} - MC_{t+s}\right) \frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})}\right] = 0 \tag{6}$$

where $x_t = G'^{-1}(z_t)$ and $z_t = \frac{P_t(i)}{P_t} \tau_t$.

The aggregate price index is in this case given by:

$$P_t = (1-\xi_p) P_t(i) G'^{-1}\left[\frac{P_t(i)\tau_t}{P_t}\right] + \xi_p \pi_{t-1}^{\iota_p} \pi_*^{1-\iota_p} P_{t-1} G'^{-1}\left[\frac{\pi_{t-1}^{\iota_p} \pi_*^{1-\iota_p} P_{t-1} \tau_t}{P_t}\right] \tag{7}$$

2.1.3 Households

Household j chooses consumption $C_t(j)$, hours worked $L_t(j)$, bonds $B_t(j)$, investment $I_t(j)$ and capital utilisation $Z_t(j)$, so as to maximise the following objective function:

$$E_t \sum_{s=0}^{\infty} \beta^s \left[\frac{1}{1-\sigma_c} (C_{t+s}(j) - \eta C_{t+s-1})^{1-\sigma_c}\right] \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_{t+s}(j)^{1+\sigma_l}\right)$$

subject to the budget constraint:

$$C_{t+s}(j) + I_{t+s}(j) + \frac{B_{t+s}(j)}{\varepsilon_t^b R_{t+s} P_{t+s}} - T_{t+s} \leq \tag{8}$$
$$\frac{B_{t+s-1}(j)}{P_{t+s}} + \frac{W_{t+s}^h(j) L_{t+s}(j)}{P_{t+s}} + \frac{R_{t+s}^k Z_{t+s}(j) K_{t+s-1}(j)}{P_{t+s}} - a(Z_{t+s}(j)) K_{t+s-1}(j) + \frac{Div_{t+s}}{P_{t+s}}$$

and the capital accumulation equation:

$$K_t(j) = (1-\delta) K_{t-1}(j) + \varepsilon_t^q \left[1 - S\left(\frac{I_t(j)}{I_{t-1}(j)} \right) \right] I_t(j) \tag{9}$$

There is external habit formation captured by the parameter η. The one–period bond is expressed on a discount basis. ε_t^b is an exogenous premium in the return to bonds, which might reflect inefficiencies in the financial sector leading to some premium on the deposit rate versus the risk free rate set by the central bank, or a risk premium that households require to hold the one period bond. δ is the depreciation rate, $S(\cdot)$ is the adjustment cost function, with $S(\gamma) = 0$, $S'(\gamma) = 0$, $S''(\cdot) > 0$, and ε_t^q is a stochastic shock to the price of investment relative to consumption goods. T_{t+s} are lump sum taxes or subsidies and Div_t are the dividends distributed by the intermediate goods producers and the labour unions.

Finally, households choose the utilisation rate of capital. The amount of effective capital that households can rent to the firms is:

$$\overline{K}_t(j) = Z_t(j) K_{t-1}(j) \tag{10}$$

The income from renting capital services is $R_t^k Z_t(j) K_{t-1}(j)$, while the cost of changing capital utilisation is $P_t a(Z_t(j)) K_{t-1}(j)$.

In equilibrium households will make the same choices for consumption, hours worked, bonds, investment and capital utilization. The first–order conditions can be written as (dropping the j index):

$$(\partial C_t) \qquad \Xi_t = \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} {L_t}^{1+\sigma_l} \right) (C_t - \eta C_{t-1})^{-\sigma_c} \tag{11}$$

$$(\partial L_t) \qquad \left[\frac{1}{1-\sigma_c} (C_t - \eta C_{t-1})^{1-\sigma_c} \right] \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l} \right) (\sigma_c - 1) L_t^{\sigma_l} = -\Xi_t \frac{W_t^h}{P_t} \tag{12}$$

$$(\partial B_t) \qquad \Xi_t = \beta \varepsilon_t^b R_t E_t \left[\frac{\Xi_{t+1}}{\pi_{t+1}} \right] \tag{13}$$

$$(\partial I_t) \qquad \Xi_t = \Xi_t^k \varepsilon_t^q \left(1 - S(\frac{I_t}{I_{t-1}}) - S'(\frac{I_t}{I_{t-1}})\frac{I_t}{I_{t-1}}\right) + \beta E_t \left[\Xi_{t+1}^k \varepsilon_{t+1}^q S'(\frac{I_{t+1}}{I_t})(\frac{I_{t+1}}{I_t})^2\right] \tag{14}$$

$$(\partial \bar{K}_t) \qquad \Xi_t^k = \beta E_t \left[\Xi_{t+1}\left(\frac{R_{t+1}^k}{P_{t+1}} Z_{t+1} - a(Z_{t+1})\right) + \Xi_{t+1}^k (1-\delta)\right] \tag{15}$$

$$(\partial u_t) \qquad \frac{R_t^k}{P_t} = a'(Z_t) \tag{16}$$

where Ξ_t and Ξ_t^k are the Lagrange multipliers associated with the budget and capital accumulation constraint respectively. Tobin's $Q_t = \Xi_t^k / \Xi_t$ and equals one in the absence of adjustment costs.

2.1.4 Intermediate labour unions and labour packers

Households supply their homogenous labour to an intermediate labour union which differentiates the labour services, sets wages subject to a Calvo scheme and offers those labour services to intermediate labour packers. Labour used by the intermediate goods producers L_t is a composite made of those differentiated labour services $L_t(i)$. As with intermediate goods, the aggregator is the one proposed by Kimball (1995). The labour packers buy the differentiated labour services, package L_t, and offer it to the intermediate goods producers.

The labour packers maximize profits:

$$\max_{L_t, L_t(i)} W_t L_t - \int_0^1 W_t(i) L_t(i) di$$
$$s.t. \left[\int_0^1 H\left(\frac{L_t(i)}{L_t}; \epsilon_t^w\right) di\right] = 1$$

where W_t and $W_t(i)$ are the price of the composite and intermediate labour services respectively, and H is a strictly concave and increasing function characterised by $H(1) = 1$. ϵ_t^w is an exogenous process that reflects shocks to the aggregator function that result in changes in the elasticity of demand and therefore in the mark up. We will constrain $\epsilon_t^w \in (0, \infty)$. Combining FOCs results in:

$$L_t(i) = L_t H'^{-1}\left[\frac{W_t(i)}{W_t} \int_0^1 H'\left(\frac{L_t(i)}{L_t}\right) \frac{L_t(i)}{L_t} di\right]$$

The labour unions are an intermediate between the households and the labor packers. Under Calvo pricing with partial indexation, the optimal wage set by the union that is allowed to re-optimise its wage results from the following optimisation problem:

$$\max_{\widetilde{W}_t(i)} E_t \sum_{s=0}^{\infty} \xi_w^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \left[\widetilde{W}_t(i) (\Pi_{l=1}^s \gamma \pi_{t+l-1}^{\iota_w} \pi_*{}^{1-\iota_w} - W_{t+s}^h\right] L_{t+s}(i)$$

$$s.t.\ L_{t+s}(i) = L_{t+s} H'^{-1}\left(\frac{W_t(i)X^w_{t,s}}{W_{t+s}}\tau^w_{t+s}\right)$$

where $\widetilde{W}_t(i)$ is the newly set wage, ξ_w is the Calvo probability of being allowed to optimise one's wage, $\tau^w_t = \int_0^1 H'\left(\frac{L_t(i)}{L_t}\right)\frac{L_t(i)}{L_t}di$ and

$$X^w_{t,s} = \left\{ \begin{array}{c} 1\ for\ s=0 \\ (\Pi^s_{l=1}\gamma\pi^{\iota_w}_{t+l-1}\pi^{1-\iota_w}_{*})\ for\ s=1,...,\infty \end{array} \right\}$$

The first-order condition is given by:

$$E_t\sum_{s=0}^{\infty}\xi^s_w\frac{\beta^s\Xi_{t+s}P_t}{\Xi_t P_{t+s}}L_{t+s}(i)\left[X^w_{t,s}\widetilde{W}_t(i) + \left(\widetilde{W}_t(i)X^w_{t,s} - W^h_{t+s}\right)\frac{1}{H'^{-1}(z^w_{t+s})}\frac{H'(x^w_{t+s})}{H''(x^w_{t+s})}\right] = 0 \tag{17}$$

where $x^w_t = H'^{-1}(z^w_t)$ and $z^w_t = \frac{W_t(i)}{W_t}\tau^w_t$.

The aggregate wage index is in this case given by:

$$W_t = (1-\xi_w)\widetilde{W}_t H'^{-1}\left[\frac{\widetilde{W}_t\tau^w_t}{W_t}\right] + \xi_w\gamma\pi^{\iota_w}_{t-1}\pi^{1-\iota_w}_{*}W_{t-1}H'^{-1}\left[\frac{\gamma\pi^{\iota_w}_{t-1}\pi^{1-\iota_w}_{*}W_{t-1}\tau^w_t}{W_t}\right] \tag{18}$$

The mark up of the aggregate wage over the wage received by the households is distributed to the households in the form of dividends (see the budget constraint of households).

2.1.5 Government Policies

The central bank follows a nominal interest rate rule by adjusting its instrument in response to deviations of inflation and output from their respective target levels:

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho}\left[\left(\frac{\pi_t}{\pi_*}\right)^{r_\pi}\left(\frac{Y_t}{Y^*_t}\right)^{r_y}\right]^{1-\rho}\left(\frac{Y_t/Y_{t-1}}{Y^*_t/Y^*_{t-1}}\right)^{r_{\Delta y}}\varepsilon^r_t \tag{19}$$

where R^* is the steady state nominal rate (gross rate), and ρ determines the degree of interest rate smoothing. ε^r_t is the exogenous monetary policy shock. Y^*_t is defined as potential output taking into account only the exogenous process for total factor productivity and the trend growth in the economy:

$$Y^*_t = \varepsilon^a_t\overline{K}^{\alpha}\left[\gamma^t\overline{L}\right]^{1-\alpha} - \gamma^t\Phi \tag{20}$$

This assumption deviates from the original Smets and Wouters (2007) where the natural output was considered in the reaction rule.

The government budget constraint is of the form

$$P_t G_t + B_{t-1} = T_t + \frac{B_t}{R_t} \tag{21}$$

where T_t are nominal lump–sum taxes (or subsidies) that also appear in household's budget constraint. Government spending is exogenous and expressed relative to the steady state output path as $\varepsilon_t^g = G_t/(Y\gamma^t)$.

2.1.6 Resource constraint

Integrating the budget constraint across households and combining with the government budget constraint and the expressions for the dividends of intermediate goods producers and labour unions gives the overall resource constraint:

$$C_t + I_t + G_t + a(Z_t)K_{t-1} = Y_t \tag{22}$$

2.2 Detrending and linearization

The model can be detrended with the deterministic trend γ and nominal variables can be replaced by their real counterparts. The non–linear system is then linearised around the stationary steady state of the detrended variables. Starred variables denote steady state values. We first describe the aggregate demand side of the model and then turn to the aggregate supply.

2.2.1 Aggregate demand side

The aggregate resource constraint is given by:

$$\widehat{y}_t \; = \; \widehat{g}_t \; + \; \frac{c_*}{y_*}\widehat{c}_t \; + \; \frac{i_*}{y_*}\widehat{i}_t \; + \; \frac{r_*^k k_*}{y_*}\widehat{u}_t. \tag{23}$$

Output ($\widehat{y}_t$) is absorbed by consumption ($\widehat{c}_t$), investment ($\widehat{i}_t$), capital-utilisation costs that are a function of the capital utilisation rate ($\widehat{u}_t$) and exogenous spending ($\widehat{g}_t$). We assume that exogenous spending follows a first–order autoregressive process with an IID–Normal error term and is also affected by the productivity shock as follows: $\widehat{g}_t = \rho_g \widehat{g}_{t-1} + \rho_{ga}\epsilon_t^a + \epsilon_t^g$. The latter is empirically motivated by the fact that in estimation exogenous spending also includes net exports, which may be affected by domestic productivity developments.

The dynamics of consumption follows from the consumption Euler equation and is given by:

$$\widehat{c}_t = \frac{1}{(1+(\eta/\gamma))} E_t\left[\widehat{c}_{t+1}\right] + \frac{(\eta/\gamma)}{(1+(\eta/\gamma))}\widehat{c}_{t-1}$$
$$-\frac{(1-\eta/\gamma)}{\sigma_c(1+(\eta/\gamma))}(\widehat{b}_t + \widehat{R}_t - E_t[\widehat{\pi}_{t+1}]) - \frac{(\sigma_c-1)(w_*^h L/c_*)}{\sigma_c(1+(\eta/\gamma))}(E_t\left[\widehat{L}_{t+1}\right] - \widehat{L}_t).$$

Current consumption ($\widehat{c}_t$) depends on a weighted average of past and expected future consumption, and on expected growth in hours worked ($E_t\left[\widehat{L}_{t+1}\right] - \widehat{L}_t$), the ex–ante real interest rate ($\widehat{R}_t - E_t[\widehat{\pi}_{t+1}]$) and a disturbance term $\widehat{b}_t$. This disturbance term represents a wedge between the interest rate controlled by the central bank and the return on assets held by the households. A positive shock to this wedge increases the required return on assets and reduces current consumption. At the same time, it also increases the cost of capital and reduces the value of capital and investment, as shown below. The disturbance is assumed to follow a first–order autoregressive process with an IID–Normal error term: $\widehat{b}_t = \rho_b \widehat{b}_{t-1} + \epsilon_t^b$.

The dynamics of investment comes from the investment Euler equation and is given by:

$$\widehat{i}_t = \frac{1}{(1+\overline{\beta}\gamma)}(\widehat{i}_{t-1} + (\overline{\beta}\gamma)\widehat{i}_{t+1} + \frac{1}{\gamma^2 S''}\widehat{Q}_t^k) + \widehat{q}_t, \tag{24}$$

where S'' is the steady–state elasticity of the capital adjustment cost function and $\overline{\beta} = (\beta/\gamma^{\sigma_c})$ where β is the discount factor applied by households. As in CEE (2005), a higher elasticity of the cost of adjusting capital reduces the sensitivity of investment ($\widehat{i}_t$) to the real value of the existing capital stock ($\widehat{Q}_t^k$). Modelling capital adjustment costs as a function of the change in investment rather than its level introduces additional dynamics in the investment equation, which is useful in capturing the hump–shaped response of investment to various shocks. Finally, $\widehat{q}_t$ represents a disturbance to the investment–specific technology process and is assumed to follow a first–order autoregressive process with an IID–Normal error term: $\widehat{q}_t = \rho_q \widehat{q}_{t-1} + \epsilon_t^\mu$.

The corresponding arbitrage equation for the value of capital is given by:

$$\widehat{Q}_t^k = -(\widehat{b}_t + \widehat{R}_t - E_t[\widehat{\pi}_{t+1}]) + \frac{r_*^k}{r_*^k + (1-\delta)} E_t[r_{t+1}^k] + \frac{(1-\delta)}{r_*^k + (1-\delta)} E_t[Q_{t+1}^k]. \tag{25}$$

The current value of the capital stock ($\widehat{Q}_t^k$) depends positively on its expected future value and the expected real rental rate on capital (r_{t+1}^k) and negatively on the ex-ante real interest rate and the risk premium disturbance.

2.2.2 Aggregate supply side

Turning to the supply side, the aggregate production function is given by:

$$\widehat{y}_t = \Phi(\alpha \widehat{k}_t + (1-\alpha)\widehat{L}_t + \widehat{A}_t) \quad (26)$$

Output is produced using capital ($\widehat{k}_t$) and labour services (hours worked, $\widehat{L}_t$). Total factor productivity ($\widehat{A}_t$) is assumed to follow a first-order autoregressive process: $\widehat{A}_t = \rho_a \widehat{A}_{t-1} + \epsilon^a_t$. The parameter α captures the share of capital in production and the parameter Φ is one plus the share of fixed costs in production, reflecting the presence of fixed costs in production.

As newly installed capital becomes only effective with a one–quarter lag, current capital services used in production are a function of capital installed in the previous period ($\widehat{\bar{k}}_{t-1}$) and the degree of capital utilisation ($\widehat{u}_t$):

$$\widehat{k}_t = \widehat{u}_t + \widehat{\bar{k}}_{t-1}.$$

Cost minimisation by the households that provide capital services implies that the degree of capital utilisation is a positive function of the rental rate of capital:

$$\widehat{u}_t = \frac{1-\psi}{\psi}\widehat{r}^k_t,$$

where ψ is a positive function of the elasticity of the capital utilisation adjustment cost function and normalized to be between zero and one. When $\psi = 1$, it is extremely costly to change the utilisation of capital and as a result the utilisation of capital remains constant. In contrast, when $\psi = 0$, the marginal cost of changing the utilisation of capital is constant and as a result in equilibrium the rental rate on capital is constant.

The accumulation of installed capital ($\widehat{\bar{k}}_t$) is not only a function of the flow of investment but also of the relative efficiency of these investment expenditures as captured by the investment–specific technology disturbance:

$$\widehat{\bar{k}}_t = (1 - \frac{i_*}{\bar{k}_*})\, \widehat{\bar{k}}_{t-1} + \frac{i_*}{\bar{k}_*}\widehat{i}_t + \frac{i_*}{\bar{k}_*}(1+\bar{\beta}\gamma)\gamma^2 S'' \widehat{q}_t.$$

Due to price stickiness as in Calvo (1983) and partial indexation to lagged inflation of those prices that can not be re–optimised as in Smets and Wouters (2003), prices adjust only sluggishly to their desired mark–up. Profit maximisation by price–setting firms gives rise to the following New–Keynesian Phillips curve:

$$\widehat{\pi}_t = \frac{1}{(1+\overline{\beta}\gamma\iota_p)}(\iota_p\widehat{\pi}_{t-1} + \overline{\beta}\gamma E_t\left[\widehat{\pi}_{t+1}\right] + \frac{1}{((\phi_p - 1)\varepsilon_p + 1)}\frac{(1-\xi_p\overline{\beta}\gamma)(1-\xi_p)}{\xi_p}(\widehat{mc}_t)) + \widehat{\lambda_{p,t}}$$

Inflation ($\widehat{\pi}_t$) depends positively on past and expected future inflation, negatively on the current price mark–up and positively on a price mark–up disturbance ($\widehat{\lambda_{p,t}}$). The price mark–up disturbance is assumed to follow an ARMA(1,1) process: $\widehat{\lambda_{p,t}} = \rho_p\widehat{\lambda_{p,t-1}} - \mu_p\epsilon_{p,t-1} + \epsilon_t^p$, where ϵ_t^p is an IID–Normal price mark–up shock. The inclusion of the MA term is designed to capture the high–frequency fluctuations in inflation.

When the degree of indexation to past inflation is zero ($\iota_p = 0$), equation (10) reverts to a standard purely forward–looking Phillips curve. The assumption that all prices are indexed to either lagged inflation or the steady state inflation rate ensures that the Phillips curve is vertical in the long run. The speed of adjustment to the desired mark–up depends among others on the degree of price stickiness (ξ_p), the curvature of the Kimball goods market aggregator (ε_p) and the steady–state mark–up, which in equilibrium is itself related to the share of fixed costs in production ($\phi - 1$) through a zero–profit condition. A higher ε_p slows down the speed of adjustment because it increases the strategic complementarity with other price setters. When all prices are flexible ($\xi_p = 0$) and the price mark–up shock is zero, the inflation equation reduces to the familiar condition that the price mark–up is constant or equivalently that there are no fluctuations in the wedge between the marginal product of labour and the real wage. The marginal cost is given by:

$$\widehat{mc}_t \;=\; (1-\alpha)\,\widehat{w}_t + \alpha\,\widehat{r}_t^k - \widehat{A}_t$$

Cost minimisation by firms will also imply that the rental rate of capital is negatively related to the capital–labour ratio and positively to the real wage (both with unitary elasticity):

$$\widehat{k}_t \;=\; \widehat{w}_t \;-\; \widehat{r}_t^k \;+\; \widehat{L}_t. \tag{27}$$

Similarly, due to nominal wage stickiness and partial indexation of wages to inflation, real wages only adjust gradually to the desired wage mark–up:

$$\begin{aligned}\widehat{w}_t \;=&\; \frac{1}{(1+\overline{\beta}\gamma)}(\widehat{w}_{t-1} + \overline{\beta}\gamma E_t\left[\widehat{w}_{t+1}\right] - (1+\overline{\beta}\gamma\iota_w)\widehat{\pi}_t + \iota_w\widehat{\pi}_{t-1} + \overline{\beta}\gamma E_t\left[\widehat{\pi}_{t+1}\right] \\ &+ \frac{(1-\xi_w\overline{\beta}\gamma)(1-\xi_w)}{\xi_w((\phi_w-1)\varepsilon_w+1)}[\frac{1}{1-\eta/\gamma}\widehat{c}_t - \frac{\eta/\gamma}{1-\eta/\gamma}\widehat{c}_{t-1} + \sigma_l\widehat{L}_t - \widehat{w}_t\,] + \widehat{\lambda_{w,t}}\end{aligned}$$

The real wage is a function of expected and past real wages, expected, current and past inflation, the wage mark up and a wage–mark up disturbance ($\widehat{\lambda_{w,t}}$). If wages are perfectly flexible ($\xi_w = 0$), the real wage is a constant mark–up over the marginal rate of substitution between consumption and leisure. In general, the speed of adjustment to the desired wage mark–up depends on the degree of wage stickiness (ξ_w) and the demand elasticity for labour, which itself is a function of the steady–state labour market mark–up ($\phi_w - 1$) and the curvature of the Kimball labour market aggregator (ε_w). When wage indexation is zero (ι_w), real wages do not depend on lagged inflation. The wage–mark up disturbance ($\widehat{\lambda_{w,t}}$) is assumed to follow an ARMA(1,1) process with an IID–Normal error term: $\widehat{\lambda_{w,t}} = \rho_w \widehat{\lambda_{w,t-1}} - \mu_w \epsilon_{w,t-1} + \epsilon_t^w$. As in the case of the price mark–up shock, the inclusion of an MA term allows us to pick up some of the high frequency fluctuations in wages.

Finally, the model is closed by adding the following empirical monetary policy reaction function:

$$\begin{aligned}\widehat{R}_t \;&=\; \rho_R \widehat{R}_{t-1} \;+\; (1-\rho_R)(r_\pi \widehat{\pi}_t \;+\; r_y \widehat{ygap}_t) \\ &\quad + r_{\Delta y}(\widehat{ygap}_t - \widehat{ygap}_{t-1}) \;+ r_t \end{aligned} \tag{28}$$

The monetary authorities follow a generalised Taylor rule by gradually adjusting the policy–controlled interest rate ($\widehat{R}_t$) in response to inflation and the output gap, defined as the difference between actual and potential output (Taylor, 1993). The output gap is given by $\widehat{ygap}_t = \Phi(\alpha \widehat{k}_t + (1-\alpha)\widehat{L}_t)$. The parameter ρ_R captures the degree of interest rate smoothing. In addition, there is also a short–run feedback from the change in the output gap. Finally, we assume that the monetary policy shocks (r_t) follows a first–order autoregressive process with an IID–Normal error term: $\widehat{r}_t \;= \rho_r \widehat{r}_{t-1} + \epsilon_t^r$.

Equations (1) to (13) determine thirteen endogenous variables: $\widehat{y}_t$, $\widehat{c}_t$, $\widehat{i}_t$, $\widehat{q}_t$, $\widehat{\bar{k}}$, $\widehat{k}_t$, $\widehat{u}_t$, $\widehat{r}_t^k$, $\widehat{mc}_t$, $\widehat{\pi}_t$, $\widehat{w}_t$, $\widehat{L}_t$ and $\widehat{R}_t$. The stochastic behaviour of the system of linear rational expectations equations is driven by seven exogenous processes and their respective disturbances: total factor productivity (A_t, ϵ_t^a), investment–specific technology (q_t, ϵ_t^q), risk premium (b_t, ϵ_t^b), exogenous spending (g_t, ϵ_t^g), price mark–up ($\lambda_t^p, \epsilon_t^p$), wage mark–up ($\lambda_t^w, \epsilon_t^w$) and monetary policy ($r_t, \epsilon_t^r$) shocks. Together with the two lagged innovations entering the ARMA processes, the model contains 29 variables, of which 7 enter with a lead term[1]. Next we turn to the estimation of the model.

[1] The original model, that includes the modelling of the natural output, contains 40 variables of which 12 appear as forward variables.

2.3 Estimation under Rational Expectations

The model presented in the previous section is estimated in Smets and Wouters (2007) under the hypothesis that agents have rational expectations. It was shown that these models, when equipped with a rich set op frictions and a general stochastic structure, are able to explain the data relatively well and these model have a forecasting performance that is comparable or even better than purely statistical VAR or BVAR models.

2.3.1 Measurement equations

The model is estimated using seven key macro–economic quarterly US time series as observable variables: the log difference of real GDP, real consumption, real investment and the real wage, log hours worked, the log difference of the GDP deflator and the federal funds rate. A full description of the data used is given in the appendix. The corresponding measurement equation is:

$$O_t = \begin{bmatrix} dlGDP_t \\ dlCons_t \\ dlINV_t \\ dlWag_t \\ lHOURS_t \\ dlP_t \\ FEDFUNDS_t \end{bmatrix} = \begin{bmatrix} \overline{\gamma} \\ \overline{\gamma} \\ \overline{\gamma} \\ \overline{\gamma} \\ \overline{l} \\ \overline{\pi} \\ \overline{r} \end{bmatrix} + \begin{bmatrix} \widehat{y}_t - \widehat{y}_{t-1} \\ \widehat{c}_t - \widehat{c}_{t-1} \\ \widehat{i}_t - \widehat{i}_{t-1} \\ \widehat{w}_t - \widehat{w}_{t-1} \\ \widehat{l}_t \\ \widehat{\pi}_t \\ \widehat{R}_t \end{bmatrix}, \qquad (29)$$

where l and dl stand for log and log difference respectively, $\overline{\gamma} = 100(\gamma - 1)$ is the common quarterly trend growth rate to real GDP, consumption, investment and wages,$\overline{\pi} = 100(\Pi_* - 1)$ is the quarterly steady–state inflation rate and is $\overline{r} = 100(\gamma^{\sigma_c}\Pi_*/\beta - 1)$ the steady–state nominal interest rate. Given the estimates of the trend growth rate and the steady–state inflation rate, the latter will be determined by the estimated discount rate. Finally, $\overline{l}$ is steady–state hours–worked. The model is estimated over the full sample period from 1966:1 till 2004:4.

The estimations are executed using Bayesian estimation methods. First, we estimate the mode of the posterior distribution by maximising the log posterior function, which combines the prior information on the parameters with the likelihood of the data. In a second step, the Metropolis–Hastings algorithm is used to get a complete picture of the posterior distribution and to evaluate the marginal likelihood of the model.

2.3.2 Prior distribution of the parameters

The priors on the stochastic processes are harmonised as much as possible. The standard errors of the innovations are assumed to follow an inverse gamma distribution with a mean of 0.10 and two degrees of freedom, which corresponds to a rather loose prior. The persistence of the AR(1) processes is beta distributed with mean 0.5 and standard deviation 0.2. A similar distribution is assumed for the MA parameter in the process for the price and wage mark–up. The quarterly trend growth rate is assumed to be Normal distributed with mean 0.4 (quarterly growth rate) and standard deviation 0.1. The steady–state inflation rate and the discount rate are assumed to follow a gamma distribution with a mean of 2.5% and 1% on an annual basis.

Five parameters are fixed in the estimation procedure. The depreciation rate is fixed at 0.025 (on a quarterly basis) and the exogenous spending–GDP ratio is set at 18%. Both of these parameters would be difficult to estimate unless the investment and exogenous spending ratios would be directly used in the measurement equation. Three other parameters are clearly not identified: the steady–state markup in the labour market (λ_w), which is set at 1.5, and the curvature parameters of the Kimball aggregators in the goods and labour market (ε_p and ε_w), which are both set at 10.

The parameters describing the monetary policy rule are based on a standard Taylor rule: the long run reaction on inflation and the output gap are described by a Normal distribution with mean 1.5 and 0.125 (0.5 divided by 4) and standard errors 0.25 and 0.05 respectively. The persistence of the policy rule is determined by the coefficient on the lagged interest rate rate which is assumed to be Normal around a mean of 0.75 with a standard error of 0.1. The prior on the short run reaction coefficient to the change in the output–gap is 0.125.

The parameters of the utility function are assumed to be distributed as follows. The intertemporal elasticity of substitution is set at 1.5 with a standard error of 0.375; the habit parameter is assumed to fluctuate around 0.7 with a standard error of 0.1 and the elasticity of labour supply is assumed to be around 2 with a standard error of 0.75. These are all quite standard calibrations. The prior on the adjustment cost parameter for investment is set around 4 with a standard error of 1.5 (based on CEE, 2005) and the capacity utilisation elasticity is set at 0.5 with a standard error of 0.15. The share of fixed costs in the production function is assumed to have a prior mean of 0.25. Finally, there are the parameters describing the price and wage setting. The Calvo probabilities are assumed to be around 0.5 for both prices and wages, suggesting an average length of price and wage contracts of half a year. This is compatible with the findings of Bils and Klenow (2004) for prices. The prior mean of the degree of indexation to past inflation is also set at 0.5 in both goods and labour markets.

2.3.3 Posterior estimates of the parameters

In the DSGE model under RE, the trend growth rate is estimated to be around 0.43, which is somewhat smaller than the average growth rate of output per capita over the sample. The posterior mean of the steady state inflation rate over the full sample is about 3% on an annual basis. The mean of the discount rate is estimated to be quite small (0.65% on an annual basis). The implied mean steady state nominal and real interest rates are respectively about 6 % and 3% on an annual basis.

A number of observations are worth making regarding the estimated processes for the exogenous shock variables. Overall, the data appears to be very informative about the stochastic processes for the exogenous disturbances. The productivity, the government spending and the wage mark–up processes are estimated to be the most persistent with an AR(1) coefficient of respectively 0.96, 0.98 and 0.97. The high persistence of the productivity and wage mark–up processes implies that at long horizons most of the forecast error variance of the real variables will be explained by those two shocks. In contrast, both the persistence and the standard deviation of the risk premium and monetary policy shock are relatively low (respectively 0.18 and 0.13).

Turning to the estimates of the main behavioural parameters, we see that in the DSGE model the mean of the posterior distribution is typically relatively close to the mean of the prior assumptions. There are a few notable exceptions. Both the degree of price and wage stickiness are estimated to be quite a bit higher than 0.5. The average duration of wage contracts is somewhat less than a year; whereas the average duration of price contracts is about 3 quarters. The mean of the degree of price indexation (0.23) is on the other hand estimated to be much less then 0.5. Also the elasticity of the cost of changing investment is estimated to be higher than assumed a priori, suggesting an even slower response of investment to changes in the value of capital. Finally, the posterior mean of the fixed cost parameter is estimated to be much higher than assumed in the prior distribution (1.62) and the share of capital in production is estimated to be much lower (0.19). Overall, it appears that the data is quite informative on the behavioural parameters as indicated by the lower variance of the posterior distribution relative to the prior distribution. Two exceptions are the elasticity of labour supply and the elasticity of the cost of changing the utilisation of capital, where the posterior and prior distributions are quite similar.

Finally, turning to the monetary policy reaction function parameters, the mean of the long-run reaction coefficient to inflation is estimated to be relatively high (2.03). There is a considerable degree of interest rate smoothing as the mean of the coefficient on the lagged interest rate is estimated to be 0.82. Policy does not appear to react very strongly to the output gap level (0.09), but does respond strongly to changes in the output gap (0.22) in the short run.

3 Kalman Filter Learning with Small Forecasting Models

Sargent and Williams (2005) showed that even if Kalman filter and constant gain Recursive Least Squares (RLS) learning are asymptotically equivalent on average, their transitory behavior may differ significantly. In particular, Kalman filter tends to result in much faster adjustment of agents' beliefs. The relative contribution to the improved fit of the initial non–rational beliefs versus time–varying coefficients generated by the adaptive learning was a major question arising from Slobodyan and Wouters (2007). Faster adjustment of beliefs allows us to move further in understanding of this particular issue.

Expecting the agents to take too many variables into account is unrealistic if one considers the adaptive learning setup as a description of real behavior by the agents. Therefore, we study what happens if forecasts are based on small models, much smaller than those implied by the REE solution.

Finally, one can never be sure that a particular model is the best. Therefore, we allow the agents to run a set of forecasting models and create combined forecasts taking past performance into account, using Bayesian Model Averaging techniques.

3.1 Kalman Filter Setup

We implement the adaptive learning within the DYNARE 3.064 MATLAB toolbox which is used to estimate and simulate DSGE models. The model is driven by the exogenous driving process w_t, which is an AR(1) process

$$w_t = \Gamma w_{t-1} + \Pi \epsilon_t. \tag{30}$$

DYNARE represents our model in the following way:

$$A_0 \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + A_1 \begin{bmatrix} y_t \\ w_t \end{bmatrix} + A_2 E_t y_{t+1} + B_0 \epsilon_t = 0, \tag{31}$$

where the vector y_t includes endogenous variables of the model.[2] The solution of the model is provided by DYNARE as

$$\begin{bmatrix} y_t \\ w_t \end{bmatrix} = \mu + T \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + R \epsilon_t. \tag{32}$$

The vector y contains state variables y^s (those appearing with a lag), forward variables y^f that appear with a lead in the model, and the so–called static variables.[3] Deviating from the rational equilibrium (RE) assumption and following Marcet and Sargent (1989) and Evans and Honkapohja (2001), we assume that the agents imagine the values of the lead variables to be a linear function of the endogenous

[2] DYNARE variable `jacobia_` contains the matrix $\begin{bmatrix} A0 & A1 & A2 & B \end{bmatrix}$.

[3] y^f and y^s could intersect.

model variables,[4],[5]

$$y^f_j = X_j \beta_j + u_j. \tag{33}$$

The agents then use the linear model (33) for forecasting, with forecasts given as

$$y^f_{j,t} = X_{j,t-1}\beta_{j,t-1} + u_{j,t}.$$

In contrast with Smets and Wouters (2007), we allow only past endogenous variables y_{t-1} in the data matrix X_{t-1}, plus constants. Thus, agents cannot access values of exogenous processes w_t. In general, every forward–looking variable is predicted using its own set of right–hand variables. We keep these forecasting models small, with not more than four variables and a constant on the right–hand side of any particular equation.[6]

The agents believe that the coefficients β (a vector obtained by stacking all β_j) follow a vector autoregressive process:

$$vec\left(\beta_t - \overline{\beta}\right) = F \cdot vec\left(\beta_{t-1} - \overline{\beta}\right) + v_t, \tag{34}$$

where F is a diagonal matrix with $\rho \leq 1$ on the main diagonal, and use Kalman filter to update their beliefs about β. Errors v_t are assumed to be *i.i.d.* with variance–covariance matrix V. In addition, we allow them to entertain a small set of forecasting models (five for the estimations reported here), and to combine the forecasts using weights that are either fixed or adjusted on–line using Bayes Information Criterion (BIC).

We can write the forecasting model (33) in the following SURE format:

$$\begin{bmatrix} y^f_{1t} \\ y^f_{2t} \\ \vdots \\ y^f_{mt} \end{bmatrix} = \begin{bmatrix} X_{1,t-1} & 0 & \dots & 0 \\ 0 & X_{2,t-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{m,t-1} \end{bmatrix} \begin{bmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \\ \vdots \\ \beta_{m,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{m,t} \end{bmatrix},$$

Data matrices X_j could contain different numbers of columns, making lengths of β_j, the vectors of coefficients in scalar forecasting model for a forward–looking variable y^f_j, also different. The errors $u_{j,t}$ are different linear combinations of

[4] In the adaptive learning literature, this equation is called the Perceived Law of Motion (PLM).

[5] This type of learning, promoted by Evans and Honkapohja (2001), is called *Euler equation* learning: the agents forecast only immediate future variables which are typically present in Euler equations of firms and/or consumers. An alternative description of learning — *long–horizon* learning — has been promoted recently by Bruce Preston: he considers agents forecasting economic variables (present in their budget constraint and exogenous to their decision–making) infinitely many periods ahead.

For a theoretical discussion on these two approaches to adaptive learning, see Preston (2005) and Honkapohja et al. (2002). For a discussion of effects of the learning type on the behavior of estimated DSGE model, see Milani (2006) and references therein.

[6] For the full list of variables included into X, see the end of Section 3.

true model errors ε_t and thus are likely to be correlated, making the variance–covariance matrix non–diagonal:

$$\Sigma = E\left[u_t \cdot u_t^T\right].$$

With the above notation, Kalman filter equations are given as

$$\begin{aligned}
\beta_{t|t} &= \beta_{t|t-1} + P_{t|t-1}X_{t-1}\left[\Sigma + X_{t-1}^T P_{t|t-1} X_{t-1}\right]^{-1} \times \left(y_t^f - \beta_{t|t-1}^T X_{t-1}\right), \quad (35)\\
& \quad \mathit{with}\ (\beta_{t+1|t} - \overline{\beta}) = F \cdot (\beta_{t|t} - \overline{\beta}).\\
P_{t|t} &= P_{t|t-1} - P_{t|t-1}X_{t-1}\left[\Sigma + X_{t-1}^T P_{t|t-1} X_{t-1}\right]^{-1} \times X_{t-1}^T P_{t|t-1},\\
& \quad \mathit{with}\ P_{t+1|t} = F \cdot P_{t|t} \cdot F^T + V.
\end{aligned}$$

Updating of the beliefs at any t depends on the data (best estimates of the state, the lead and the exogenous variables at time $t-1$) and on the initial beliefs. Best estimates are filtered values of the model variables taken from the likelihood Kalman filter. In principle, one could use smoothed rather than filtered estimates, re–smoothing every period and re–estimating past beliefs. This would represent a more consistent usage of available information, but is computationally very intensive and is not performed here.

3.2 Generating Initial Beliefs

In order to perform the iterations of the beliefs Kalman filter described in (35) we need to specify $\beta_{1|0}$, $P_{1|0}$, Σ, and V. All of them are derived based on the correlations between the model variables, implied by the rational expectations equilibrium for the currently evaluated parameter vector. In terms of Slobodyan and Wouters (2007), the initial beliefs are model consistent.

As is well known, the estimates in a standard problem with heteroscedastic errors are

$$\begin{aligned}
\widehat{\beta}_{GLS} &= \left(X^T\Sigma^{-1}X\right)^{-1} X^T\Sigma^{-1}y,\\
Var\left[\widehat{\beta}_{GLS}\right] &= \left(X^T\Sigma^{-1}X\right)^{-1}.
\end{aligned}$$

Using the fact that $\widehat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T y$ is unbiased, we use the theoretical moments under RE and set

$$\beta_{1|0} = E\left[X^T X\right]^{-1} \cdot E\left[X^T y\right].$$

Given $\beta_{1|0}$, we calculate Σ as

$$\Sigma = E\left[\left(y_t^f - X_{t-1}\beta_{1|0}\right)\left(y_t^f - X_{t-1}\beta_{1|0}\right)^T\right],$$

again using RE theoretical moments in the process. Finally, $P_{1|0}$, initial guess about the mean square forecast error of the state, and V, the assumed variance–covariance matrix of shocks v_t driving the state process, are both taken to be proportional to $Var\left[\widehat{\beta}_{GLS}\right]$:

$$P_{1|0} = \gamma \cdot \left(X^T \Sigma^{-1} X\right)^{-1}, \quad (36)$$

$$V = \sigma \cdot \left(X^T \Sigma^{-1} X\right)^{-1}. \quad (37)$$

Given knowledge of theoretical moments and of Σ, the matrix $\left(X^T\Sigma^{-1}X\right)^{-1}$ could be readily calculated.

This initialization leaves just three parameters, γ, σ, and ρ, to be estimated or calibrated. Following an approach of Sargent and Williams (2005), it is possible to show that for $\rho = 1$, if one can ignore the term $X_{t-1}^T P_{t-1|t-2} X_{t-1}$ relative to Σ in (35), then selecting

$$P_{1|0} = \gamma \cdot \left(X^T \Sigma^{-1} X\right)^{-1},$$

$$V = \gamma^2 \cdot \left(X^T \Sigma^{-1} X\right)^{-1},$$

leads to asymptotic (close to the steady state) equivalence of the mean dynamics of Kalman filter learning and of the constant gain RLS learning with the gain equal to γ. For $\rho \neq 1$, one could derive similar relationship between P_0 and V that makes the Kalman filter and the constant gain learning asymptotically equivalent. However, in the estimation step, we find that it is numerically more stable to assume (36-37) and to estimate (or calibrate) γ, σ, and ρ separately.

3.3 Beliefs and likelihood construction

In contrast to low–dimensional models studied by Milani (2005), Sargent, Williams, and Zha (2006), or Vilagi (2007), our set–up exhibits a clear distinction between the endogenous model variables and the observable variables which are used to estimate the model. Therefore, we use output from the Kalman filter, used to construct the likelihood function for a particular combination of parameters, on both sides of the updating equation (36-37).[7]

[7] In terms of Hamilton (1994), we use $\widehat{y}_{t-1|t-1}$ on the right and $\widehat{y}_{t|t}^f$ on the left. In principle, as time t progresses, the agents could revise their past filtered estimates and thus adjust values of ϕ_t

Given that our agents use very small forecasting models, we allow them to include constants even though both left– and right–hand side variables in the forecasting equations have zero means. This makes sense if they believe that certain variable could deviate from their unconditional means for a long time, with such belief clearly vindicated by the Great Inflation period.

We allow the agents to use several models at the same time, track their performance, and use a variant of the Bayesian Model Averaging (BMA) technique to produce an aggregate forecast that is used to inform their actions. In particular, for every forecasting model M_i, the agents track the value of

$$B_{i,t} = t \cdot \ln \det \left(\frac{1}{t} \sum_{i=1}^{t} u_i u_i^T \right) + \kappa_i \cdot \ln t,$$

where κ_i is number of degrees of freedom in forecasting model M_i, and u_i the i^{th} model forecasting errors. This expression is a generalization of the sum of squared errors adjusted for degrees of freedom using Bayesian Information Criterion (BIC) penalty.

Given values of $B_{i,t}$, the weight of a model i at time t is proportional to $\exp\left\{-\frac{1}{2}B_{i,t}\right\}$. For linear models with normal priors and normal errors, this procedure is asymptotically equivalent to weighting the models using posterior odds ratio. These weights are used to form the aggregate beliefs vector β_t^{aggr}. The models estimated using this way of generating time–varying forecasting model weights are denoted with 'BIC' in the remainder of the paper.

There are many alternative approaches to model averaging. In this paper, we also report estimations when the weights on all models are fixed at $\frac{1}{N}$, where N is the number of models used. These estimations are marked by 'EW'.

The beliefs generated in the beliefs Kalman filter step (35) and aggregated as described above are then used to generate expectations of forward–looking variables, $E_t y_{t+1}^f$, as a linear function of y_t^f. Plugging these expectations into (31), we solve the purely backward–looking equations to obtain a representation

$$\begin{bmatrix} y_t \\ w_t \end{bmatrix} = \mu_t + T_t \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + R_t \epsilon_t.$$

Thus obtained time–dependent matrices replace DYNARE–produced μ, R, and T, and are then used in the main Kalman filtering step used to calculate the model likelihood. This is the only major intervention needed to introduce adaptive learning into DYNARE (another one is generating initial beliefs, described earlier).

used in the past. In other words, in every period the agents would use the smoothed estimates of the model variables, and revise the whole sequence of beliefs held in past. This procedure would make a better use of the available information; however, our current procedure uses only filtered estimates.

During updating, the transition matrix T_t (derived as if a single forecasting model at a time were used) is restricted to the stable domain by a version of a projection facility: if the largest eigenvalue of T_t is outside of the unit circle, we retain last period β. A standard projection facility (checking roots of the forecasting VAR) cannot be implemented consistently, as the relationship between lead variables, forming the left–hand side of the PLM, and the right–hand side variables including state, forward, and even static variables depends on the solution of the model. On the other hand, T_t could be interpreted as the forecasting VAR for all model variables, including lead, state, and static.[8]

Slobodyan and Wouters (2007) found that such discontinuous adjustments of the beliefs lead to numerical problems during estimations, especially in the optimization step. We do observe discontinuities in the likelihood function related to the projection facility in this paper as well, but they seem to represent a much smaller problem. We can speculate that the averaging of several forecasting models reduces the importance of these discontinuities compared to Slobodyan and Wouters (2007).

The set of forecasting models used in all estimations that are reported in this paper, includes:[9]

1. AR(1): every forward–looking variable is predicted based on its own lagged value;
2. AR(1) + 2: in addition to own lag, lagged interest rate and inflation are added to the RHS of every forecasting equation;
3. AR(2): every forward–looking variable is predicted based on two own lags;
4. AR(1) + 1: in addition to own lag, inflation is added to the RHS of every forecasting equation;
5. AR(1) + 3: in addition to own lag, interest rate, inflation, and output are added to the RHS of every forecasting equation.

[8] Some small forecasting models used here *are* VARs. For this models, an unstable vector autoregression often means that the corresponding T_t is also unstable.

[9] All equations in all models include a constant, which is also assumed to behave according to (34).

4 Estimation results under learning with small models

4.1 Marginal likelihood of the model

Table 1 compares the outcomes for the marginal likelihood of the various models under learning with the results of the rational expectations model. Under rational expectations, the simplified Smets & Wouters (2007) model that we consider in this paper (using the TFP–based output gap concept) produces a slightly worse marginal likelihood of -926 vs. -922 for the original model (based on the natural output gap concept).

First of all, the table shows that the model under KF learning produces significantly higher marginal likelihoods than the model under RE. Within the models under KF learning, the versions where beliefs of the five forecasting models are combined with a constant and equal weight (EW), produces consistently the best results. Also, the setups of the Kalman filter in which the belief coefficients are allowed to follow an autoregressive process (ρ estimated) produces consistently better results than the setup where these coefficients were assumed to follow a random walk process. Fixing the gain and sigma parameters of the Kalman filter (fixed at the posterior mean, $\gamma = 0.031$ and $\sigma = 0.003$, which was characterised by a large uncertainty) do not impose any cost on the estimation outcomes.[10]

Allowing for time varying weights based on the BIC criterium as a measure of the past forecasting performance of the individual forecasting models, does not translate in an improvement of the overall model likelihood. Clearly, the additional time variation in the beliefs that is introduced via this selection process is not improving the overall fit. This result is probably not so surprising: the five small forecasting models produce relatively similar forecasts and in such a situation one can argue that a simple equal weight combination might be the more appropriate combination method for forecasts (see Van Dijk et al. (2007) for a similar conclusion on forecast combination methods). It would be interesting to see how this conclusion is changed if we consider alternative forecasting models that would

[10] In a first set of estimations, we used random walk coefficients model ($\rho = 1$) and estimated σ and γ. Consistently with findings of Sargent, Williams, and Zha (2006), we find that the estimated σ tends to be larger than γ^2. Unrestricted MCMC tends to generate parameter draws with extremely high σ, which results in very volatile beliefs. On the other hand, γ tends to be estimated very imprecisely and usually includes zero into the HDP interval.

Therefore, we decided to perform a set of estimations with γ and σ fixed at the values consistent with usual posterior modes. At the same time, we allowed ρ to be estimated, as the models clearly preferred it to be slightly below one. We suggest that this is because when the beliefs are perceived to be a random walk, we do not use a consistent Kalman filtering to deal with the problem. For example, imposing diffuse prior Kalman filtering on the beliefs would be prohibitely complex.

produce more diversity in the forecasts. The BIC selection criterium is of course also heavily influenced by the penalty for the degrees of freedom of the various models, and it is not straightforward whether and how these corrections should be taken into account.

The table also illustrates that these results are not very sensitive to the specific choice of the initial beliefs. If we allow the initial beliefs to adjust over a presample period (30 quarters before the start of the actual likelihood evaluation corresponding with the period 1958-1965), the results remain robust. The beliefs over this presample period were quite stable but nevertheless there is some influence of presample data on the initial beliefs.

In order to understand where the improvement of the marginal likelihood comes from, we tried to evaluate the relative contribution of the specific belief assumptions on the one hand and the time variation produced by the KF learning on the other hand. Therefore, we estimated the model with fixed beliefs corresponding with the initial beliefs of the small forecasting models (under both BIC and EW forecast combinations). This exercise shows that the rational expectations assumption is restrictive: by just replacing the model consistent expectations (which are but linear functions) by a different set of simple (and fixed) expectations that use a much smaller information set, the model fit is already improved significantly: from -926 for the RE beliefs to -920 (BIC initial beliefs) and -916 (EW initial beliefs). When the KF updating of these beliefs is also allowed for, the marginal likelihood further drops to -911 for BIC selection and -909 for EW, respectively. Kalman filter updating generates beliefs that allow the model to better match the data.

We also estimated the KF model with single forecasting models. The results showed quite some difference in marginal likelihood depending on the specific forecasting model used by the agents, but the model with beliefs formed based on the AR(2) specification performed best and yielded a marginal likelihood that was similar to the best performing equal weights model.

Table 1: Model comparison in terms of Marginal Likelihood.

REE model (natural output gap)	-922
REE model (TFP based output gap)	-926
KF Learning with 5 small models: same sample for beliefs and model estimation	
5 models, BIC selection, γ and σ estimated, $\rho = 1$	-917
5 models, EW combination, γ and σ estimated, $\rho = 1$	-910
5 models, BIC selection, γ and σ fixed, ρ estimated	-911
5 models, EW combination, γ and σ fixed, ρ estimated	-909
KF Learning with 5 small models: longer sample for beliefs than for model estimation	
5 models, BIC selection, γ and σ estimated, $\rho = 1$	-916
5 models, EW combination, γ and σ estimated, $\rho = 1$	-910
5 models, BIC selection, γ and σ fixed, ρ estimated	-910
5 models, EW combination, γ and σ fixed, ρ estimated	-909
No learning, constant beliefs	
5 models, constant beliefs from BIC selection	-920
5 models, constant beliefs from EW combination	-916

Figure 1 compares the likelihood evaluation over time for the best performing BIC and EW models relative to the RE model, with negative cumulative difference indicating better fit of the learning models relative to the RE one.[11] The learning models perform better than the REE model around the late seventies but loose in the beginning of the eighties, and start to improve again since the beginning of the nineties. It is interesting to note that the improvement in the likelihood over the nineties, correspond with improved one period ahead predictions of the learning models for inflation and wage behaviour over this period, presented in Figures 2 and 3. This result is remarkable given that the RE model did already extremely well in terms of out of sample predictions for inflation and wages over this period (See Smets and Wouters 2007 for the out of sample prediction performance of the RE model).

[11]Here and later, unless otherwise noted, "the best performing BIC and EW models" are the models with the same time period used to update the beliefs and evaluate the models, and with γ and σ fixed but ρ estimated.

Figure 1. Cumulative likelihood the best performing BIC and EW models over time, relative to the RE model.

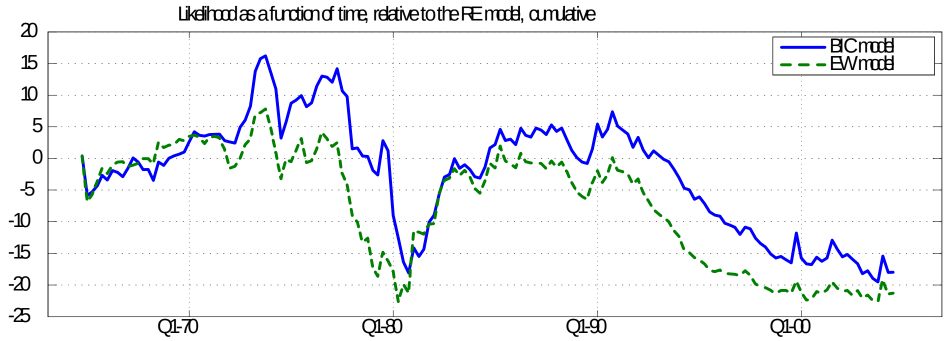


Figure 2: One period ahead prediction performance for inflation.

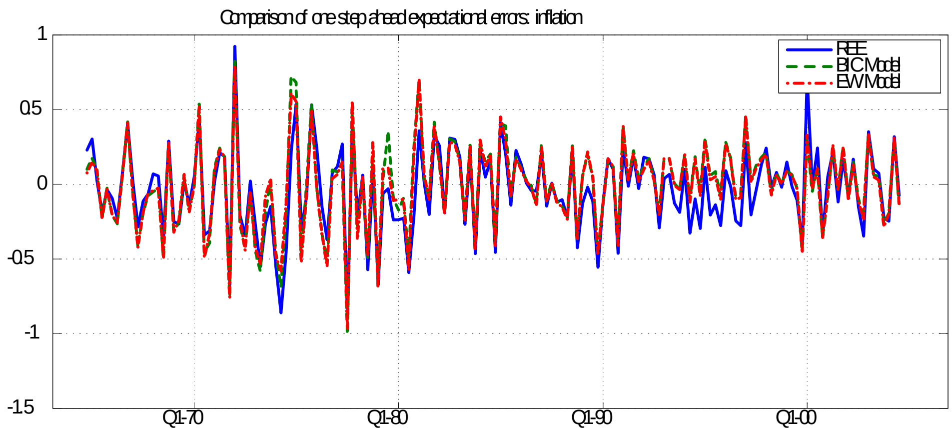
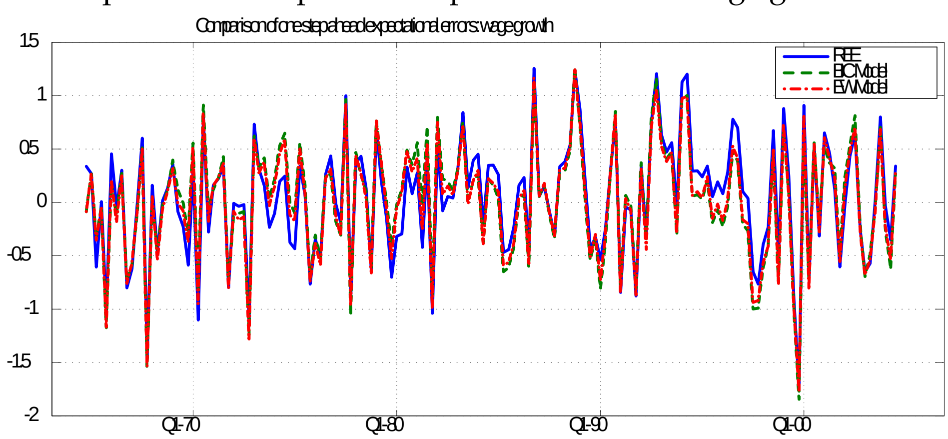


Figure 3: One period ahead prediction performance for wage growth.



4.2 Posterior distribution of the parameters

The estimated parameters of the model with KF learning deviate from the RE estimates in a consistent and very interesting way. The most important changes are observed for the stochastic processes that describe the exogenous price and wage mark up shocks. These exogenous processes are assumed to follow an ARMA(1,1) process. Under the RE assumption, these processes typically have both a very persistent autoregressive component (0.87 for the price mark up (ρ_p), and 0.97 for the wage mark up shock(ρ_w)). In the model with KF learning, this autoregressive component declined to values between 0.5 and 0.7 for the price mark up shock and 0.59 and 0.69 for the wage mark up shock. These values fall clearly outside of the range of possible values implied by the posterior distribution of the RE model. At the same time, the moving average coefficients (μ_p,μ_w) are very close to the corresponding autoregressive coefficients (ρ_p, ρ_w), which suggests that the price and wage mark up shock are in fact following a pure iid process in the models under learning. The priors for these four coefficients are centred around 0.5, which might explain why they are not estimated to be close to zero and stay around 0.5 instead. One can also observe that the posterior distributions for both the autocorrelation and the moving average terms are very similar to the prior distribution, which suggests that the data are not able to pin down these parameters with a high precision in the model with learning. On the other hand, under RE, the posterior distribution for these parameters was very much concentrated around high values.

Guided by the results on mark–up shocks, we performed estimation of the model assuming that both price and wage mark–up shocks are *i.i.d.*, using the same sample for beliefs and model estimation, equal weights, and fixed γ and σ but ρ estimated. The marginal likelihood is the same as for this estimation in the Table 1: -909. All structural rigidity parameters remain essentially the same. We conclude that adaptive learning completely replaces the persistence introduced into the RE model through the price and wage mark–up shocks.

The degree of wage indexation (ι_w) also decreases systematically under learning: 0.34 to 0.39 versus 0.59 under RE, while the wage stickiness (ξ_w) tends to increase but much less significantly. The opposite applies for the price setting: here the price stickiness tends to decrease while the indexation tends slightly upwards. Except for the wage indexation and to a lesser degree price stickiness, these changes remain small, but the overall direction is towards less structural rigidities. We also observe a significant decline in investment adjustment cost (φ), but an increase in the gradualism of the monetary policy rule (ρ_r). The other exogenous processes and the parameters describing the endogenous frictions of the model do not show any systematic changes.

Our results confirm to some extent the results obtained in Milani (2006). He found that learning could completely explain the observed persistence in inflation and consumption behaviour, and that there was no need for price stickiness or habits in the model under learning. Our results confirm that learning is crucial for understanding the inflation dynamics of prices and wages: we do not require exogenous persistence in the shocks in order to explain the observed inflation dynamics and we also require a lower degree of wage indexation in the model, but the estimated degree of price and wage stickiness are not significantly lower than under RE. There is also some evidence that learning might be important for explaining the investment dynamics, but it does not replace the degree of habit persistence. To compare our results with Milani (2006), note that he does not differentiate between consumption and investment, so that his habit persistence refers to both expenditures categories at the same time.

These results differ from our previous estimates under learning with large forecasting models (Slobodyan and Wouters 2007). In that model, the estimated coefficients did not change significantly, although a close look at the result also shows that under VAR learning (meaning that agents only use the seven observed time series in their forecasting model) the posterior distribution for the autocorrelation in the price and wage mark up shock gives slightly more weight to lower values than is typically observed in the RE model. Our results here are however completely confirmed by the DSGE–VAR estimation of the RE model presented in Slobodyan and Wouters (2007), where the mean of the posterior for both autocorrelations were estimated at 0.56 (price mark up) and 0.74 (wage mark up).

Table 2: Model comparison in terms of estimated parameters.

	φ	λ	ξ_w	ι_w	ξ_p	ι_p	ρ_r	ρ_p	μ_p	ρ_w	μ_w
REE model: natural output gap	5.49	0.71	0.74	0.59	0.66	0.23	0.82	0.89	0.70	0.97	0.85
REE model: TFP-based output gap	5.63	0.77	0.71	0.59	0.70	0.22	0.84	0.87	0.73	0.97	0.88
KF: same sample beliefs											
5 models, BIC selection	4.56	0.75	0.76	0.36	0.60	0.23	0.90	0.69	0.57	0.59	0.48
5 models, equal weights	3.17	0.79	0.74	0.38	0.60	0.29	0.90	0.57	0.53	0.65	0.45
5 models, BIC, ρ est	4.61	0.76	0.79	0.34	0.64	0.25	0.88	0.48	0.48	0.56	0.46
5 models, EW, ρ est	3.73	0.79	0.77	0.38	0.64	0.24	0.90	0.48	0.49	0.62	0.46
KF: long sample beliefs											
5 models, BIC selection	4.04	0.78	0.75	0.37	0.60	0.20	0.90	0.70	0.60	0.68	0.55
5 models, equal weights	2.91	0.82	0.73	0.39	0.59	0.25	0.91	0.59	0.52	0.69	0.48
5 models, BIC, ρ est	4.37	0.75	0.77	0.36	0.64	0.22	0.89	0.57	0.54	0.69	0.58
5 models, EW, ρ est	3.90	0.78	0.77	0.39	0.65	0.18	0.89	0.50	0.53	0.68	0.51
Constant beliefs											
BIC initial beliefs	4.84	0.66	0.67	0.51	0.53	0.12	0.88	0.96	0.60	0.96	0.78
EW initial beliefs	4.93	0.77	0.69	0.38	0.60	0.70	0.89	0.25	0.42	0.96	0.72

5 Time variation generated by the learning process

5.1 Implied time variation in the beliefs

KF learning leads to important time variation in the coefficients of the individual forecasting models. This time variation is similar across models, but the interpretation is easier for the simple autoregressive models and therefore we plot the results for the AR(1) and AR(2) model in Figure 4. The graph illustrates the time variation in the coefficients of the forecasting models for five of the seven forward variables in the model. These results are representative for all of the five forecasting models. The graphs show the behaviour of both the constant and the sum of the two lagged coefficients in the beliefs. First, it is clear that the constants, or the mean expected or perceived values, vary a lot for all of the forward variables. These constants in the beliefs play a very important role as they relax the restrictions imposed in the RE model that all real variables grow at the common trend growth rate and that inflation is centred around the fixed inflation objective of the central bank. So the fluctuations in the constants of the belief equations can be interpreted as deviations of private sector expectations from these unconditional means. The constants in the real variables fluctuate over the cycle reflecting the past growth rates observed in the individual variables. Clearly, the constant for the expected investment rate is the most cyclical, while the constants for consumption, labour and the real wage reflect more the long term growth rates in these variables which deviate quite persistently from the imposed common productivity growth rate in the model. For inflation, the constant also reflects the trend in the past observed inflation rate. The expected mean inflation rate rose during the seventies and started to decline only slowly after the disinflation of the early eighties. The constant term in the inflation beliefs stabilised around zero since the mid nineties, meaning that the expected mean inflation of the private agents varied around the constant inflation objective of the central bank since then.

The autoregressive coefficients are more stable for most of the variables except for the inflation expectations.[12] The perceived inflation persistence peaked around the mid seventies and again around 1980 and from then on, declined gradually and stabilised around 0.6 since the mid eighties. This perceived inflation persistence by the private sector plays a very important role in the model. Shocks in the inflation rate were perceived as much more persistent in the seventies then they were in the more recent period. This reflects the fact that monetary policy and the inflation objective of the central bank became much more credible over the last twenty years. We will discuss in the following sections in detail how this perceived inflation persistence affects the impulse responses of shocks and how

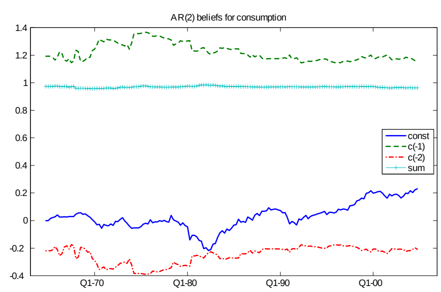
[12]Coefficients of the perceived AR(2) processes for real wages, consumption, investment, and output suggest that the true data generating process might probably be best described as AR(1) in first differences.

they can be helpful to understand the great inflation in the seventies. The perceived inflation persistence as estimated in our forecasting models corresponds with the statistical properties of the observed inflation process over this period. For instance, Cogley, Primiceri and Sargent (2007) obtain a very similar pattern for the persistence in the inflation gap.

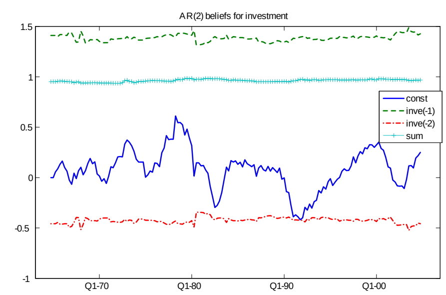
The extremely high perceived inflation persistence in the mid and late seventies also explains why the updating in the beliefs during these years sometimes leads to explosive outcomes. As it is standard in the learning literature, the projection facility in our estimation process cancels the updates in the beliefs that would result in unstable dynamics for the inflation process.

Figure 4: Time variation in the belief coefficients of the AR(2) model

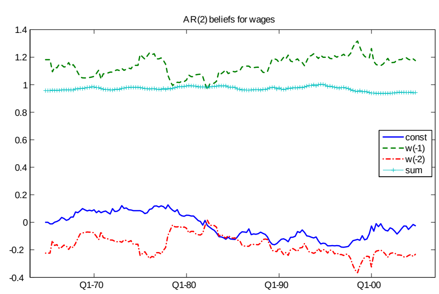
Consumption



Investment

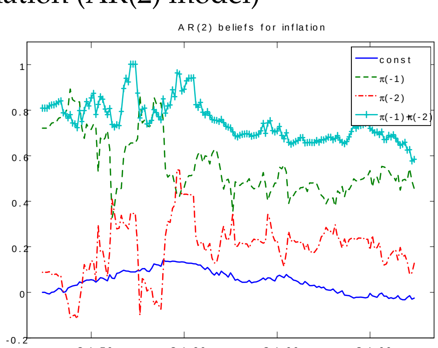


Real wage

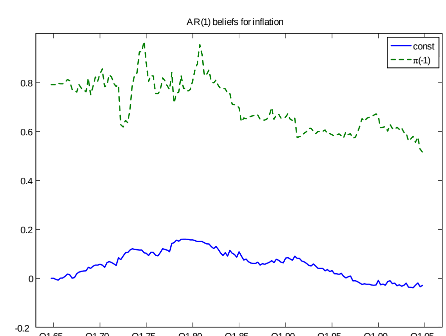


Labour

Inflation (AR(2) model)

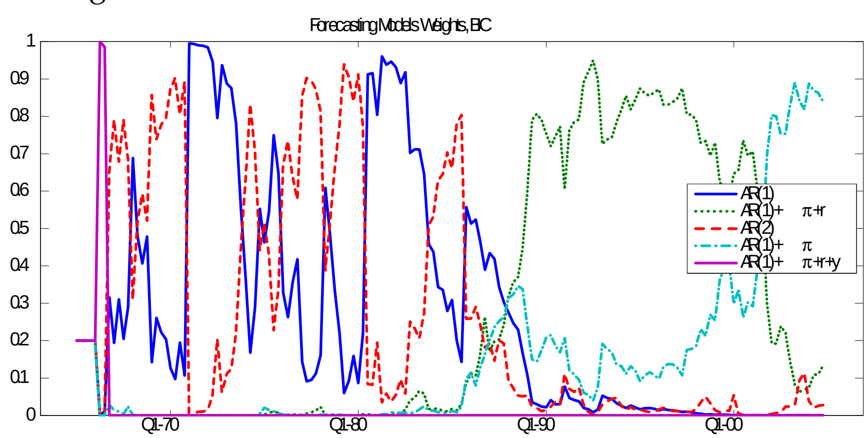


Inflation (AR(1) model)



In the case where we combine the individual model forecast based on the BIC selection criterium, the aggregate forecasting model will not only change because of the KF updating of the individual models but also because the weights appointed to the individual models change over time. Figure 5 shows how these weights change over time for the best performing BIC model. In the beginning of the estimation period, the simple AR(1) and AR(2) models perform best and are the only ones retained according to the BIC criterium. Later on, especially after the decline in the persistence of inflation, the forecasting models that use more information from the other variables, like the inflation rate and the interest rate, tend to dominate the pure autoregressive models. Note that the largest model that includes output in addition to inflation, the interest rate and the own lags, does never receive a significant weight in the aggregate model because it is punished to much for the degrees of freedom under the BIC selection criterium. With simple averaging of the individual forecasting models (equal weights), it is interesting to note that the aggregate model yields forecasts that are competitive to the best individual forecast model over the complete sample.

Figure 5: Weights of the individual model under the BIC selection criterium



5.2 Implied time variation in the IRF

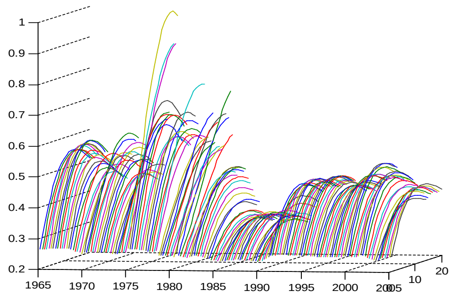
The transmission of the various shocks in the model depends of course strongly on the way private agents form their expectations. Therefore it is interesting to study how the IR functions depend on the information set that agents include in their expectations and on the Kalman filter updating of these beliefs. In Figure 6 and 7, we plot the time varying IRF for the productivity shock, the intertemporal risk premium shock, the monetary policy shock and the wage mark up shock for respectively the EW learning model and the learning model where expectations are only formed on the single AR(2) forecasting model. Only the effects on output and inflation are shown. We calculate the IRF for given belief coefficients at each point in time and disregard the updating of these beliefs that might be caused by the shock. In doing so, these pseudo-IRF might underestimate the persistence and the magnitude of the responses.

For all shocks, it is clear that the reaction of inflation depends crucially on the perceived persistence of inflation by the private agents. Inflation reacted much stronger and more persistently to the shocks in the seventies when inflation was perceived as very persistent. This picture applies for both the productivity shock, the demand shock and the wage mark up shock and for both assumptions on the belief models (EW or the single AR(2) model). For the monetary policy shock, the same result applies when expectations are based on the single AR(2) model. As in Slobodyan and Wouters (2007), we observe that the reaction of inflation to the monetary shock is much more gradual on impact but more persistent afterwards in comparison to the typical impulse response in the RE model. For beliefs based on the combination of the five models, a more complicated picture appears. During the seventies, inflation reacted positively to the monetary policy shock on impact and turned negative only after several quarters. This result is explained by the important correlation between the interest rate and future inflation in these historical episodes, which is taken up in the forecasting models where the interest rate is part of the information set that agents use to form beliefs. Unexpected interest rate increases are then corresponding with rising inflation expectations. The model is not able to identify the exact source of these events, and therefore classifies the shocks as monetary policy innovations. This result is very similar to the traditional price puzzle in SVAR models, and can only be solved if more information is included in the private agents' expectation models. It is also remarkable that the influence of the monetary policy shock on inflation becomes very small over the recent period.

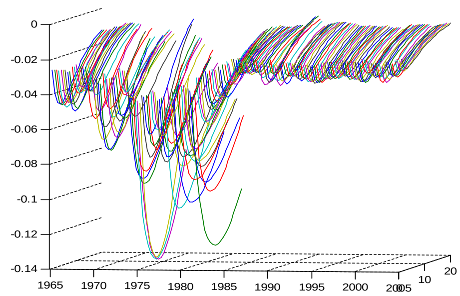
The impact of the shocks on output displays less time variation, although the impulses are somewhat stronger and more persistent in the mid seventies as it is the case for inflation. Some of the impulse responses are however strongly dependent on the specification of the forecasting model. If the expectations are purely based on own lags, some of the shocks, like the productivity and the wage mark up shock, clearly have a much more persistent effect on the economy.

Figure 6: IRF for the EW learning model

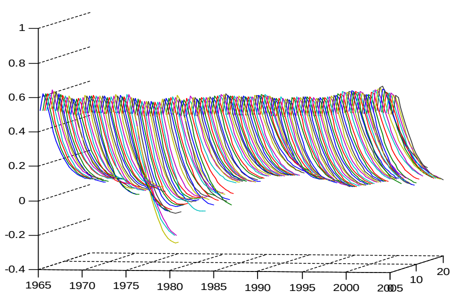
Productivity shock on output



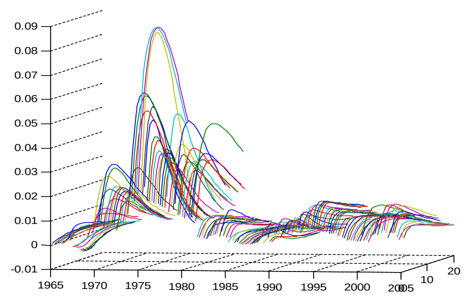
Productivity shock on inflation



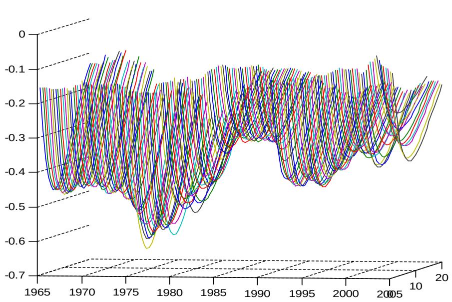
Risk premium shock on output



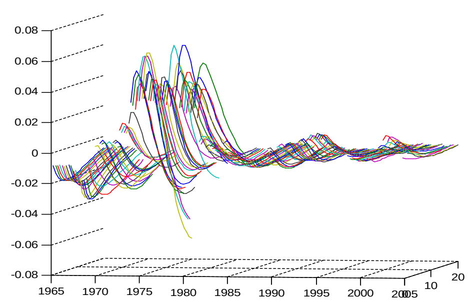
Risk premium shock on inflation



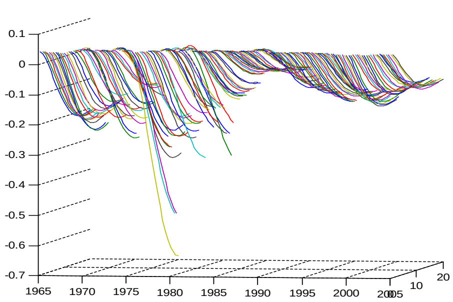
Monetary policy shock on output



Monetary policy shock on inflation



Wage mark up shock on output



Wage mark up shock on inflation

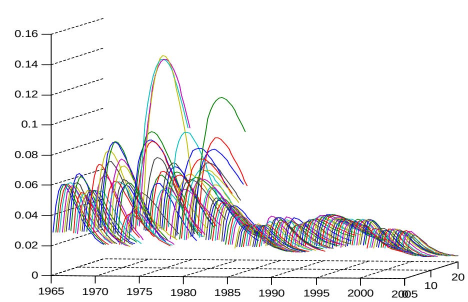
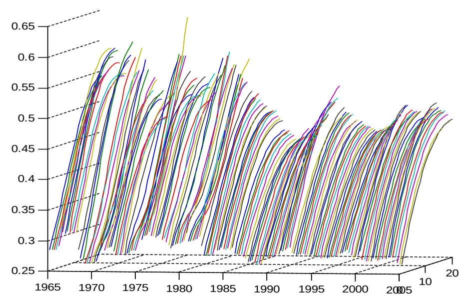
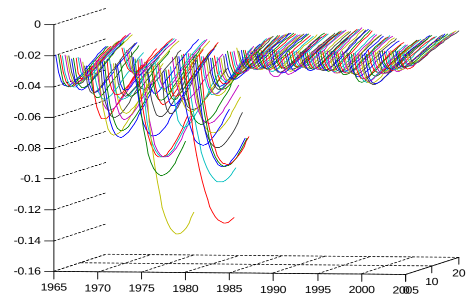


Figure 7: IRF for the AR(2) learning model

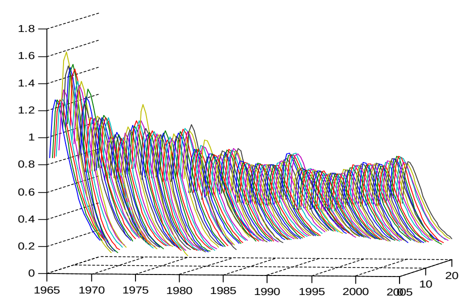
Productivity shock on output



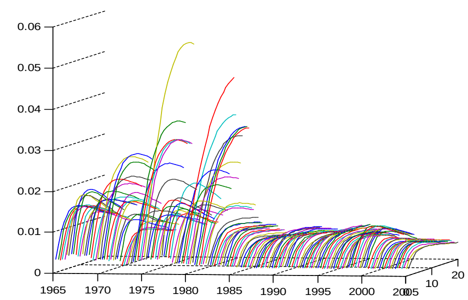
Productivity shock on inflation



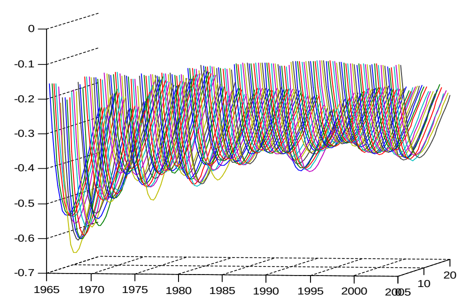
Risk premium shock on output



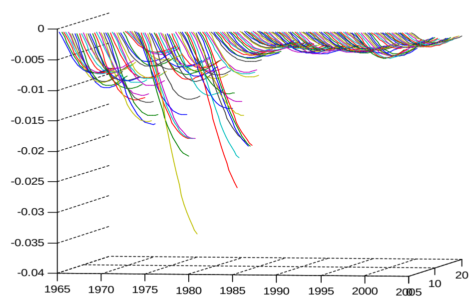
Risk premium shock on inflation



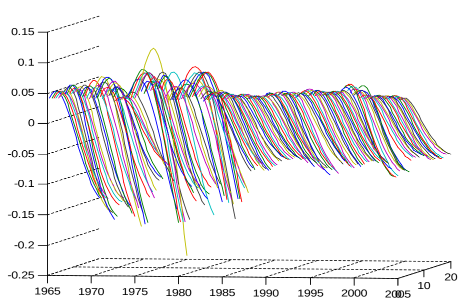
Monetary policy shock on output



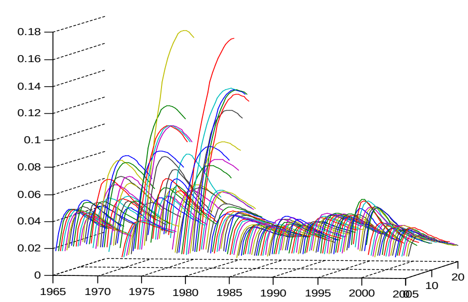
Monetary policy shock on inflation



Wage mark up shock on output



Wage mark up shock on inflation



5.3 Implied time variation in the variance decomposition

Given the time variation in the way agents formed expectations in the model and the effect of this on the transmission mechanism of the different shocks, it is interesting to evaluate how this contributed to the overall volatility in the economy. The results are most outspoken for inflation. The model produces both a higher mean inflation and a much higher inflation volatility in the seventies than in the period since 1984, see Table 3.[13] The mean inflation is of course directly related to the time varying constant in the belief equations of the private sector. The higher volatility is explained by the higher perceived inflation persistence and the stronger and more persistent reaction of inflation to all the shocks in the seventies compared to the more recent period. Averaging over the sub–periods before and after 1984, the model explains a drop in inflation volatility from 0.6 to 0.3, which corresponds almost exactly with the observed drop in inflation volatility in the historical data. Also in terms of the mean inflation rate, the model explains a large share of the observed decline. These results clearly illustrate the crucial role of inflation expectations to explain the great inflation experience of the seventees. The series of upward inflation shocks that arose in the mid seventies led to an upward revision in the mean expected inflation rate by the private agents and at the same time they also revised their perceived inflation persistence which reinforced the impact of the negative shocks on inflation even further. This revision in the inflation expectations of the private sector happened independently of the monetary policy behaviour, as the policy rule in our model is assumed to be constant over the complete estimation period. In the beginning of the eighties, restrictive monetary policy shocks caused agents to revise downward their expectations about future mean inflation and the perceived inflation persistence, so that inflation gradually converged towards the inflation objective of the central bank. The crucial mechanism in this explanation of the great inflation is the interaction between the way inflation expectations are formed and the specific series of historical shocks that appear over time. This interpretation suggests that monetary policy makers should be extremely careful about inflation expectations and how these react to positive inflation shocks.

For the real output growth rate, the model is able to replicate the increase in the average growth rate over the two sub–periods, but it does not explain the great moderation in the volatility of the growth rates.

[13]To produce these numbers, 500 draws from the MCMC were randomly selected. At every parameter draw, the time–varying μ, T, and R implied by the changing beliefs, were saved. Then this time–varying VAR was simulated 500 times to produce 500 hypothetical alternative histories for the estimation period. Before– and after–84 means and standard deviations were then averaged over all histories, and then over all parameter draws.

Table 3: Mean and volatility in inflation and output growth

	Before 1984				After 1984			
	Data		Model EW		Data		Model EW	
	mean	st.dev.	mean	st.dev	mean	st.dev	mean	st.dev
Inflation	1.4273	0.5964	1.1152	0.5880	0.5904	0.2400	0.6272	0.2966
Output growth	0.3801	1.1245	0.3551	0.9491	0.4965	0.5398	0.5167	0.9331

Figure 8 provides more information on the time variation of the implied model variance over time. The figure shows the theoretical one period ahead and unconditional forecast error variance implied by the time varying model coefficients.[14] A drop in one period ahead variance of inflation further illustrates the results discussed above for the two sub–periods. The unconditional inflation forecast variance exhibits double peak in mid–70es to early 80es, clearly reflecting close to random walk beliefs about inflation during this period. For the output level, the pattern is different, and one period ahead forecast error is high both in the beginning and the end of the sample. The unconditional variance of output level is clearly influenced by the implied persistence of inflation, as evidenced by the double peak contemporaneous with that of inflation. It shows some evidence of a moderation in unconditional volatility since 1984. On average, there is a significant decrease in the unconditional variance over the subsample before and after 1984, a trend which is absent in the one period ahead forecast variance. This decline is also partially reversed after 1992 and unconditional volatility peaks again in 2000. This time pattern is consistent with the time variation in the IRF that was displayed above. The absence of a declining trend in the one period ahead forecast variance is compatible with the results in Table 3. The introduction of learning does not seem to be contributing singificantly to the moderation in real growth rates. This moderation is mainly due to the declining volatility of the realized shocks since 1984[15].

The divergence between the one period variance and the unconditional variance of output is related to the declining autocorrelation in growth rates implied by the learning dynamics. This feature is illustrated in Figure 10 and discussed later on.

[14]The major difference between these results and those presented in the Table 3 is that here all the variances are computed at a particular point in time, taking into account only T_t and R_t. Thus, the time–varying nature of the constant term μ_t is ignored, unlike in Table 3 results.

In addition, Table 3 presents variance calculated over a specific time interval. This variance is asymptotically equal to the mean (over the same time interval) of unconditional point–in–time variances presented in the right panel of Fig 8 only if μ_t, T_t and R_t were constant over time.

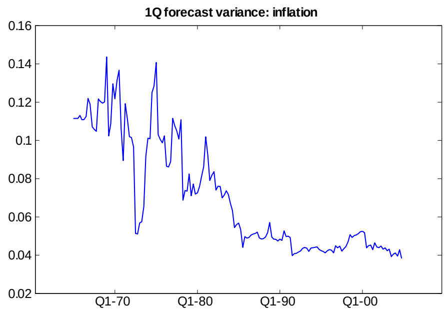
[15]Note that the results in Table 3 are calculated for random shocks, and therefore these statistics do not take into account the decline in the realized historical innovations over these subperiods.

Figure 8: Time varying variance implied by the learning model (EW)

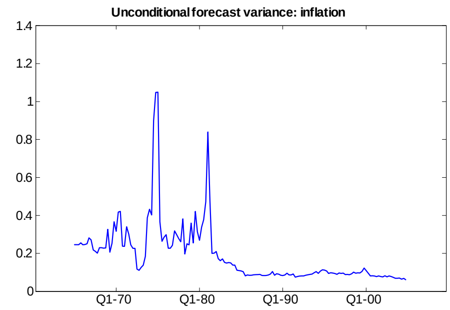
1Q forecast variance

Unconditional forecast variance

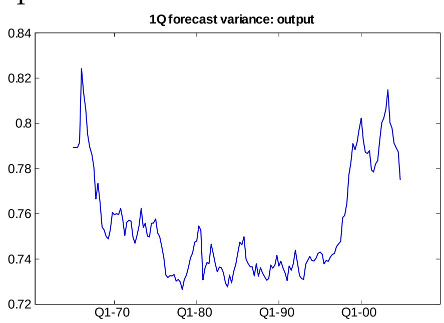
Inflation



Inflation



Output level



Output level

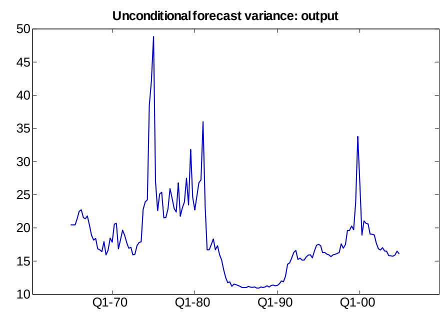


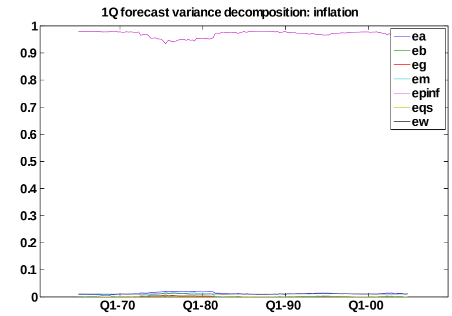
Figure 9 summarizes the variance decomposition for inflation and output level. While the variance decomposition one period ahead is quite stable, there is substantial variation in the unconditional variance decomposition. In the case of inflation, the relative importance of the different shocks changed with the perceived inflation persistence. The short run volatility, typically generated by the price mark–up shock, became less dominant when the inflation process was perceived as very persistent and all shocks affected inflation in a more persistent way. In the mid–nineties, we observe a rise in the contribution of the wage shocks. This period corresponds with the improved fit of the learning model compared to the RE model, which suggests that something special happened in that period with real wages and inflation during that period. The decomposition for output shows that productivity and monetary policy became less dominant in the late eighties, while the contribution of the demand shocks increased. The impact of the wage mark up shock on output is very volatile. These results are in line with the time variation observed in the Impulse Response Functions.

Figure 9: Time varying variance decomposition implied by the learning model (EW)

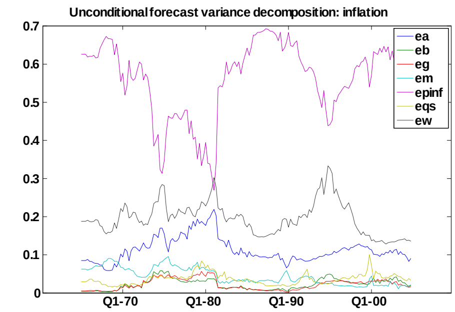
1Q forecast error variance

Unconditional forecast variance

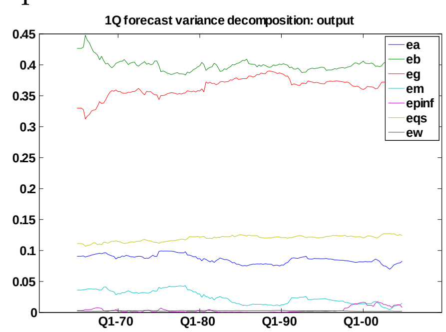
Inflation



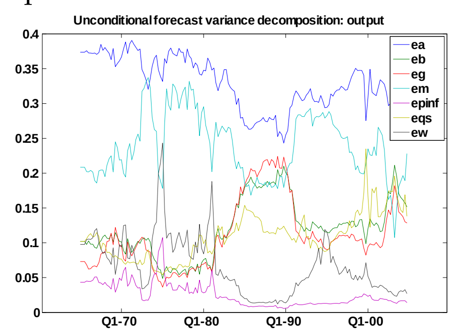
Inflation



Output



Output

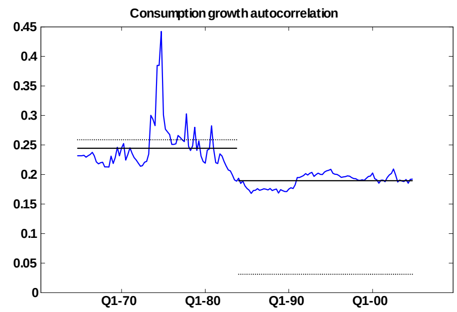


To understand further the changes to the transmission mechanism of the model brought about by the adaptive learning, we investigate unconditional second moments of growth rates of real variables, such as consumption, investment, output, and real wages, implied by the time–varying beliefs.[16] Figure 10 presents implied autocorrelations. All rates demonstrate a sharp peak around 1975, a marked decline towards 1985, and a limited pick–up after 1990. Dashed line in every graph represents the autocorrelation coefficient derived from the observed time series, and solid line averages over the theoretical autocorrelation. Implied autocorrelations for the investment and output growth rates are always (but for a short period around 1975) lower than the observed value, the opposite result is obtained for the wage growth (note, however, that both the data and the model imply very low correlation for the wage growth rate), and consumption growth is first less and later more persistent than in the data. Our results confirm the results of Eusepi and Preston (2008) on the potentially important role of learning dynamics in the propagation mechanism of the shocks.

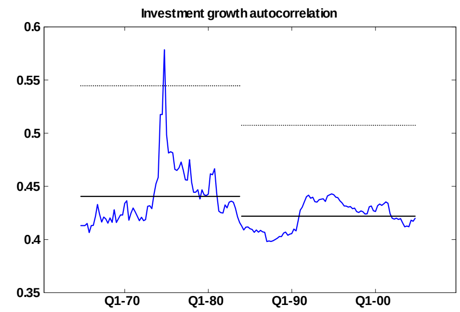
[16] We selected the variables that are model counterparts of the variables observed in first differences.

Figure 10: Time varying autocorrelations implied by the learning model (EW)

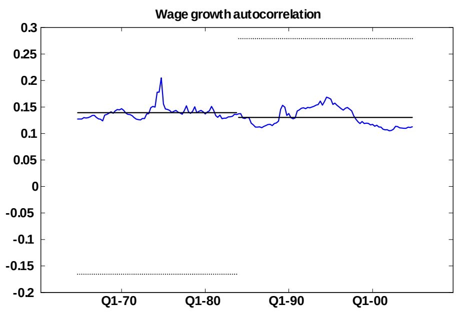
Consumption growth



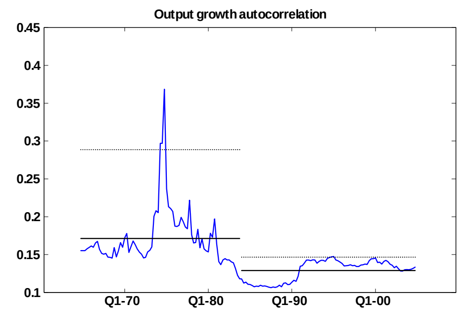
Investment growth



Wage growth



Output growth



Unconditional variances of the growth rates (not shown) tell a similar story: after a peak around 1975, a drop–off turns into an increase after 1990 at least for investment and output growth. Not surpricingly, investment growth is the most volatile and shows most time varying variance. The pattern for the output growth rate is compatible with the absence of moderation that we report in the Table 3.

6 Inflation expectations and survey evidence

Given the crucial role of inflation expectations, it is important to check whether the inflation expectations implied by the model are confirmed by the historical survey evidence on inflation expectations. In order to provide some evidence on this, we calculated the correlation between the inflation expectations generated by the RE model and EW learning model with the expectations observed in the Survey of Professional Forecasters. For the period 1970-2004, we observe that the correlation between the expectations of the learning model and the RE model expectations is almost perfect and both series are very highly correlated with the SPF series. In first differences, the learning expectations are slightly higher correlated with the survey series.

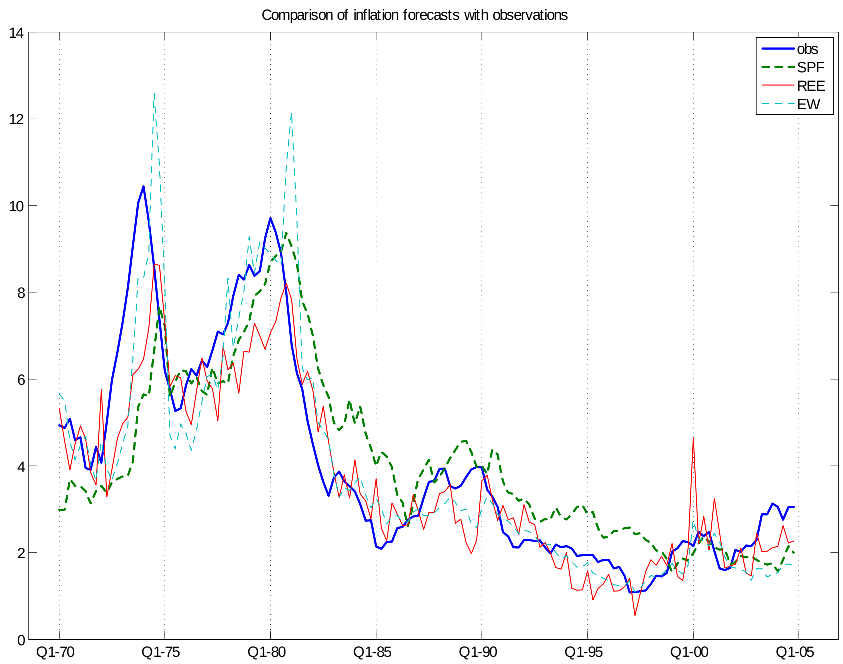
Table 4 : Correlation between model beliefs and SPF-survey beliefs about future inflation

	RE beliefs	Learning EW beliefs
correlation in levels	0.87	0.87
correlation in first difference	0.20	0.26

Figure 11 presents a comparison of the SPF expectations, expectations produced by the DSGE model estimated under RE and adaptive learning with equal weighting of forecasting models. We also include time series for observed inflation. SPF expectations are about one year ahead inflation starting from the quarter following the one in which the question is asked, and all other time series are transformed into this format. There are three periods in which survey expectations deviate systematically from the observed inflation series: run–up to high inflation in seventies, the disinflation period in the early eighties and the mid–nineties. During both inflation run–up and the disinflation periods, the RE inflation expectations and the learning inflation expectations react very similarly and adapt much faster than the survey series. Of course, the mechanism that are behind these expectations can be very different in the two models: in the RE model it is probably driven more by the persistence of the exogenous shocks, while under the learning model the persistence is mostly accounted for by the learning mechanism. Nevertheless, the resulting persistence in the expectations underestimates the one observed in the survey data. For the mid–nineties, the learning model produces higher inflation expectations which are somewhat closer to the ones observed in the SPF survey (and to the actual inflation), while the RE model underestimates the SPF expectations and actual inflation. Higher expected inflation rates during this period was important for improving the model fit for inflation and wages over this period as we saw before.

At one–year horizon, model–based expectations predict observed inflation better than the SPF: standard deviation of the difference is 1.47, 1.08, and 1.16 for SPF, RE, and adaptive learning with equal weights, respectively, over the whole sample. After 1985Q1, adaptive expectations become significantly more precise than the rational expectation one, with standard deviations being 0.77, 0.70, and 0.57, respectively.

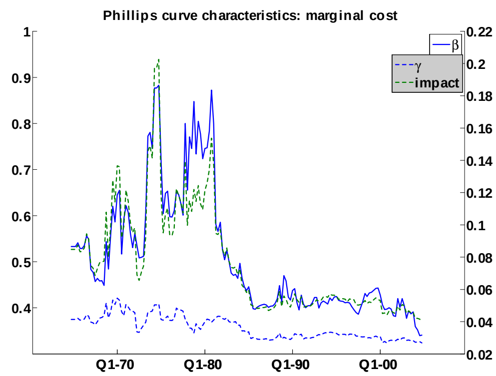
Figure 11: Inflation expectations from the models and the SPF-survey



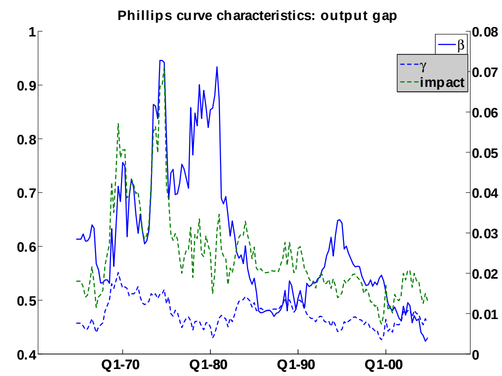
7 Implications for the Phillips Curve

Our framework allows us to investigate whether the adaptively learning agents would have perceived a 'flattening' of the Phillips curve over the last two decades. Flattening of the Phillips curve has been observed for many alternative specifications, cf. Atkeson (2001), Stock and Watson (2006), Borio and Filardo (2007), and others. Using the time–varying VAR representation of the adaptive learning model with equal weighting of forecasts exemplified in matrices T_t and R_t, we projected current inflation on the current measure of economic conditions (marginal cost or output gap) and past inflation, at each point in time. The results are presented in Figure 12. The left panel shows that if our agents would have used real marginal costs as a measure of macroeconomic conditions, they would indeed perceive a marked flattening of the Phillips curve beyond 1985 (dashed line, right axis). For the agents using output gap instead (right panel), the conclusion would not have been so clear, as there are two periods of significant flattening separated by a marked increase in the slope of the Phillips curve in early eighties. However, if the agents were interested in effect of a sustained four quarters change in economic conditions on the one year ahead inflation (dotted line, left axis), their conclusion would have been clear: over time, the one year ahead impact on inflation drops significantly. In case of marginal costs used as a proxy for economic conditions, the impact is mostly drived by the perceived inflation persistence (solid line, left axis), while in the output gap case the slope of the Phillips curve also plays a not insignificant role.

$\pi_t = \beta\pi_{t-1} + \gamma mc_t$



$\pi_t = \beta\pi_{t-1} + \gamma ygap_t$



8 Conclusions

The hypothesis of model-consistent expectations, especially in the context of a medium-scaled DSGE model, implies that economic agents are extremely well informed both about the structure of the model and the type of shocks that are hitting the economy at each point in time. Therefore, it should not surprise that models with simpler, and probably more realistic, assumptions about the expectations mechanism can improve the empirical fit of these models. In addition, our results suggest that there might be an important role for learning in these expectations: agents update their belief models dependent on the realized past data and by doing so their reaction to exogenous shocks change considerably over time. This process is particular relevant to understand the changing dynamics of the inflation process. Even under a constant monetary policy rule, the beliefs of the private agents about the mean and the persistence of the inflation process can vary substantially over time. The additional dynamics from the learning process substitute for the persistence in the exogenous price and wage shocks and the backward-looking indexation in the wage setting, which are both very important in the rational expectations version of the model.

The specification of the small belief models may of course be criticized as being ad-hoc. We tried to take into account that problem by allowing agents to consider different small models and to weight them depending on their past forecasting performance. Still, the belief models that we consider might be too restrictive. Introducing evidence from surveys about expectations might help to pin down the relevant information set used by agents and to overcome this problem.

Two other extensions of the paper are on our research agenda. The learning dynamics can potentially also contribute to an explanation of the great moderation on the real side of the economy. At this stage, our belief models for consumption and investment do not seem to capture fully the declining autocorrelation in the growth rates of these variables. Beliefs that do take up this trend, should be able to explain at least part of observed decline in the real volatility. Secondly, we would like to test the time-variation that is introduced in our model through the learning dynamics, against a more general and unrestrictive time-varying VAR model. We are confident that at least for the inflation equation, our model does a good job in reproducing the reduced form dynamics.

9 References

Atkeson. A. (2001), “Are Phillips Curves Useful for Forecasting Inflation?” *Federal Reserve Bank of Minneapolis Quarterly Review*, 25 (1), p. 2-11.

Borio, C. and A. Filardo (2007), “Globalisation and Inflation: New Cross–Country Evidence on the Global Determinants of Domestic Inflation”, BIS Working Paper 227.

Christiano, L.J., M. Eichenbaum and C.L. Evans (2005), “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy” *Journal of Political Economy*, Vol. 113(1), p. 1–45.

Cogley, T., G. E. Primiceri and T. J. Sargent (2008), “Inflation–Gap Persistence in the U.S.” NBER Working Paper 13749.

Del Negro, M. and F. Schorfheide (2006). "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)," Working Paper 2006-16, Federal Reserve Bank of Atlanta.

Del Negro, M., F. Schorfheide, F. Smets and R. Wouters (2007), "On the Fit of New Keynesian Models", *Journal of Business & Economic Statistics*, Vol. 25(2), p. 123-143.

Eichenbaum, M. and J. Fisher (2007), “Estimating the Frequency of Reoptimisation in Calvo-style Models.” *Journal of Monetary Economics*, 54(7), p. 2032-47.

Eusepi, S. and B. Preston (2008), ““Expectations, Learning and Business Cycle Fluctuations,” mimeo.

Evans, G.W. and S. Honkapohja (2001), “Learning and Expectations in Macroeconomics".

Giannitsarou, C., and E. Carceles–Poveda (forthcoming), “Adaptive Learning in Practice", *Journal of Economic Dynamics and Control*

Honkapohja, S., K. Mitra, and G. W. Evans (2002), “Notes on agents’ behavioral rules under adaptive learning and recent studies of monetary policy", mimeo.

Kimball, M.S. (1995), “The Quantitative Analytics of the Basic Neomonetarist Model”, *Journal of Money, Credit, and Banking*, 27(4), p. 1241–77.

McCallum, B.T. (2006), “E-Stability vis-a-vis Determinacy Results for a Broad Class of Linear Rational Expectations Models," NBER Working Papers 12441.

Milani, F. (2005), “Learning, Monetary Policy Rules, and Macroeconomic Stability", mimeo.

Milani, F. (2006), “A Bayesian DSGE Model with Infinite-Horizon Learning: Do ‘Mechanical’ Sources of Persistence Become Superfluous?", *International Journal of Central Banking*, Iss. 6.

Orphanides, A., and J. C. Williams (2005), “Inflation scares and forecast–based monetary policy", *Review of Economic Dynamics*, Vol. 8, 498-527.

Orphanides, A., and J. C. Williams (2003), “Imperfect Knowledge, Inflation Expectations, and Monetary Policy", NBER Working Paper No. W9884.

Orphanides, A., and J. C. Williams (2007), "Robust Monetary Policy With Imperfect Knowledge", ECB Working Paper No. 764.

Preston, B. (2005), "Learning About Monetary Policy Rules When Long-Horizon Expectations Matter", *International Journal of Central Banking*, Vol. 1.

Sargent, T., and N. Williams (2005), "Impacts of Priors on Convergence and Escape from Nash Inflation", *Review of Economic Dynamics*, Vol. 8(2): 360-391.

Sargent, T. N. Williams and T. Zha (2006). "Shocks and Government Beliefs: The Rise and Fall of American Inflation," *American Economic Review*, Vol. 96(4), p. 1193-1224.

Slobodyan, S. and R. Wouters (2007), "Adaptive Learning in an Estimated Medium–Size DSGE Model", mimeo.

Smets, F. and R. Wouters (2003), “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area”, *Journal of the European Economic Association*, Vol.1(5), p. 1123–75.

Smets, F. and R. Wouters (2005), “Comparing Shocks and Frictions in U.S. and Euro Area Business Cycles: A Bayesian DSGE Approach”, *Journal of Applied Econometrics*, Vol. 20(2): 161–83.

Smets, F. and R. Wouters (2007), "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach", *American Economic Review*, Vol 97(3), p. 586-606.

Stock, J. H. and M. W. Watson (2006). “Why Has U.S. Inflation Become Harder to Forecast?” NBER Working Paper 12324.

Vilagi, B. (2007), Adaptive Learning and Macroeconomic Persistence: Comparing DSGE models of the Euro Area. Mimeo.

Williams, N. (2003), "Adaptive Learning and Business Cycles", mimeo.

A Data appendix

The model is estimated using seven key macro-economic time series: real GDP, consumption, investment, hours worked, real wages, prices and a short-term interest rate. GDP, consumption and investment are taken from the US Department of Commerce - Bureau of Economic Analysis databank. Real Gross Domestic Product is expressed in Billions of Chained 1996 Dollars. Nominal Personal Consumption Expenditures and Fixed Private Domestic Investment are deflated with the GDP-deflator. Inflation is the first difference of the log of the Implicit Price Deflator of GDP. Hours and wages come from the BLS (hours and hourly compensation for the NFB sector for all persons). Hourly compensation is divided by the GDP price deflator in order to get the real wage variable. Hours are adjusted to take into account the limited coverage of the NFB sector compared to GDP (the index of average hours for the NFB sector is multiplied with the Civilian Employment (16 years and over) . The aggregate real variables are expressed per capita by dividing with the population over 16. All series are seasonally adjusted. The interest rate is the Federal Funds Rate. Consumption, investment, GDP, wages and hours are expressed in 100 times log. The interest rate and inflation rate are expressed on a quarterly basis corresponding with their appearance in the model.