Journal of Economic Dynamics & Control 36 (2012) 26–46

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Learning in an estimated medium-scale DSGE model

Sergey Slobodyan [a], Raf Wouters [b,*]

[a] CERGE-EI, Politickych veznu 7, 111 21 Prague 1, Czech Republic
[b] National Bank of Belgium/Université catholique de Louvain, Boulevard de Berlaimont, B-1000 Brussels, Belgium

ARTICLE INFO

Article history:
Received 18 November 2008
Accepted 9 January 2011
Available online 10 August 2011

JEL classification:
C11
D84
E30
E52

Keywords:
Constant-gain adaptive learning
Medium-scale DSGE model
DSGE-VAR

ABSTRACT

We evaluate the empirical relevance of learning by private agents in an estimated medium-scale DSGE model. We replace the standard rational expectations assumption in the Smets and Wouters (2007) model by a constant-gain learning mechanism. If agents know the correct structure of the model and only learn about the parameters, both expectation mechanisms produce very similar results, and only the transition dynamics that are generated by specific initial beliefs seem to improve the fit. If, instead, agents use only a reduced information set in forming the perceived law of motion, the implied model dynamics change and, depending on the specification of the initial beliefs, the marginal likelihood of the model can improve significantly. These best-fitting models add additional persistence to the dynamics and this reduces the gap between the IRFs of the DSGE model and the more data-driven DSGE-VAR model. However, the learning dynamics do not systematically alter the estimated structural parameters related to the nominal and real frictions in the DSGE model.

1. Introduction

In this paper, we evaluate the potential role of adaptive learning in an estimated medium-sized DSGE model. In Smets and Wouters (2003, 2007), it was shown that these models, when equipped with a rich set of frictions and a general stochastic structure, explain the data relatively well. However, the DSGE-VAR approach as applied in Del Negro et al. (2007) shows that these models are still misspecified along various dimensions: the marginal likelihood of the models increases significantly if the theoretical cross-equation restrictions are relaxed and the impulse responses of different shocks provide conflicting evidence on the role of nominal price rigidity. One potential source of misspecification in these models might be the stringent assumption of rational or model-consistent expectations. This assumption implies that economic agents, when forming their expectations about future outcomes, know exactly the structural model, its parameters, and the stochastic structure. Endowing the agents with so much knowledge can hardly be considered realistic; therefore, it is important to check the consequences of relaxing this assumption.

We evaluate empirically the fit of a DSGE model while allowing the agents to form their expectations under imperfect knowledge. More specifically, we assume that private agents use adaptive learning: expectations of the forward-looking variables are obtained as linear functions of past model variables. Coefficients of these linear functions, commonly known as beliefs, are re-estimated every period using a constant-gain (perpetual) learning algorithm. The beliefs about the relationship between expectations and current and past variables adapt to the patterns recently observed in the data. Several authors have suggested that adaptive learning can enhance the propagation mechanism of the DSGE models and

* Corresponding author.
E-mail addresses: sergey.slobodyan@cerge-ei.cz (S. Slobodyan), rafael.wouters@nbb.be (R. Wouters).

0165-1889/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jedc.2011.01.016

generate the persistence that is otherwise caused by these models' frictions or by the dynamics in the exogenous stochastic processes. For instance, Orphanides and Williams (2003, 2005a) illustrate how adaptive learning can lead to inflation scares or to increased inflation persistence. Milani (2007) 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.

We extend this previous work by estimating the learning process in a medium-scale DSGE model and by testing whether the additional dynamics in the expectations process can overcome the remaining misspecification problems of the rational expectations model. Furthermore, we investigate systematically the role of initial beliefs and the information set in our learning models. Initial beliefs are important because they introduce some additional transition dynamics to the learning process. The specific form of the initial beliefs is very difficult to discover because they depend on historical observations that are not directly taken into account in the likelihood function. We apply several procedures in this paper to estimate these initial beliefs. They can be based on pre-sample data information. Alternatively, one can search for initial beliefs that maximise the likelihood of the in-sample data. Here, one can assume that the initial beliefs are consistent with the final estimated model, or search for the specific initial beliefs that optimise the in-sample likelihood through their impact on the transition dynamics.

The dynamics generated by the learning process are also crucially influenced by the assumptions about the information set that is used in forming the beliefs. The standard approach, which remains close to the Rational Expectations (RE) assumption, is that agents know the reduced-form model but have to estimate its parameters. In this case, the agents will use the correct minimum state variable (MSV) representation to estimate their expectations regressions. In our application, this assumption implies that agents use a state vector containing 20 variables, many of which are unobserved, forcing the agents to use estimates generated by the recursive Kalman filter. In addition, like in the traditional adaptive learning approach as applied in Evans and Honkapohja (2001), this approach assumes that agents know the exogenous driving processes exactly. Alternatively, we will also consider the case in which agents use only a restricted information set to form their expectations. In this setup, the most natural limitation is that agents use only the observed data in their belief regressions. We label this approach as VAR learning to differentiate it from the standard MSV learning. We further differentiate between a situation in which the agents additionally learn about the constants (the inflation target, the real interest rate and the deterministic trend growth rate) and one where we assume that the 'correct' values of these parameters are known.

Our results suggest that learning is a plausible alternative to the RE hypothesis for estimating medium-scale DSGE models, but at the same time we identify some limitations. First, we observe that when agents are assumed to know the correct structure of the model and only learn about the parameters, both learning and rational expectations result in a similar fit with little time-variation induced by the learning dynamics. This is correct as long as the initial beliefs from which the learning process starts are consistent with the RE-equilibrium dynamics. This result is not surprising: when agents start with correctly specified beliefs and new innovations in the data are entering as white noise disturbances, there is no reason for them to revise their beliefs substantially. Secondly, allowing agents to start from different initial beliefs, either based on pre-sample experience or in-sample optimisation, generates substantial learning dynamics. With optimised initial beliefs, the likelihood of the overall model improves considerably. This improved fit can be driven by the transition from the initial beliefs to the equilibrium beliefs or by the standard updating due to new information. The relative contribution of these two forces is difficult to identify: with initial beliefs far away from the equilibrium and/or with a high gain parameter, the updating process becomes highly sensitive to specific model parameters and the evaluation of the posterior distribution becomes highly complex. Thirdly, reducing the information set that agents are supposed to use in their expectation models can further improve the marginal likelihood of the overall model. The combination of specific initial beliefs with a limited information set fits the data almost as well as the best-fitting DSGE-VAR approach. Finally, the learning models that fit the data better than the model with rational expectations tend to add some additional persistence to the DSGE model, in particular following a monetary policy shock, that reduces the gap between the IRFs of the DSGE model and the more data-driven DSGE-VAR approach. We also observe that the additional dynamics that are introduced by the learning process do not systematically alter the estimated structural parameters related to the nominal and real frictions in the DSGE model.

The structure of the paper is as follows. First, we use DSGE-VAR approach to indicate some of the potential misspecifications in the medium-sized DSGE model based on the work of CEE (2005) and estimated on US data in Smets and Wouters (2007). We introduce our assumptions about the learning process in Section 3. In Section 4, we evaluate the potential role of learning in the model by studying the volatility and persistence of the simulated data for different specifications of the beliefs and for different values of the learning horizon and gain parameter. In Sections 5 and 6, we turn to estimating the model under MSV and VAR learning, respectively.

2. Evidence of misspecification in the rational expectations 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.[1] Following CEE (2005), this DSGE model contains many frictions that affect both

[1] An overview of the linearised model is provided in the Appendix. We refer to Smets and Wouters (2003, 2007) for the formal presentation of the model.

nominal and real decisions of households and firms. The specification of the model 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, the standard Dixit–Stiglitz aggregator is replaced 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 Fisher (2007), the introduction of this real rigidity allows us to estimate a more reasonable degree of price and wage stickiness. The model is closed with a Taylor rule that allows for inertia in the monetary policy reaction. The output gap in this rule is defined as the deviation of output from its flexible economy counterpart, which implies that the flexible economy must be solved simultaneously with the rest of the model. Finally, the model also contains seven stochastic shocks to technology, preferences and policy behaviour.

The model can be detrended with the deterministic trend 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. The estimations are executed using Bayesian methods. First, the mode of the posterior distribution is estimated by maximising the log posterior function, which combines the prior information on the parameters with the likelihood of the data.[2] 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. 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. The number of structural shocks matches with the number of observables that is used in estimation.

In Smets and Wouters (2007), this model was estimated under the hypothesis that agents have rational expectations. It was shown that these models, equipped with a rich set of frictions and a general stochastic structure, explain the data relatively well and have a forecasting performance that is comparable or even better than purely statistical VAR or BVAR models.

However, Del Negro et al. (2007) show that these models are still misspecified along various dimensions. In order to illustrate this, these authors use an alternative estimation strategy, the DSGE-VAR approach. In this approach, the DSGE model solution is approximated by a vector autoregression in the observed variables and the cross-equation restrictions that are implied by the DSGE model are used as priors on the empirical VAR coefficients. Deviations between posterior and prior estimates of the coefficients can be considered as an indication of misspecification in the DSGE model. By introducing and estimating the hyperparameter (λ), which represents the prior tightness, the procedure provides an overall assessment of the DSGE model restrictions. In Table 1, the results of the DSGE-VAR approach are summarised and compared to the standard Bayesian estimation results of Smets and Wouters. A fourth-order VAR approximation is used in the DSGE-VAR exercise. The DSGE-VAR with the optimal marginal likelihood gives more or less equal weight to the data and DSGE model ($\lambda = 1$). The marginal likelihood of the DSGE-VAR model is considerably higher than for the DSGE model, -897.8 versus -922.2,[3] meaning that the out-of-sample prediction power of the model is further improved when the DSGE restrictions are relaxed. The mean and the 5 and 95 percentiles of the posterior distribution of the parameters obtained from the Metropolis–Hastings algorithm are reported. The two estimation strategies yield similar estimates for the structural coefficients and confirm the importance of the various real and nominal frictions in the model. In general terms, the coefficients under the DSGE-VAR approach are estimated less precisely and tend to be centred more around the prior distribution, while the standard errors of the structural shocks are typically estimated to be smaller.

The impulse response functions generated by the two estimation procedures are more informative. They illustrate significant deviations between the DSGE model and the more data-driven DSGE-VAR approach and provide more economic intuition on the nature of the misspecification induced by the former approach. Fig. 1 shows one interesting dimension on which the standard DSGE model tends to be misspecified if we accept the DSGE-VAR model as the benchmark. Following a monetary shock, the DSGE model predicts a relatively quick response of inflation with a peak response within the year following the shock. The timing of the peak effect is very similar to the output one. In the DSGE-VAR model, inflation typically responds more gradually with only a very weak, if any, response in the first couple of quarters, followed by a more persistent decline in inflation afterwards. This type of inflation response is standard for many of the SVAR experiments on monetary policy shocks. This gradual and persistent reaction of inflation following a monetary

[2] See Smets and Wouters (2007) for the discussion of the priors.

[3] The reported marginal likelihood in Table 1 deviates from the one presented in Smets and Wouters (2007) because we do not use a training sample in this exercise.

Table 1
Posterior estimates for benchmark RE-DSGE and DSGE-VAR model.

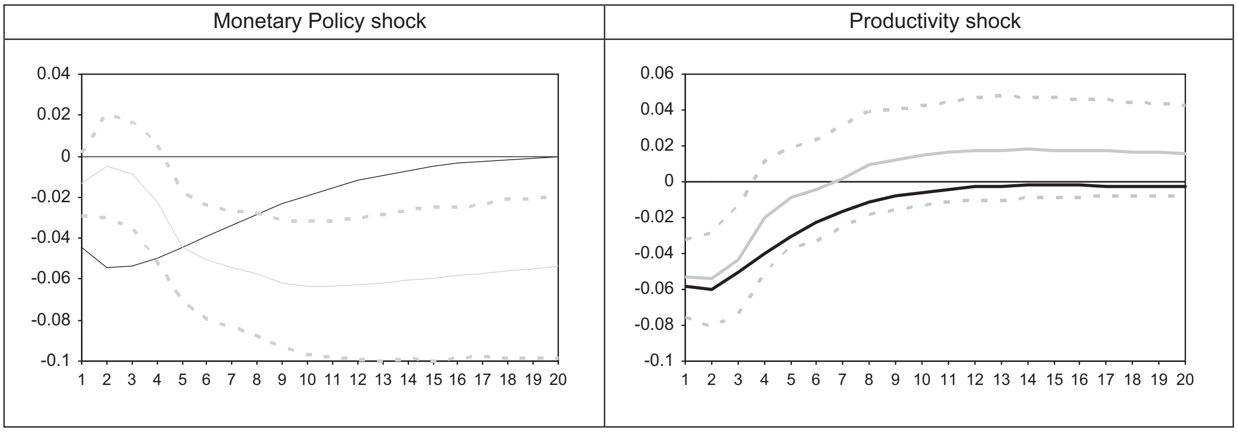
	Prior distribution			Posterior distribution					
				RE-DSGE			DSGE-VAR ($\lambda=1$)		
	Type	Mean	St. dev.	Mean	5%	95%	Mean	5%	95%
St. dev. of the innovations[a]									
σ_a	IG	0.1	2	0.46	0.41	0.51	0.40	0.35	0.47
σ_b	IG	0.1	2	0.24	0.20	0.28	0.16	0.10	0.21
σ_g	IG	0.1	2	0.53	0.48	0.58	0.38	0.32	0.43
σ_q	IG	0.1	2	0.45	0.37	0.53	0.48	0.36	0.62
σ_r	IG	0.1	2	0.24	0.22	0.27	0.18	0.15	0.21
σ_p	IG	0.1	2	0.14	0.11	0.17	0.15	0.12	0.20
σ_w	IG	0.1	2	0.24	0.21	0.28	0.19	0.13	0.23
Persistence of the exogenous processes: $\rho=AR(1)$, $\theta=MA(1)$									
ρ_a	B	0.5	0.2	0.96	0.94	0.98	0.93	0.89	0.99
ρ_b	B	0.5	0.2	0.22	0.08	0.36	0.50	0.20	0.76
ρ_g	B	0.5	0.2	0.98	0.96	0.99	0.77	0.61	0.96
ρ_q	B	0.5	0.2	0.71	0.62	0.81	0.50	0.32	0.68
ρ_r	B	0.5	0.2	0.15	0.04	0.24	0.16	0.04	0.25
ρ_p	B	0.5	0.2	0.89	0.81	0.97	0.56	0.21	0.85
ρ_w	B	0.5	0.2	0.97	0.95	0.99	0.74	0.56	0.98
θ_p	B	0.5	0.2	0.70	0.55	0.86	0.54	0.26	0.83
θ_w	B	0.5	0.2	0.85	0.76	0.94	0.53	0.18	0.85
a_g[b]	N	0.5	0.25	0.52	0.38	0.67	0.52	0.35	0.69
Structural parameters of endogenous decisions[c]									
φ	N	4.0	1.5	5.74	3.97	7.42	3.89	2.37	5.69
σ_c	N	1.5	0.37	1.38	1.17	1.59	1.20	0.86	1.54
η	B	0.7	0.1	0.71	0.64	0.78	0.63	0.51	0.73
σ_l	N	2.0	0.5	1.84	0.92	2.79	1.77	0.81	2.77
ξ_p	B	0.5	0.1	0.65	0.56	0.74	0.64	0.57	0.72
ξ_w	B	0.5	0.1	0.71	0.60	0.81	0.73	0.62	0.85
ι_p	B	0.5	0.15	0.24	0.10	0.38	0.43	0.20	0.75
ι_w	B	0.5	0.15	0.59	0.39	0.78	0.52	0.29	0.76
ψ	B	0.5	0.12	0.55	0.36	0.72	0.53	0.35	0.74
Φ_p	N	1.25	0.12	1.61	1.48	1.74	1.55	1.41	1.69
ρ_R	B	0.75	0.1	0.81	0.77	0.85	0.78	0.74	0.85
r_π	N	1.5	0.25	2.04	1.75	2.33	1.76	1.33	2.06
r_y	N	0.12	0.05	0.09	0.05	0.13	0.11	0.04	0.20
$r_{\Delta y}$	N	0.12	0.05	0.22	0.18	0.27	0.21	0.17	0.27
$\overline{\pi}$	G	0.62	0.1	0.79	0.61	0.96	0.68	0.50	0.85
β[d]	G	0.25	0.1	0.17	0.08	0.26	0.23	0.10	0.36
$\overline{l}$	N	0.0	2.0	0.53	−1.3	2.33	−0.11	−1.5	1.26
γ	N	0.4	0.1	0.43	0.41	0.45	0.39	0.26	0.53
α	N	0.3	0.05	0.19	0.16	0.22	0.19	0.15	0.22

[a] The Inverse Gamma distribution is defined by the degree of freedom.
[b] The effect of TFP innovations on exogenous demand.
[c] See the model appendix for the definition of the parameters.
[d] The estimated coefficient is equal to $100(\beta^{-1}-1)$.

policy shock contrasts with the immediate and short-lived effect of the productivity shock on inflation. This sharp difference in the inflation response to monetary policy and technology shocks is widely recognised in the empirical literature and identified as a stylised fact that is difficult to explain with the standard models based on nominal rigidities and rational expectations.[4]

This observed misspecification of the rational expectation model will serve as a benchmark to evaluate the performance of the models under learning in later sections. Intuitively, there are several arguments which suggest that learning dynamics might be able to overcome this problem. A first argument is that the relative importance of expectations in the transmission mechanism of a shock might be very different depending on the nature of the shock. Productivity shocks have a direct and immediate impact on the marginal cost of the price setter. On the other hand, the impact of monetary policy shocks on the marginal cost and the price decision is more indirect and many of the standard frictions in the DSGE model have precisely the objective of mitigating the influence of interest rates and demand shocks on the marginal cost.

[4] This observation is also documented in Altig et al. (2005) and Dupor et al. (2009) and Paciello (2009b). Paciello (2009a) develops a Rational Inattention model to explain this finding.



Grey line: the benchmark DSGE-VAR IRF (mode in bold and 90% interval).
Black line: the REE-DSGE IRF.

Fig. 1. IRF of a monetary policy and a productivity shock on inflation: DSGE versus DSGE-VAR.

Therefore, the transmission of the monetary policy shock will be much more sensitive to the formation of expectations than the transmission of the productivity shock, and it will not only depend on the expectations about future inflation but also on the expectations about future demand and cost developments. Secondly, productivity is an observable variable at the firm level and is central in the information set of the price-setting firm. On the other hand, monetary policy shocks and the derived demand effects are typical macro-variables and their effects build up slowly over time, so the impact on the firm's decisions will be much more sensitive to the correct interpretation of the information set. Under RE, the exact identification of the shocks and their transmission towards all the relevant variables benefits from the general equilibrium perspective and the implied cross-equation restrictions. This efficiency might be lost under learning and, as a consequence, expectations will respond less or more slowly to the shock. Finally, under learning, the expectations or the perceived law of motion will gradually adjust to the observed developments. Unexpected innovations in a specific variable will tend to affect the perceived mean and persistence of the underlying process. As a result, these adjustments in beliefs will reinforce the persistence of the response to the shocks. Orphanides and Williams (2003) show that adaptively learning agents would tend to increase inflation expectations in response to a positive inflation shock relative to the RE case. Updates in agents' beliefs are necessarily gradual, forcing a more gradual monetary policy response initially, and a slower return of inflation to the steady state relative to the RE benchmark, later on. Thus, adaptive learning might prove useful in matching the DSGE-VAR evidence referred to above.

3. Learning setup

3.1. Updating of beliefs

We implement the adaptive learning approach in the DYNARE toolbox which is developed to estimate and simulate DSGE models. The structural model is represented in the following way:

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

where the vector y_t includes endogenous variables of the model.[5] The model is driven by the exogenous processes x_t,

$$x_t = \rho x_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}.$$

Including the innovation ε_t in the vector of exogenous processes $w_t = (x_t^T, \varepsilon_t^T)^T$, the result can be written as an AR(1) process

$$w_t = \Gamma w_{t-1} + \Pi\varepsilon_t. \tag{2}$$

The solution of the model under rational expectations is provided by DYNARE as

$$\begin{bmatrix} y_t \\ w_t \end{bmatrix} = \mu + T\begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + R\varepsilon_t. \tag{3}$$

[5] Dynare is available for downloading at http://www.dynare.org. Version 3.064 of Dynare was used. The variable `jacobia_` contains the matrix $[A_0\ A_1\ A_2\ B_0]$. See Juillard (1996) for the original documentation.

Eq. (2) forms part of the system (3). Eq. (3) is only notationally different from the Minimum State Variable (MSV) solution, which for a system consisting of (1) and (2) is usually written as

$$y_t = a + by_{t-1} + cw_t.$$

The vector y contains a subset of state variables y^s and variables that appear with a lead in the model, y^f.[6] Deviating from the rational equilibrium (RE) assumption and following Marcet and Sargent (1989) and Evans and Honkapohja (2001), we assume that the agents forecast future values of the lead variables using a linear function of the states and exogenous driving processes,

$$y_t^f = \alpha_{t-1} + \beta_{t-1}^T \begin{bmatrix} y_{t-1}^s \\ w_t \end{bmatrix}, \tag{4}$$

where β_{t-1}^T does not necessarily coincide with the REE reduced form coefficients b and c, but the functional form of the relationship between y_t^f, y_{t-1}^s, and w_t exactly corresponds to the MSV REE reduced form.[7,8] Finally, the agents' beliefs about the reduced form coefficients α and β are updated using a constant-gain algorithm. The constant-gain algorithm is equivalent to Weighted Least Squares where the weight given to the more recent data points is higher. This weighting scheme allows agents to "forget" information from the distant past and this makes constant-gain learning preferable to Recursive Least-Squares (RLS) if agents believe that their environment is non-stationary.[9]

In our model, there are 12 forward-looking variables, 11 endogenous state variables, and 9 exogenous stochastic processes (with 2 moving average terms counted as exogenous processes). Therefore, α_{t-1} is a 12×1 vector, while β_{t-1} is a 20×12 matrix. Every period, the agents are updating their beliefs in a constant-gain RLS step:

$$\phi_t = \phi_{t-1} + gR_t^{-1} Z_{t-1} (y_t^f - \phi_{t-1}^T Z_{t-1})^T, \tag{5a}$$

$$R_t = R_{t-1} + g(Z_{t-1} Z_{t-1}^T - R_{t-1}). \tag{5b}$$

Here we denoted the data vector that the agents use in their regressions as $Z_t = (1, (y_{t-1}^s)^T, w_t^T)^T$, and gathered the beliefs into a single matrix $\phi^T = (\alpha, \beta^T)$.[10]

3.2. Initial beliefs

Eq. (5) allow us to track the agents' beliefs over time, if both the data and the initial beliefs are known. As it turns out that the results are very sensitive to the initial beliefs, we describe their selection in detail. We distinguish four ways of determining the initial beliefs: the first three are consistent with some REE, and the fourth is based on regression estimates with pre-sample data.

The first three ways of selecting the initial beliefs all use Eq. (6) below to calculate ϕ_0 and R_0. For any REE, given by (3), one could derive a matrix of second moments of the model variables, Ω. The moments imply a relation between the forward-looking variables y^f and the variables Z used in forecasting equation, $y_t^f = \phi_0^T Z_{t-1}$. Corresponding rows and columns of Ω also give us initial conditions for the second moments matrix R_0. The formulae for ϕ_0 and R_0 are given by

$$\phi_0 = E[Z_{t-1} Z_{t-1}^T]^{-1} \cdot E[Z_{t-1} (y_t^f)^T], \tag{6a}$$

$$R_0 = E[Z_{t-1} Z_{t-1}^T], \tag{6b}$$

where the expectations $E[\,]$ are derived using Ω. The difference between the three initialisation procedures lies in which REE, and thus which Ω matrix, is used to derive ϕ_0 and R_0.

Denote the parameter vector that is used to derive the model equations by θ. Denote $\tilde{\theta}$ a vector of parameters for an auxiliary model which generates matrix $\Omega(\tilde{\theta})$ that is then used for evaluating the initial beliefs in (6). The first three ways of deriving initial beliefs can then be characterised as follows:

- In the first procedure, $\theta = \tilde{\theta}$ at all times. In other words, initial beliefs always correspond to the REE which is implied by the estimated parameter vector θ which changes in the posterior maximisation and MCMC steps. We think that this procedure is the closest to the pure rational expectations, as the only source of differences with the REE is related to the temporary deviations of beliefs from their REE values caused by in-sample data fluctuations and the related stochasticity of the constant-gain learning.

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

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

[8] 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 behaviour of estimated DSGE model, see Milani (2006) and references therein.

[9] As discussed below, our results are not sensitive to the choice of constant-gain learning.

[10] We discuss whether the constant should be present in Z_t and ϕ below.

- In the second procedure, $\tilde{\theta}$ is fixed while θ changes. In principle, $\tilde{\theta}$ can be selected to be any parameter vector; in practice, here we take $\tilde{\theta}$ to be the posterior mode of the model estimated on pre-sample data running from 1955Q1 to 1964Q4 (and using the Rational Expectations assumptions). If pre-sample correlations of variables are very different from those in-sample, initial beliefs derived in this way would differ a lot from the ones consistent with the in-sample REE. This procedure has a natural interpretation of agents using prior information to initialise their forecasting functions.
- Finally, following Sargent et al. (2006), we use optimised initial beliefs. This means that during the posterior maximisation step, we optimise the posterior probability with respect to both $\tilde{\theta}$ and θ.[11] At the MCMC step, $\tilde{\theta}$ is kept fixed at the resulting posterior mode values, while θ changes. This is the third way of generating the initial beliefs. By disconnecting the initial beliefs completely from the pre-sample observations, this approach runs the risk of overfitting the data with initial beliefs being driven by spurious correlations in the sample data.

Regression-based initial beliefs, our fourth initialisation approach, are obtained by running a regression of y_t^f on Z_t using pre-sample data. Usually, we pick the point estimate rather than a random point from the distribution of regression estimates, as proposed by Carceles-Poveda and Giannitsarou (2007).

In all these procedures, we assume that the agents know both the law of motion (2) of the exogenous driving processes and the standard deviations of the innovations ε_t.[12] This is a standard assumption in the learning literature. An alternative approach is to estimate exogenous processes separately and then use the current beliefs about (2) in the updating step (5). We do not pursue this route here.

3.3. Beliefs and likelihood construction

In contrast to the low-dimensional models studied by Milani (2006, 2007), Sargent et al. (2006), or Vilagi (2007), our setup exhibits a clear distinction between the endogenous model variables and the observable variables which are used to estimate the model. The link between the two concepts is described by the measurement equations. All variables in the MSV REE solution are model variables which are not, in fact, observed. Therefore, we use the output from the Kalman filter, used to evaluate the likelihood function for a particular combination of parameters, on both sides of the updating Eq. (5).[13]

All endogenous model variables have zero means. Therefore, the MSV solution does not include the constant. Our baseline estimations take this into account and use $\tilde{Z}_t = ((y_{t-1}^s)^T, w_t^T)^T$ as the data vector. However, if we assume that the agents are also (implicitly) learning the values of the growth rates or the inflation target, we include the constant into (5).

Given current beliefs, it is possible to derive the value of $E_t y_{t+1}^f$ as a function of a constant, y_t, and w_t. One can then solve Eq. (1) for $(y_t^T, w_t^T)^T$ and derive a time-varying VAR representation of the model:

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

The values of μ_t, T_t, and R_t are then used to form expectations of the next period model variables in the Kalman filter. Thus, the estimation of a DSGE model under adaptive learning reduces to calculating a time-varying law of motion for the model and plugging it into the Kalman filter step, leaving the rest of the DYNARE toolbox largely untouched.

The learning setup just described allows an easy introduction of non-MSV learning. For example, we could allow our agents to forecast values of y_{t+1}^f using only observable variables, y_t^{obs}, or their model counterparts. Some of these variables are not in the state vector.[14] Given that DYNARE provides Eqs. (3) and (1) using the whole vector y, not just its state subset, derivation of μ_t, T_t, and R_t does not depend on using MSV or non-MSV solution. In the rest of the paper, we refer to this type of learning with misspecified beliefs as the VAR-learning case.

This type of under-parameterisation of the belief equations implies a much bigger departure from rational expectations. One might object to this assumption on the grounds that it does not fulfil a lower bound on rationality (Marcet and Nicolini, 1993), and that it provides too many degrees of freedom for the macromodeler. On the other hand, misspecified or under-parameterised expectations may be considered the rule rather than the exception, with agents selecting among parsimonious forecasting models based on their relative forecast performance (Branch and Evans, 2006; Adam, 2005).

[11] To make this way of initialising initial beliefs consistent with the assumption that the agents know parameters of exogenous processes, during the posterior maximisation step we restrict a subset of $\tilde{\theta}$ which describes standard errors of innovations and ARMA coefficients of exogenous processes w_t, to be the same as in θ.

[12] In the third way of selecting the initial beliefs, this means that the auxiliary model shares with the main model the parameter values related to the exogenous processes.

[13] In terms of Hamilton (1994), we use $\hat{y}_{t-1|t-1}^s$ and $\hat{w}_{t|t}$ on the right and $\hat{y}_{t|t}^f$ on the left. In principle, as time t progresses, the agents could revise their past filtered estimates and thus adjust values of ϕ_t 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 the past. This procedure would make better use of the available information; however, our current procedure uses only filtered estimates.

[14] The observable variables include first differences of consumption, output, investment, real wages, and levels of inflation, interest rates, and hours worked. These variables (together with a constant) could be used to construct the PLM, but none of them is in the state vector. Alternatively, we report estimations where we used model-based consumption, output, investment, real wages, inflation, interest rates, and hours worked in the PLM instead. In this set, hours are not in the state vector.

From the empirical perspective, there is also evidence in favour of small forecasting models. Orphanides and Williams (2005b), and Branch and Evans (2006) find that small-scale models provide a good fit for the inflation dynamics and approximate well the Survey of Professional Forecasters.

The procedure just described makes T_t a complicated function of the data, current parameters, and beliefs. There is nothing preventing T_t from being explosive (having one or several eigenvalues outside of a unit circle) for one or several periods. We follow the standard approach in the learning literature for handling this type of unstable dynamics by using a projection facility that excludes updating in such circumstances (see for instance Evans and Honkapohja, 2001).[15]

4. Simulation exercises

Before moving to the estimation of the models under learning, we illustrate the potential role of learning in these medium-scale DSGE models through a set of simulation experiments. We generate random data from the model under RE and under different learning setups and compare the properties of the simulated data to understand the impact of learning. Furthermore, we also consider the influence of the learning mechanism on the impulse response functions of the stochastic shocks.

Four different assumptions about the learning process are evaluated in these simulation experiments:

- the RE model as the benchmark model;
- MSV learning with perfect information about the constants;
- MSV learning where agents learn the constants as well; and
- VAR learning with beliefs about observed variables and a constant only.

In all these experiments, the structural parameters of the model, including the parameters of the exogenous stochastic processes, remain constant and equal to the mode of the estimated RE model. For each learning mechanism, we consider different values of the gain parameters (0.01, 0.02 and 0.05) corresponding roughly to a regression with forgetting half-length of 69, 34 and 14 periods.[16] In order to understand the influence of the initial beliefs on the simulation outcomes, we run 1000 simulations for each model; each simulation starts from a different initial PLM (initial beliefs) derived from the REE evaluated at a random draw from the posterior distribution of the estimated model. Each simulation run is 1000 periods long. We report the mean statistics for the first 150 observations and the last 150 observations in order to distinguish the learning dynamics during the "transition" period, directly dependent on the specific initialisation of the beliefs, from the dynamics due to the "permanent" time-variation generated by the learning process.[17] In all these simulations, we impose a projection facility and do not update the belief coefficients if the model dynamics become explosive under the new beliefs. Observations that push the eigenvalues of the system above one are therefore disregarded in the learning process.

4.1. Simulated second moments

In Table 2, we report the volatility and persistence of the generated data for output, inflation and the short rate. The first two columns describe the output for the RE model.

The outcomes for the MSV learning model with a small gain parameter (0.01) remain almost identical to the outcomes under RE. The additional variation induced by the beliefs updating does not significantly increase the volatility or the persistence of the observable variables. For small gain parameters, it makes no difference whether the constant is included in the belief regressions or not. For higher values of the gain parameter, the standard deviations and the autocorrelations of the simulated variables start to increase. This tendency is still moderate during the transition period, where the influence of the initial beliefs and the initial state vector remains very strong, but it becomes very pronounced for the permanent dynamics. The increase is also stronger if agents are learning about the constants in the model. These effects are still moderate for a gain of 0.02 (half forgetting in 8.5 years) but become very large if the gain parameter is further increased to 0.05 (only 3.5 years are needed to halve the weight of a data point in a regression). In relative terms, it is the volatility of inflation and the interest rate that experiences the strongest increase. In the simulations with a higher gain, the percentages of observations that are disregarded by the projection facility are also higher: while only 0.04% (0.05% for the model with a constant) of the observations are disregarded for a gain of 0.01, this percentage increases to 0.37% (0.69%) for a gain of 0.02 and 13.8% (24.14%) for a gain of 0.05. With a higher gain, the roots of the model are more frequently approaching one, which also explains why the standard deviations for the level variables tend to grow quickly in these cases. An additional source of increased variability for larger gain parameters are the so-called "exits" or "large

[15] See the Working Paper version of this paper (Slobodyan and Wouters, 2009) for a more detailed discussion of the particular form of our projection facility and other computational issues that occurred during the implementation of the learning dynamics.

[16] For a constant-gain learning with the gain parameter g, weight of a data point t periods ago is given as $g(1-g)^t$. This weight decreases by 50% in $T = -\frac{\ln 2}{\ln(1-g)} \approx \frac{0.69}{g}$ periods.

[17] Statistics based on the simulation of one long sample are close to the permanent statistics reported for the permanent dynamics here.

Table 2
Simulation results for different learning models and gain parameters.

	RE model	MSV beliefs		
		$\gamma = 0.01$	$\gamma = 0.02$	$\gamma = 0.05$
Standard deviations				
Δy	0.94/0.94	0.94/0.94	0.94/0.95	0.98/1.89
π	0.50/0.51	0.48/0.51	0.50/0.55	0.60/2.40
R	0.54/0.55	0.54/0.55	0.54/0.57	0.59/2.50
Autocorrelations				
Δy	0.27/0.28	0.26/0.28	0.26/0.28	0.30/0.46
π	0.82/0.82	0.80/0.83	0.81/0.83	0.83/0.81
R	0.89/0.89	0.89/0.90	0.89/0.89	0.91/0.91

	MSV beliefs+cte			VAR beliefs+cte		
	$\gamma = 0.01$	$\gamma = 0.02$	$\gamma = 0.05$	$\gamma = 0.01$	$\gamma = 0.02$	$\gamma = 0.05$
Standard deviations						
Δy	0.94/0.95	0.94/0.96	0.99/2.56	0.98/0.88	1.24/1.12	1.66/2.93
π	0.48/0.52	0.50/0.57	0.61/4.25	0.76/0.56	1.20/1.35	1.86/8.68
R	0.54/0.56	0.54/0.59	0.60/5.16	0.69/0.60	0.89/1.51	1.45/9.39
Autocorrelations						
Δy	0.26/0.28	0.26/0.29	0.31/0.52	0.39/0.30	0.51/0.39	0.59/0.55
π	0.80/0.83	0.81/0.83	0.86/0.84	0.85/0.84	0.86/0.88	0.89/0.93
R	0.89/0.90	0.89/0.90	0.93/0.93	0.92/0.90	0.93/0.94	0.96/0.97

deviations". As is often observed in simulated data under adaptive learning, occasionally time series can deviate from their long-run averages to a significant degree.[18]

Turning to the VAR-learning model, we observe significant deviations even for a small gain of 0.01. The standard deviations and the persistence increase for most of the variables especially during the transition dynamics. With higher gain, all standard deviations and correlations tend to increase and this increase is much more pronounced than under the MSV beliefs. The unit root bound and the projection facility are more frequently hit under VAR learning. The percentage of observations with the projection facility increases from 0.19% for a gain of 0.01 to 3.21% as the gain increases to 0.02 to 23.88% when the gain equals 0.05.

In Fig. 2(a) and (b), the typical behaviour of the simulated series are illustrated under different learning assumptions. The two variables shown are the output growth rate and the inflation rate. The MSV learning model with a small gain produces the standard stationary series. As the gain increases to 0.05, in addition to the extra volatility referred to above, we observe occasional large jumps or exits which quickly revert towards the neighborhood of the mean value. As Fig. 2(b) makes clear, under VAR-learning the exits could already be observed for a much smaller gain of 0.02.

A detailed analysis of the simulation outcomes reveal that the probability that the beliefs end up in the domain subject to frequent exits or projection facility hits depends, to some extend, on the nature of the initial beliefs. With initial beliefs derived from a parameter draw of the posterior distribution with a low likelihood, the chance that the projection facility is activated is higher. In other words, the higher the distance between the initial beliefs and the model-consistent ones, the higher the probability that the learning process will produce unstable dynamics. This finding also explains the computational issues that arise in the empirical applications where initial beliefs are allowed to deviate significantly from the model-consistent expectations.

4.2. Impulse responses with simulated beliefs

The impact of learning, and especially the role of the initial beliefs, on the model dynamics can also be illustrated by looking at the impulse response functions. Fig. 3(a) and (b) shows the IRF of inflation to a monetary policy and a productivity shock. The graphs contain the mode and the deciles of the distribution of the IRF evaluated at the start of the simulation where the beliefs are initialised through random draws from the posterior distribution of the RE model.[19] This distribution, therefore, illustrates the impact of imposing different initial beliefs on the model keeping everything else constant. For MSV learning, pseudo-impulse responses are located very close to the IRFs of the RE model. Similarly to Williams (2003), we observe very little effect of MSV learning with "realistically" distributed initial beliefs on the impulse responses or second moments of the simulated time series.

[18] In the Working Paper version, we discuss extensively the occurrence of these exit or large deviation dynamics.

[19] The graph contains pseudo impulse responses as in Williams (2003): given initial beliefs, a one-standard deviation shock is introduced, and behaviour of the model is then traced while the beliefs themselves change.

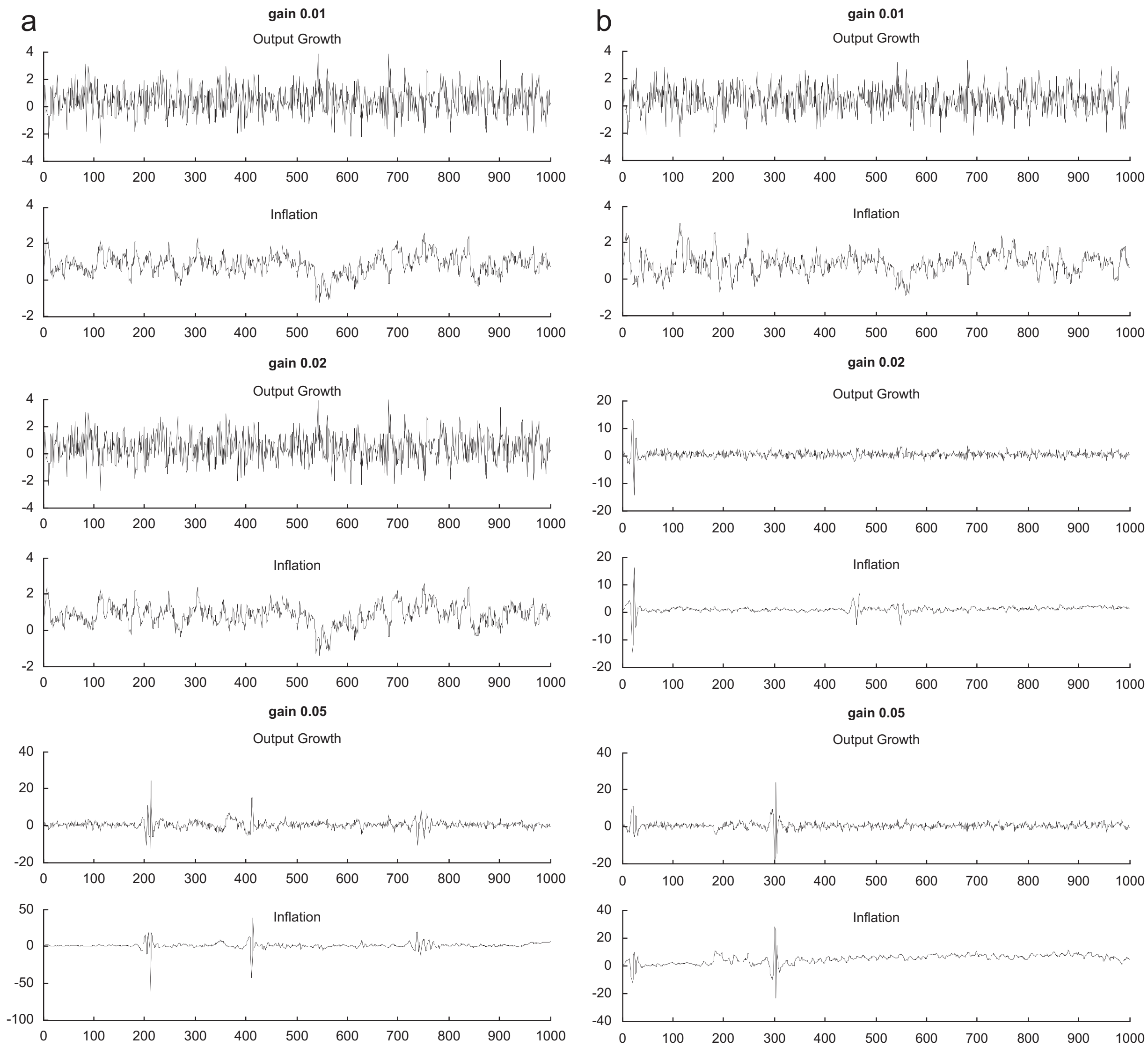


Fig. 2. (a) Simulation profile for MSV learning with different gains: gain 0.01, gain 0.02, gain 0.05 and (b) simulation profile for VAR learning and different gains: gain 0.01, gain 0.02, gain 0.05.

The impulse responses under the VAR learning differ from the IRFs under RE considerably more than under MSV learning and their decile dispersion is wider. In both cases, the dispersion of responses to the monetary policy shock decreases slower than for the productivity shock. Under VAR learning, the agents' Perceived Law of Motion (and therefore, their expectations) deviates substantially from the RE reduced form, which explains this difference. For a monetary policy shock, inflation expectations are affected much less initially, but the effect is longer lived afterwards. The same dynamic profile applies for the productivity shock but in that case the overall impact on inflation is relatively even stronger in the VAR learning case. These differences in the underlying beliefs are mainly due to two reasons. First, under VAR learning, the agents no longer include explicitly the monetary policy shock in their belief equations. Expectations will still react to the policy shock because the interest rate and the other observed variables are in the information set. But the precise nature of the shock is not immediately recognised and used in the expectations formation. The reaction of the expectations reflects only the perceived unconditional correlation between the forward variables and the observed information set, and not the correlation conditional on the specific shock. Given that inflation and the interest rate tend to be positively correlated on average, the negative influence of a policy tightening on inflation expectations is weakened. Second, the beliefs under VAR learning imply that agents assume that the inflation process is more persistent than under RE or MSV beliefs. The coefficient on past inflation in the inflation forecasting equation is close to 0.8 in the initial beliefs while it is around 0.2 for the REE. The resulting persistence in the realised inflation under VAR learning is higher than in the RE case. This higher persistence under VAR learning might substitute for the structural frictions in the model. The estimation exercises will provide more information on this issue.

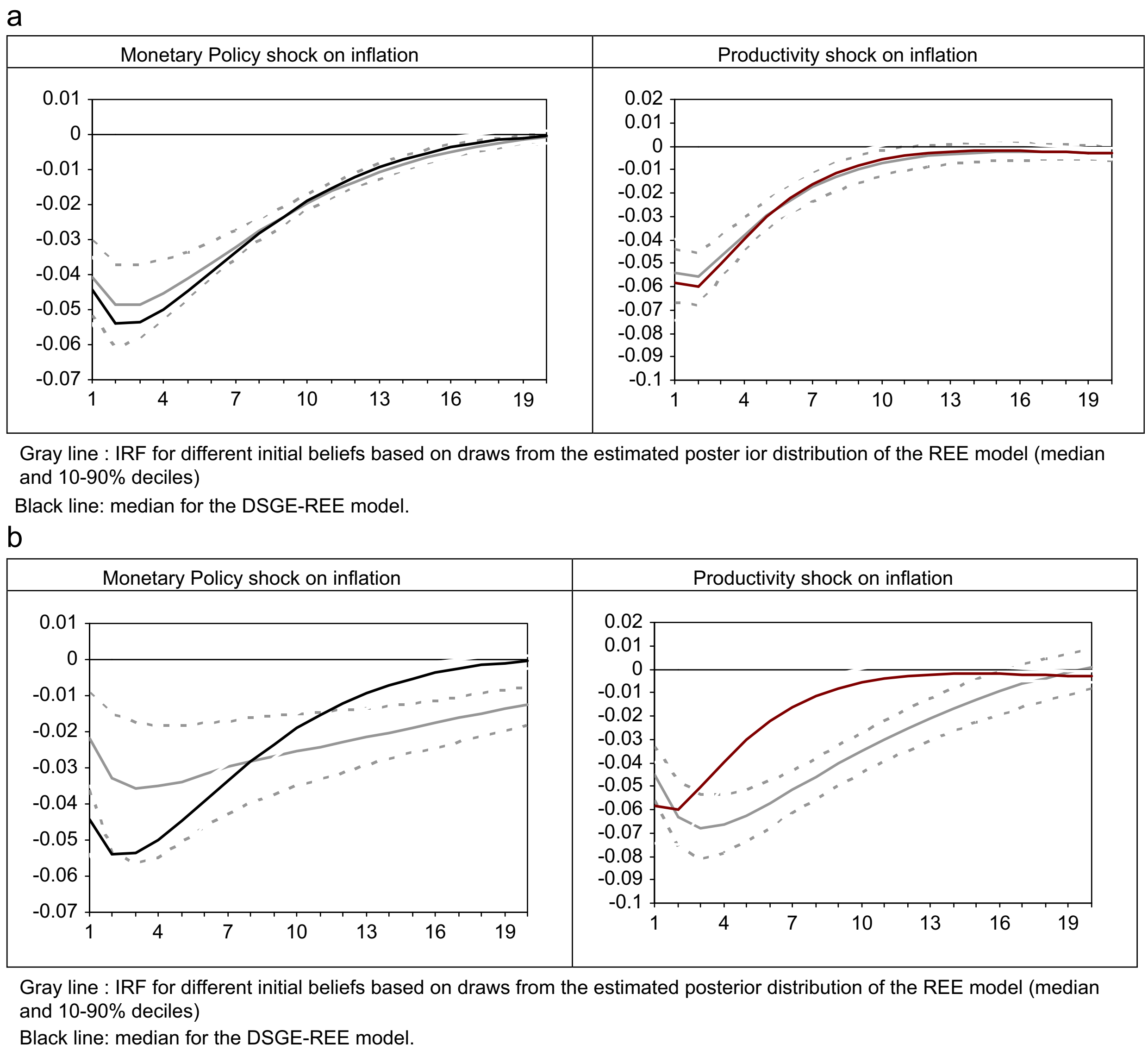


Fig. 3. (a) IRF for a monetary policy and a productivity shock on inflation: MSV learning for different initial beliefs (gain=0.02), and (b) IRF for a monetary policy and a productivity shock on inflation: VAR learning for different initial beliefs (gain=0.02).

The impact of subsequent learning on these IRFs is very minor: for the MSV learning and the gain values we consider, a one-standard error shock does not generate any significant change in the beliefs. With VAR learning and gains of 0.02 or 0.05 there is already a noticeable impact on the impulse responses; again, we attribute the difference to the different information content of agents' PLM under these two learning types, and correspondingly different expectations. An additional source of the difference is related to the fact that the RE model-based beliefs under VAR learning, in fact, are not correct even on average as the equilibrium beliefs are given by the so-called Restrictive Perceptions Equilibrium (RPE). Therefore, a temporary effect related to the evolution of the beliefs from their REE-based values towards RPE values may play a role, especially for relatively large gains. The temporary effect is illustrated in Fig. 4, where we track behaviour of the inflation-on-inflation belief coefficient during a simulation run with gain equal to 0.01. With MSV learning, this coefficient (and others) does not deviate much from its REE value; in contrast, the evolution from the RE-model-consistent beliefs towards the RPE values is apparent under the VAR learning. In addition, the beliefs are more volatile with VAR learning.

5. Estimation under MSV learning

In this section, we present the estimation results for the case of MSV learning, in which agents use the correct reduced form of the model to form expectations about future variables but have to learn the exact values of the parameters. We use constant-gain (perpetual) Least-Squares learning and estimate the gain parameter jointly with the rest of the model

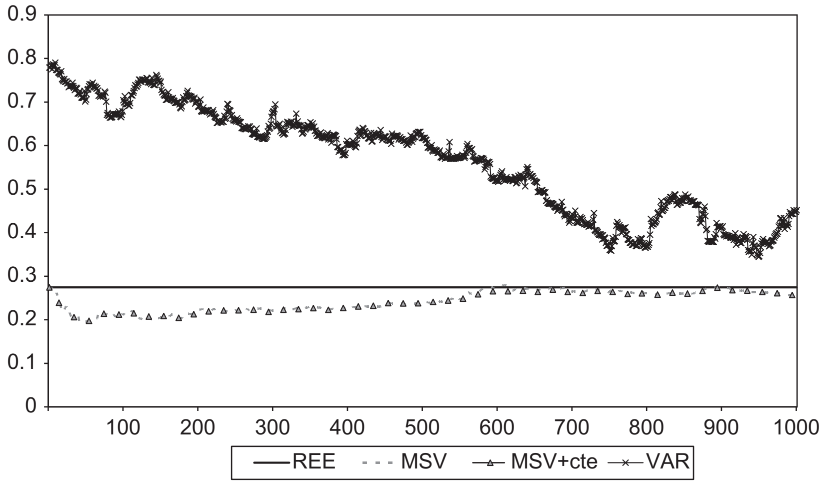


Fig. 4. Simulated belief coefficients under different learning mechanisms: coefficient of lagged inflation in the inflation belief regression (starting from the beliefs of the REE model and gain = 0.01).

Table 3
Comparison of the estimation outcomes for various learning models.

	Marginal likelihood	Gain parameter
Benchmark RE-DSGE model	−922.75	
DSGE-VAR model ($\lambda = 1$)	−897.78	
Learning models with MSV beliefs and initial beliefs defined as		
Model-consistent	−922.56	0.012 [0.001–0.034]
Model-consistent with constant	−922.65	0.012 [0.002–0.036]
Optimised	−910.97	0.017 [0.006–0.021]
Pre-sample	−944.37	0.024 [0.015–0.030]
Learning models with VAR beliefs and initial beliefs defined as		
Model-consistent with constant	−921.65	0.001 [0.000–0.004]
Optimised	−904.29	0.001 [0.001–0.003]
Pre-sample	−938.17	0.016 [0.011–0.021]

parameters. Results remain robust under standard Recursive Least-Squares instead of constant-gain learning. We distinguish between the case where agents have full information about the constants—parameters that determine the steady state deterministic growth rate, inflation rate and real interest rate, and the case where they also have to learn about these constants. In practice, this means that we consider belief regression with and without a constant term. In order to illustrate the sensitivity of the results to the assumptions about the initial beliefs, we consider the four alternative setups described in detail in Section 3.2.

In all these cases, the priors on the parameters are the same as in the RE model. As we also estimate the gain coefficient, there is one additional prior: a Gamma distribution with mean 0.035 and standard deviation 0.03. This implies that the prior mode for the gain is slightly less than 0.01, but the prior is quite uninformative so that the gain parameter can take on higher and lower values as well. The marginal likelihood and the estimated gain coefficient for the different model versions are summarised in Table 3. We concentrate most of the discussion on the models that yield a marginal likelihood that is similar or better than the RE model.

5.1. MSV learning with a model-consistent initialisation of beliefs

In this learning specification, the forecasting equations use the complete set of variables that make up the MSV solution of the model under RE, and the initial beliefs are consistent with the REE of the estimated model. The estimation results for the posterior distribution of the parameters and the marginal likelihood of the model are extremely close to the REE estimates that were presented in Table 1.[20]

[20] Comparing marginal likelihoods across models is complicated and can be very sensitive to the way the priors are defined (see Del Negro and Schorfheide, 2008). In this application however, the close similarity between the RE and the learning model allows us to conclude that their relative posterior probability is very close to the prior probability.

This similarity is not really surprising for two main reasons. First, the initial beliefs are consistent with the REE implied by the estimated model. And second, the information available to the agents for updating of the belief parameters comes close to the information available to the rational agents. In every period, the agents use currently best (filtered) estimates of all the variables appearing in the MSV solution; they also are assumed to know the parameters of the exogenous shock processes. The belief coefficients are updated by regressing the forward variables up to time t on the best estimates for the exogenous processes for time t and on the best estimates of the lagged values of the endogenous state variables. But given that the REE model does rather well in fitting the data, without any strong evidence of instability over sub-samples and without any remaining correlation in the estimated innovations, there is no reason for the model under learning to deviate systematically from the belief parameters implied by the RE model. As a result, the time-variation in the beliefs from which the learning model could benefit is negligible in this setup. Consistent with the above observations, the gain parameter is estimated rather imprecisely: the 90% probability interval from the posterior distribution for the gain are 0.001 and 0.034 with the mode around 0.012.

As can be seen from Table 3, inclusion of a constant in the belief equations, which reflects alternative assumption on the agents' knowledge about the model constants, does not matter in this setup. The estimated parameters and the marginal likelihood are insensitive to the presence of a constant in the belief equations.

As mentioned previously, we also checked for sensitivity of our results to the assumption of constant-gain learning, which implies that the agents believe their world to be non-stationary. In a stationary world, they would use OLS regression, which amounts to using the same updating equations (5), but with decreasing gain given as $1/(t_0+t)$ instead of a constant g. The constant t_0 is introduced in order to avoid large swings of beliefs at small values of t; initial gain $1/t_0$ is estimated instead of the constant-gain g, using the same prior. This Recursive Least-Squares (RLS) estimation delivers a marginal likelihood of -923.39, which is extremely close to the one under RE or model-consistent MSV learning. The estimated parameters also remain essentially the same. Initial gain $1/t_0$ equals 0.012, as in posterior mode of the constant gain estimation. Over the sample, the gain in RLS decreases to $1/(t_0+160)=0.004$; the whole [0.004;0.012] interval is thus within the 90% maximum probability interval of the model-consistent MSV constant-gain estimation which is [0.001;0.034]. We thus conclude that replacing constant-gain least-squares learning with RLS does not change the results.

5.2. MSV learning with optimised initial beliefs

Our second specification of the learning process derives the initial beliefs from a REE of a model which is not necessarily the same as the model estimated in-sample. These initial beliefs are chosen to optimise the in-sample fit of the model with learning. As described previously, we estimate two models simultaneously: the 'initial belief' model is used only to construct the initial beliefs, and the 'real' model is utilised to evaluate the data. Consistent with the hypothesis, retained in all our learning models, that economic agents know the parameters of the exogenous stochastic processes that drive the economy, we estimate only the behavioural parameters of the initial belief model and impose the same stochastic parameters in both models. We derive the initial beliefs from an alternative structural model in order to save on the number of estimated parameters.[21]

As one could expect, the model with optimised initial beliefs outperforms all other MSV-learning models in terms of marginal likelihood.[22] This setup also fits the data better than the RE model and reduces the gap with the DSGE-VAR model likelihood by half. We noted that the difference in the marginal likelihood was smaller than the difference between the posterior modes of these two models. The likelihood function of the model with optimised initial beliefs is characterised by an irregular surface that creates severe computational issues: the surface is extremely steep in many directions and the optimisation process gets easily stuck at some local optimum. In addition, the MCMC sampling converges very slowly. The relatively high value of the gain, which is estimated at 0.017 with a posterior distribution that varies between 0.006 and 0.021, is the prime source of these complications. A high gain implies that the coefficients of the forecasting equations are very volatile. Small changes in the parameters of the model can result in large adjustments to the updated beliefs, which can lead to extreme consequences for the likelihood. To illustrate the role of the high gain, we considered the same model with the gain fixed at a small value of 0.002. Most of the computational problems disappear in this case. Fixing the gain has no cost in terms of marginal likelihood, although the posterior mode of this model is considerably lower than for the one with the estimated gain. This result suggests that the benefit in terms of the marginal likelihood and the additional explanatory power are mainly derived from the specific initial beliefs (which differ from the REE), and less from the time-variation induced by adaptive learning.

The implications of the estimated parameters and beliefs for the IRF of the productivity and monetary policy shock are illustrated in Fig. 5. At the start of the sample, IRFs correspond to the estimated initial beliefs. The time-variation in the IRFs is driven by the updating of the beliefs. For comparison, the constant IRFs of the RE model are plotted on the same graph as a thick black line.

---

[21] For MSV learning without constant, dimensionality of the beliefs space is $20\times12=240$ coefficients in the forecasting functions plus $20\times19/2+20=210$ elements of the second moments matrix. Estimating a separate model for the beliefs adds only 20 extra parameters.

[22] This marginal likelihood is calculated conditional on the estimated beliefs: during the MCMC sampling process, we keep the initial beliefs fixed at the estimated mode. We use this procedure to make the marginal likelihood of this model directly comparable to the other models estimated in this paper.

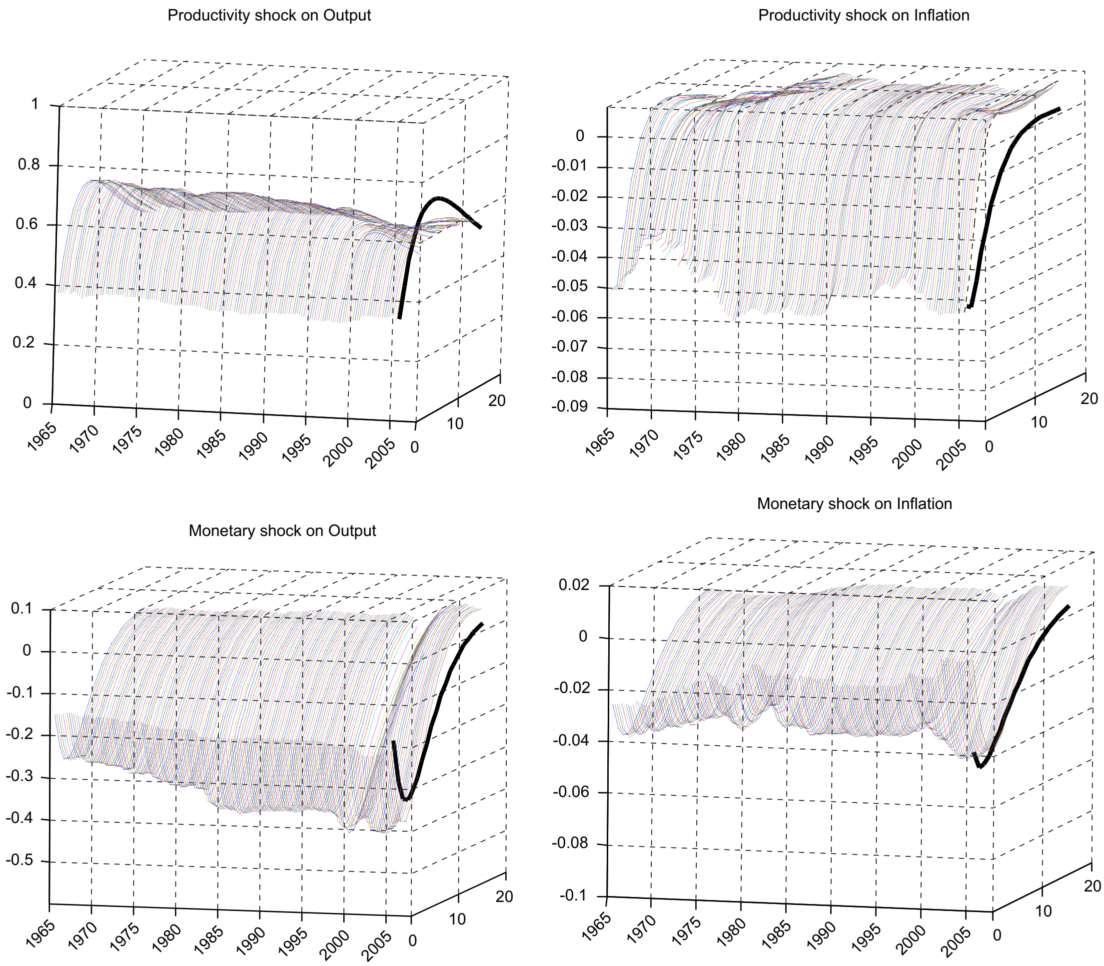


Fig. 5. Impulse Response Functions for the MSV model with optimised initial beliefs.

Although there are large changes in the underlying individual belief coefficients, the IRFs are relatively stable over time. For both shocks, the impact on output is in general similar to that under the RE. However, while for productivity the impact is declining slightly over time, the effect of the monetary policy shock tends to increase as time progresses. The effect of the productivity shock on inflation is more time-varying. The impact effect on inflation is similar to that in the RE model, but it disappears faster in the learning model than in the RE model. Exactly the opposite holds for the monetary policy shock on inflation: in the learning model, inflation tends to react more gradually, with the peak effect arriving several quarters later than in the RE model. This difference in the inflation response is especially striking at the end of the sample, where the contemporaneous reaction of inflation is almost zero and the peak effect occurs more than 1 year later (and definitely after the peak response in output).[23]

Note that this more gradual response of inflation to the monetary shock is not induced by the structural parameters that influence the inflation dynamics: relative to the model estimated under RE, the price and wage stickiness is lower while the indexation parameters are the same in the 'real' model with learning. On the other hand, the 'initial belief' model is characterised by higher stickiness and higher indexation. The gradual reaction response of inflation to the policy shock is further enforced by the updating of the beliefs over time. It is also very interesting to see how the beliefs tend to postpone the reaction of inflation to the monetary policy shock while they lower the persistence in the inflation reaction following the productivity shock.

Milani (2006) finds that the structural inertia in his model is systematically reduced when learning is introduced. We can only partially confirm his results: the estimated price and wage stickiness and the adjustment cost of investment are all lower under MSV learning, the indexation parameters are more or less unchanged, but the habit persistence coefficient

[23] The impact of the optimised beliefs under MSV learning is much stronger than that discussed in the simulation session (see Fig. 3(a)), where the beliefs were drawn from the posterior distribution of the RE model. The estimated "initial belief" parameters do actually deviate considerably from the RE model.

is larger with learning. The persistence of the exogenous shocks also stays largely the same with the exception of the persistence of the investment shock which becomes much lower.

5.3. MSV beliefs with pre-sample model initialisation

An alternative way to derive the initial coefficients of the MSV beliefs is to estimate a RE model over a sample that precedes the actual estimation period. Therefore, we estimated a RE model over the pre-sample period (1948:2–1964:4), and used the moment matrices implied by the resulting REE to fix the initial beliefs for the estimation of the MSV learning model over the sample (1966:1–2005:4).

Restricting the initial beliefs to be consistent with the pre-sample REE reduces the estimated marginal likelihood relative to our benchmark case with in-sample REE-consistent beliefs (see Table 3). The estimated gain is quite high with a mode at 0.024. As explained in Section 5.2, such a high gain generates a complex likelihood function that leads to important computational problems. MSV learning tends to generate second-moment matrices with tiny smallest eigenvalues, necessitating usage of the ridge correction mechanism. In such cases, beliefs are usually strongly adjusted in the early in-sample periods, and overfitting of the initial data might become an issue.

5.4. MSV beliefs with pre-sample regression initialisation

The fourth method of generating initial beliefs uses a least-squares regression of the forecasting equations using the pre-sample data. These regressions include a set of unobserved variables that can only be produced by an estimated model. We use the filtered series from the pre-sample RE model to generate the data needed to run the regressions.

Under this approach, the initial beliefs are such that beliefs adjust even more early on in the sample than with the pre-sample REE model initialisation from Section 5.3. Depending on the length of the pre-sample, the estimated gain parameter is extremely high, varying between 0.04 and 0.06, and this again leads to various computational difficulties. For example, optimisation routines are finding local rather than global optima, MCMCs do not converge or converge very slowly, and the approximation of the marginal likelihood yields very low values relative to the mode of the likelihood.

6. Estimation with VAR learning

Up to now, we have considered models in which private agents know the correct specification of the model, but learn about the values of the belief parameters. In the VAR learning specification, we drop the assumption that private agents know the correct specification, and instead assume that agents use only the limited list of variables in their belief equations. The potential implications of this misspecification in the beliefs on the model dynamics and the implied model dynamics have been discussed in the simulation exercises already. We assume that agents use the same list of seven observable variables as we do in the estimation of the model. In order to form expectations about the forward variables in the model, agents run regressions on the seven observed variables and a constant. However, we assume that these regressions are specified in levels and not in first differences, which imply that agents use the filtered values for the level variables of the observables, while we use the first differences of these variables in our measurement equations. In this section we only consider applications in which the belief equations contain a constant, meaning that private agents have to learn not only on the slope parameters of the belief regression but also the levels which depend on the steady-state inflation, growth and real interest rate. As in the case of MSV learning, we again consider different ways of constructing the initial beliefs at the start of the sample.

6.1. VAR learning with model-consistent initial beliefs

The estimated parameters for this variant are summarised in Table 4, while Fig. 6 shows the implied IRF functions.

First of all, VAR learning with a model-consistent initialisation of the beliefs generates a marginal likelihood that is slightly higher than that for the benchmark MSV learning and the REE model. However, contrary to the model with MSV learning and model-consistent initial beliefs, where the estimated learning gain is quite high and estimated imprecisely, the estimated gain for the misspecified VAR beliefs is extremely small with a narrow posterior distribution. The other estimated parameters remain very close to the ones in the benchmark and the REE model. One exception is the very low degree of price indexation.

The small gain parameter implies that the IRFs remain stable over time. The output reaction is relatively similar to the REE model. Relative to the MSV learning case, the response of inflation to the productivity shock is further enhanced but remains quite short-lived. For the monetary policy shock, the reaction of inflation is quite different: the overall response is small, and the impact effect is decreasing over time while the persistence in the response is increasing over time. Both trends are similar to those observed for MSV learning with optimised initial beliefs.

Two arguments help to understand the low estimate for the gain coefficient. Numerical simulations of the E-stability of the ordinary differential equation (ODE) under VAR learning with constants show that for parameter values close to the REE and the initial beliefs close to the RE-model-consistent ones, dynamics under learning are expected to remain stable.[24]

[24] See the WP version for a discussion of the E-stability and mean dynamics of this VAR learning model.

Table 4
Posterior estimates for learning models.

	MSV beliefs		VAR beliefs		VAR beliefs	
	Optimised initialisation		RE-initialisation		Optimised initialisation	
	Mean	Initial belief	Mean	Initial belief	Mean	Initial belief
St. dev. of the innovations[a]						
σ_a	0.47		0.46		0.47	
σ_b	0.25		0.29		0.27	
σ_g	0.53		0.53		0.53	
σ_q	0.61		0.92		0.93	
σ_r	0.24		0.25		0.23	
σ_p	0.14		0.08		0.07	
σ_w	0.23		0.28		0.29	
Persistence of the exogenous processes: $\rho = AR(1)$, $\theta = MA(1)$						
ρ_a	0.96		0.95		0.95	
ρ_b	0.23		0.22		0.14	
ρ_g	0.96		0.98		0.97	
ρ_q	0.33		0.54		0.55	
ρ_r	0.05		0.15		0.16	
ρ_p	0.88		0.88		0.93	
ρ_w	0.95		0.95		0.80	
θ_p	0.78		0.59		0.59	
θ_w	0.64		0.81		0.58	
a_g[b]	0.41		0.51		0.52	
Structural parameters of endogenous decisions						
φ	4.33	3.81	4.18		3.74	3.02
σ_c	0.99	1.30	1.34		1.06	1.14
η	0.80	0.69	0.73		0.74	0.73
σ_l	1.61	1.97	2.03		2.31	1.44
ξ_p	0.63	0.77	0.68		0.62	0.62
ξ_w	0.63	0.68	0.72		0.73	0.57
ι_p	0.22	0.51	0.14		0.10	0.26
ι_w	0.53	0.59	0.55		0.53	0.43
ψ	0.38	0.62	0.49		0.49	0.40
Φ_p	1.60	1.25	1.63		1.59	1.37
ρ_R	0.84	0.64	0.81		0.86	0.57
r_π	1.91	1.70	1.98		1.84	1.70
r_y	0.13	0.03	0.09		0.12	0.04
$r_{\Delta y}$	0.19	0.15	0.22		0.20	0.17
$\overline{\pi}$	0.61	0.57	0.79		0.75	0.74
β[c]	0.26	0.25	0.16		0.21	0.21
$\overline{l}$	0.92	1.26	0.33		−0.54	−0.49
γ	0.43	0.45	0.42		0.42	0.42
α	0.18	0.19	0.19		0.18	0.26

[a] The Inverse Gamma distribution is defined by the degree of freedom.
[b] The effect of TFP innovations on exogenous demand.
[c] The estimated coefficient is equal to $100(\beta^{-1}-1)$.

However, divergence is often observed if the simulations are started with beliefs about constants which are not equal to zero. Simulations show that beliefs about constants are especially volatile. Therefore, if some data point makes the agents assume a particularly large constant value in some forecasting equations, further evolution of the beliefs might become unstable. In the estimation procedure, such outcomes would produce very low likelihood values. The very low value of the estimated gain guarantees that within the 160 periods (length of the sample used for estimation) beliefs about constants do not reach values which are likely to trigger instability. An alternative explanation for the low estimated gain is that the average evolution of beliefs from their RE-model-consistent values towards the RPE values is inconsistent with the data in our sample; a low gain then guarantees that the beliefs do not move in this undesirable direction too fast.

As in the MSV case, we also performed a RLS estimation for the model-consistent VAR learning. Once again, we find that this learning algorithm delivers results very similar to constant-gain learning: the marginal likelihood is -921.44, the estimated initial gain $1/t_0$ equals 0.001, and over the sample interval the gains are all within the 90% maximum probability interval of the constant-gain estimation.

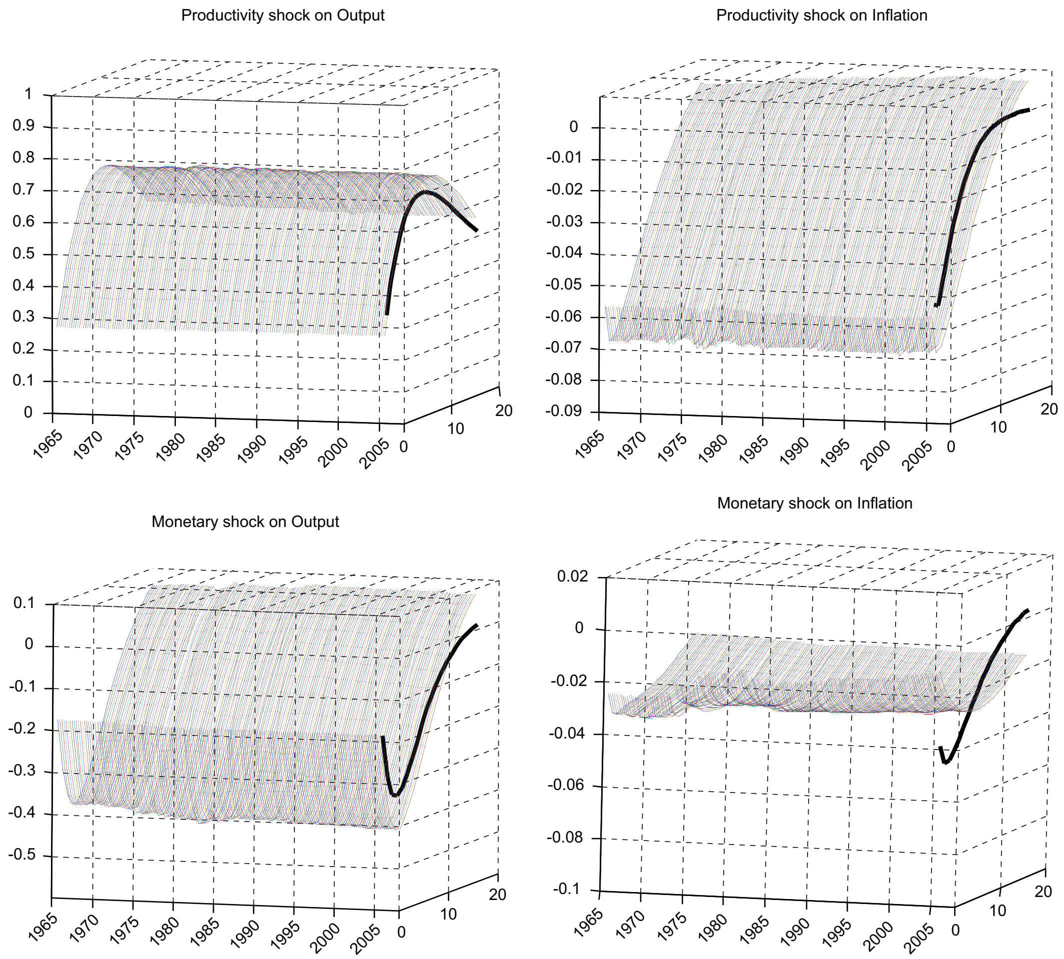


Fig. 6. Impulse response functions for the VAR beliefs with model consistent initial beliefs.

6.2. VAR learning with optimised initial beliefs

This specification of the learning process produces the best marginal likelihood and outperforms substantially the REE model (see Table 3). The marginal likelihood comes close to the values that are produced by the best fitting DSGE-VAR model. The structural parameters of the model are again close to the REE model and the benchmark learning model. The gain parameter is estimated to be very small and varying between 0.001 and 0.003 (Fig. 7).

The IRFs are relatively stable and similar to those of VAR learning with model-consistent beliefs. The inflation response to the monetary policy shock is again interesting: while in the beginning of the sample the response is very close to those of the REE model, the reaction gradually adjusts, driven by the beliefs updating, and at the end of the sample the IRF again becomes much more gradual and persistent and resembles closely the typical inflation response to the monetary policy shock in the DSGE-VAR model.

When looking at the individual belief coefficients of the inflation expectations, two trends can be observed: the positive coefficient of lagged inflation tends to decline over time, while the negative coefficient to the lagged interest rate increases. Both trends are important to understand the time-variation in the inflation response to the monetary policy shock. A higher but negative value for the coefficient on the interest rate in the inflation expectations reduces the negative impact effect of the policy tightening. The impact of the policy shock is further reduced if the perceived inflation persistence declines.

The more pronounced reaction of inflation to the productivity shock under VAR learning is also induced by the restrictions on the belief equations implicit in the set of variables the agents are using for forecasting, as the RE model with the parameters estimated under VAR learning does not produce a higher impact effect. The relevant belief coefficients remain quite stable over time as one would expect with small gain parameters. But the models with optimised initial beliefs tend to generate slightly more time-variation than the setups with model-consistent beliefs. This applies for both the MSV and the VAR learning exercises. One interpretation for this result is that optimised initial beliefs work by producing initial second-moments matrices which, combined with constant-gain updating, make it possible to capture more accurately the time-varying autocorrelations of the observable variables.

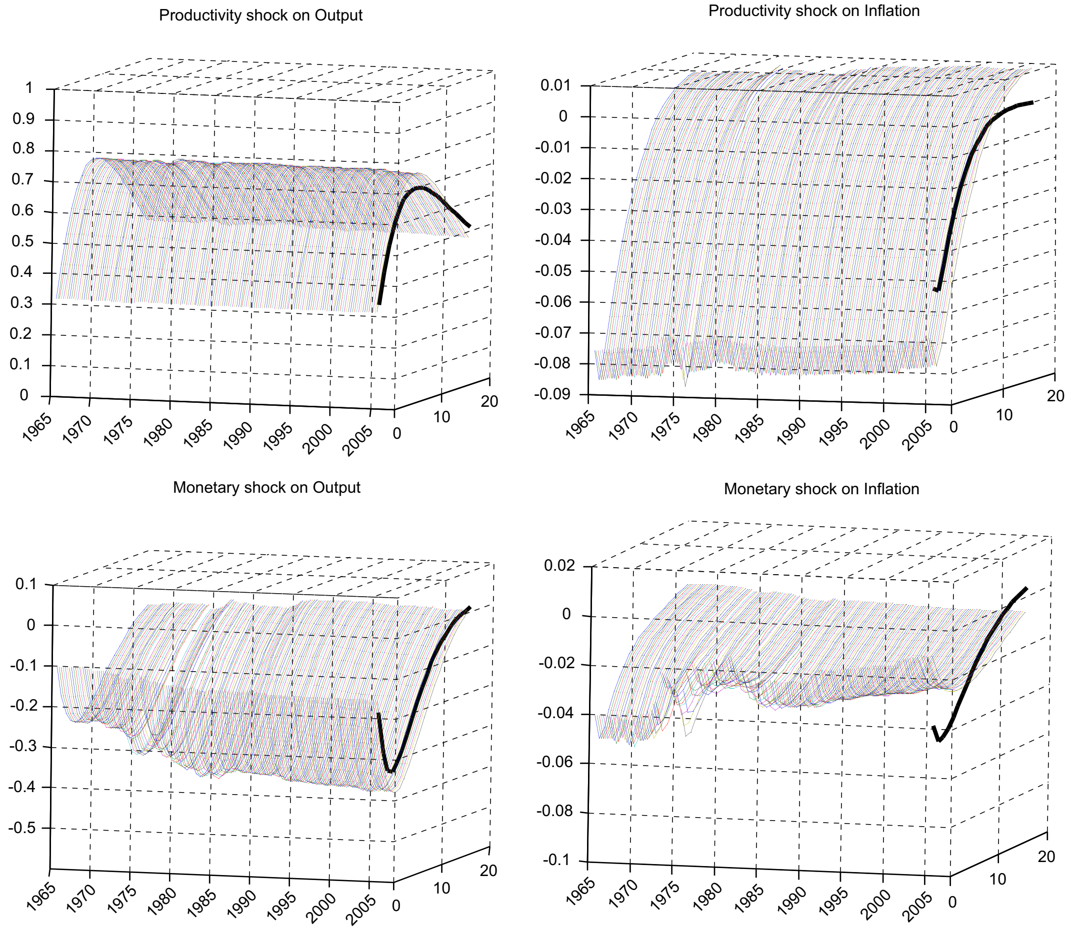


Fig. 7. Impulse response functions for the VAR beliefs with optimised initial beliefs.

The high marginal likelihood of this model delivers strong evidence in favour of beliefs that deviate significantly from the rational expectations hypothesis of model-consistent expectations. Re-estimation of this model with initial optimised beliefs that are kept constant over the rest of the sample delivers a very similar marginal likelihood. This suggests that as regards the improved model fit, the initial beliefs are more important than the beliefs updating. The small gain also suggests that the beliefs are not materially changed according to the constant-gain learning schedule. However, one should note that the updating is more important under the VAR beliefs than under the MSV beliefs even for very low gain parameters, because the residuals in the forecasting equations are larger and probably contain a more systematic component than under the MSV beliefs. This conjecture is also confirmed by the important changes in the IRF of inflation to the monetary policy shock. Therefore, the relative role of the restricted initial belief on the one hand and the time-variation in the belief coefficients on the other hand remains an open issue and it will be interesting to test whether alternative and more efficient learning procedures (such as Kalman filter) can provide more information on this.[25]

6.3. VAR learning with pre-sample-based initial beliefs

The VAR learning with initial beliefs derived from the RE model estimated using the pre-sample data does a poor job in terms of marginal likelihood. As for the model with MSV learning and pre-sample-based beliefs, the learning gain is very high (0.016) in order to allow the beliefs to adjust and to move away from the imposed initial beliefs. We also checked whether the choice of the pre-sample period was responsible for the observed divergence between the initial beliefs and the beliefs consistent with the in-sample data. If we use only data from 1955 up to 1965 to construct initial beliefs (instead of 1948–1965), the model fit improves. We conclude that the data points before 1955 affected the data-generating process

[25] Slobodyan and Wouters (forthcoming), applying Kalman filter learning, confirm indeed a more important role for time-variation in the belief coefficients.

in a way that is at odds with the one that prevailed in later periods, at least under the assumption that the agents are to include only the seven observable variables in their forecasting equations.

7. Conclusions

The above results illustrate that several of the models with learning fit the data equally well or even better than the RE model. The best-performing learning models generate marginal likelihoods that come close to that of the optimal DSGE-VAR model, which proves that the model-consistent expectations imposed by the rational expectations hypothesis are too restrictive. Specific initial beliefs contribute significantly to this result: the best-performing models are the ones where the initial beliefs are optimised to explain the in-sample data, consistent with previous findings in the literature. Restricting the set of variables used in the forecasting equations to a limited list of observed macro-variables generates learning models that explain the data better than models with MSV beliefs that use the complete set of observed and unobserved state variables implied by the REE. We generally observe a considerable updating in the belief equations in response to the systematic forecast errors. However, the best-fitting models tend to have a rather low estimated gain, which confirms that it is the initial beliefs, different from the RE-model consistent ones, that bring about improvements in the marginal likelihood, rather than the beliefs updating of the constant-gain learning type per se. On the other hand, choosing initial beliefs that are too far removed from the optimal ones leads to higher estimated gains that facilitate evolution towards better forecasting equations. Our results leave open a possibility that using alternative learning algorithms such as the Kalman filter, which can converge faster than constant-gain learning (see Sargent and Williams, 2005) could modify the tentative conclusion that initial beliefs are more important than the learning process itself.

In terms of IRFs, our discussion was limited to the implications for output and inflation of a productivity and a monetary policy shock. The implications of the other shocks still need to be documented. The implications as regards the productivity and the monetary policy shock are very promising: the learning models are able to generate an inflation response to productivity shocks that is very rapid and short-lived, while the response to monetary shocks is slow but very persistent. These findings also overcome some of the major shortcomings of the DSGE models, as indicated by the DSGE-VAR methodology for identifying misspecification. Having forecasting equations that differ significantly from those implied by the REE seems to be the key to this result.

The additional dynamics that are introduced by the learning process do not systematically alter the estimated structural parameters of the DSGE model. This result contradicts earlier claims in the literature, in particular Milani (2007), but is again in line with the results from the DSGE-VAR methodology which indicate misspecification but no systematic bias in the structural parameters.

Acknowledgments

We would like to thank the editor, two anonymous referees, and participants in multiple conferences and seminars for their suggestions and comments. First author expresses sincere gratitude to the NBB for hospitality and support which made this project possible.

Appendix A. Data appendix

The model is estimated using seven key macroeconomic 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 database. Real Gross Domestic Product is expressed in Billions of Chained 1996 Dollars. Nominal Personal Consumption Expenditure and Fixed Private Domestic Investment are deflated by the GDP deflator. Inflation is the first difference of the log of the Implicit GDP Price Deflator. 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 by the Civilian Employment (16 years and over)). The aggregate real variables are expressed per capita by dividing by 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.

Appendix B. Model appendix

In this appendix, we summarise the log–linear equations of the model. For a more detailed presentation, we refer to the discussion in SW 2007.

- Consumption Euler equation for the non-separable utility function:

$$\widehat{c}_t = c_1 E_t[\widehat{c}_{t+1}] + (1-c_1)\widehat{c}_{t-1} + c_2(\widehat{L}_t - E_t[\widehat{L}_{t+1}]) - c_3(\widehat{R}_t - E_t[\widehat{\pi}_{t+1}] + \widehat{\varepsilon}_t^b),$$

with $c_1 = 1/(1+\overline{\eta})$, $c_2 = c_1(\sigma_c - 1)(wL/C)/\sigma_c$, $c_3 = c_1(1-\overline{\eta})/\sigma_c$ where $\overline{\eta}$ is the external habit parameter adjusted for trend growth $\overline{\eta} = (\eta/\gamma)$, σ_c is the inverse of the intertemporal elasticity of substitution. $\widehat{\varepsilon}_t^b$ is the exogenous AR(1) risk premium process.

- Investment Euler equation:

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

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

- Value of the capital stock:

$$\widehat{Q}_t^k = -(\widehat{R}_t - E_t[\widehat{\pi}_{t+1}] + \widehat{\varepsilon}_t^b) + q_1 E_t[r_{t+1}^k] + (1-q_1)E_t[Q_{t+1}^k],$$

with $q_1 = r_*^k/(r_*^k + (1-\delta))$ where r_*^k is the steady-state rental rate to capital, and δ is the depreciation rate.

- Aggregate demand equals aggregate supply:

$$\widehat{y}_t = \frac{c_*}{y_*}\widehat{c}_t + \frac{i_*}{y_*}\widehat{i}_t + \widehat{\varepsilon}_t^g + \frac{r_*^k k_*}{y_*}\widehat{u}_t = \Phi_p(\alpha\widehat{k}_t + (1-\alpha)\widehat{L}_t + \widehat{\varepsilon}_t^a),$$

with Φ_p reflecting the fixed costs in production which corresponds to the price markup in steady state. $\widehat{\varepsilon}_t^g$, $\widehat{\varepsilon}_t^a$ are the AR(1) processes representing exogenous demand components and the TFP process.

- Price-setting under the Calvo model with indexation:

$$\widehat{\pi}_t - \imath_p \widehat{\pi}_{t-1} = \pi_1(E_t[\widehat{\pi}_{t+1}] - \imath_p\widehat{\pi}_t) - \pi_2\widehat{\mu}_t^p + \widehat{\varepsilon}_t^p,$$

with $\pi_1 = \overline{\beta}\gamma$, $\pi_2 = (1-\xi_p\overline{\beta}\gamma)(1-\xi_p)/[\xi_p(1+(\Phi_p-1)\varepsilon_p)]$, with ξ_p and $\imath_p$ respectively the probability and indexation of the Calvo model, and ε_p the curvature of the aggregator function. The price markup $\widehat{\mu}_t^p$ is equal to the inverse of the real marginal cost $\widehat{mc}_t = (1-\alpha)\widehat{w}_t + \alpha\widehat{r}_t^k - \widehat{\varepsilon}_t^a$.

- Wage-setting under the Calvo model with indexation:

$$\widehat{\pi}_t^w - \imath_w\widehat{\pi}_{t-1} = \pi_1(E_t[\widehat{\pi}_{t+1}^w] - \imath_w\widehat{\pi}_t) - \pi_3\widehat{\mu}_t^w + \widehat{\varepsilon}_t^w,$$

with $\pi_3 = (1-\xi_w\overline{\beta}\gamma)(1-\xi_w)/[\xi_w(1+(\phi_w-1)\varepsilon_w)]$ and wage markup $\widehat{\mu}_t^w = \widehat{w}_t - w_1\widehat{c}_t + (1-w_1)\widehat{c}_{t-1} - \sigma_l\widehat{L}_t$ with $w_1 = 1/(1-\overline{\eta})$. Both the price and wage markup shocks $\widehat{\varepsilon}_t^p, \widehat{\varepsilon}_t^w$ follow an ARMA(1,1) process.

- Capital accumulation equation:

$$\widehat{\overline{k}}_t = k_1\widehat{\overline{k}}_{t-1} + (1-k_1)\widehat{i}_t + k_2\widehat{\varepsilon}_t^q,$$

with $k_1 = (1-(i_*/\overline{k}_*))$, $k_2 = (i_*/\overline{k}_*)(1+\overline{\beta}\gamma)\gamma^2 S''$. Capital services used in production is defined as: $\widehat{k}_t = \widehat{u}_t + \widehat{\overline{k}}_{t-1}$

- Optimal capital utilisation condition:

$$\widehat{u}_t = (1-\psi)/\psi\widehat{r}_t^k,$$

with ψ is the elasticity of the capital utilisation cost function.

- Optimal capital/labor input condition:

$$\widehat{k}_t = \widehat{w}_t - \widehat{r}_t^k + \widehat{L}_t,$$

- Monetary policy rule:

$$\widehat{R}_t = \rho_R\widehat{R}_{t-1} + (1-\rho_R)(r_\pi\widehat{\pi}_t + r_y(\widehat{ygap}_t) + r_{\Delta y}\Delta(\widehat{ygap}_t)) + \widehat{\varepsilon}_t^r$$

with $ygap_t = (\widehat{y}_t - \widehat{y}_t^{flex})$ where $\widehat{y}_t^{flex}$ refers to the output level in the flexible price and wage economy, abstracting from markup shocks.

The following parameters are not identified by the estimation procedure and are therefore calibrated: $\delta = 0.025$, $\varepsilon_p = 10$, $\varepsilon_w = 10$, $\phi_w = 1.5$.

References

Adam, K., 2005. Learning to forecast and cyclical behavior of output and inflation. Macroeconomic Dynamics 9, 1–27.

Altig, D., Christiano, L., Eichenbaum, M., Linde, J., 2005. Firm-specific capital, nominal rigidities and the business cycle. NBER Working Paper Series no. 11034.

Branch, W., Evans, G., 2006. A simple recursive forecasting model. Economics Letters 127, 264–295.

Carceles-Poveda, E., Giannitsarou, C., 2007. Adaptive learning in practice. Journal of Economic Dynamics and Control 31, 2659–2697.

Christiano, L.J., Eichenbaum, M., Evans, C.L., 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of Political Economy 113, 1–45.

Del Negro, M., Schorfheide, F., 2008. Forming priors for DSGE models (and how it affects the assessment of nominal rigidities). Journal of Monetary Economics 55, 1191–1208.

Del Negro, M., Schorfheide, F., Smets, F., Wouters, R., 2007. On the fit of new Keynesian models. Journal of Business & Economic Statistics 25, 123–143.

Dupor, B., Han, J., Tsai, Y.-C., 2009. What do technology shocks tell us about the New Keynesian Paradigm? Journal of Monetary Economics 56, 560–569

Eichenbaum, M., Fisher, J., 2007. Estimating the frequency of Re-optimization in Calvo-style models. Journal of Monetary Economics 54, 2032–2047.

Evans, G.W., Honkapohja, S., 2001. Learning and Expectations in Macroeconomics. Princeton University Press.

Hamilton, J., 1994. Time Series Analysis. Princeton University Press.
Honkapohja, S., Mitra, K., Evans, G.W., 2002. Notes on agents' behavioral rules under adaptive learning and recent studies of monetary policy. Mimeo.
Juillard, M., 1996. Dynare: a program for the resolution and simulation of dynamic models with forward variables through the use of a relaxation algorithm. CEPREMAP, Couverture Orange, 9602.
Kimball, M.S., 1995. The quantitative analytics of the basic neomonetarist model. Journal of Money, Credit, and Banking 27, 1241–1277.
Marcet, A., Nicolini, J.P., 1993. Recurrent hyperinflations and learning. American Economic Review 93, 1476–1498.
Marcet, A., Sargent, T.J., 1989. Convergence of least squares learning mechanisms in self-referential linear stochastic models. Journal of Economic Theory 48, 337–368.
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.
Milani, F., 2007. Expectations, learning and macroeconomic persistence. Journal of Monetary Economics 54, 2065–2082.
Orphanides, A., Williams, J.C., 2003. Imperfect knowledge, inflation expectations, and monetary policy. NBER Working Paper No. W9884.
Orphanides, A., Williams, J.C., 2005a. Inflation scares and forecast-based monetary policy. Review of Economic Dynamics 8, 498–527.
Orphanides, A., Williams, J.C., 2005b. The decline of activist stabilization policy: natural rate misperceptions, learning and expectations. Journal of Economic Dynamics and Control 29, 1927–1950.
Paciello, L., 2009a. Monetary policy activism and price responsiveness to aggregate shocks under rational inattention. Mimeo, Einaudi Institute for Economics and Finance, Rome.
Paciello, L. 2009b. Does inflation adjust faster to aggregate technology shocks than to monetary policy shocks? Mimeo, Einaudi Institute for Economics and Finance, Rome.
Preston, B., 2005. Learning about monetary policy rules when long-horizon expectations matter. International Journal of Central Banking 1 (2).
Sargent, T., Williams, N., 2005. Impacts of priors on convergence and escape from nash inflation. Review of Economic Dynamics 8, 360–391.
Sargent, T., Williams, N., Zha, T., 2006. Shocks and government beliefs: the rise and fall of American inflation. American Economic Review 96, 1193–1224.
Slobodyan, S., Wouters, R., 2009. Learning in an estimated medium-scale DSGE model. CERGEI-EI Working Paper No. 396.
Slobodyan, S., Wouters, R. Estimating a medium-scale DSGE model with expectations based on small forecasting models. American Economic Journal: Macroeconomics, forthcoming.
Smets, F., Wouters, R., 2003. An estimated dynamic stochastic general equilibrium model of the Euro area. Journal of the European Economic Association 1, 1123–1175.
Smets, F., Wouters, R., 2007. Shocks and frictions in US business cycles: a Bayesian DSGE approach. American Economic Review 97, 586–606.
Vilagi, B., 2007. Adaptive learning and macroeconomic persistence: comparing DSGE models of the Euro Area. Mimeo, Magyar Nemzeti Bank.
Williams, N., 2003. Adaptive learning and business cycles. Mimeo.