LEARNING IN AN ESTIMATED MEDIUM-SCALE DSGE MODEL

Sergey Slobodyan
Raf Wouters

CERGE-EI

396

Learning in an Estimated Medium-Scale DSGE Model

Sergey Slobodyan
Raf Wouters

Learning in an Estimated
Medium–Scale DSGE Model

Sergey Slobodyan   and   Raf Wouters

CERGE–EI*   **National Bank of Belgium**[†]

Abstract

In this paper we evaluate the empirical relevance of learning by private agents in an estimated medium–scale DSGE model. We replace the standard rational expectation 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 result in a similar fit, and only the transition dynamics that are generated by specific initial beliefs are responsible for the differences between the two approaches. If, in addition, agents use only a reduced information set in forming the perceived law of motion, the implied model dynamics change and for some initial beliefs the marginal likelihood of the model is further improved. The learning models with the highest posterior probabilities add some additional persistence to the DSGE model that reduce the gap between the IRFs of the DSGE model and the more data-driven DSGE-VAR model. However, 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.

Keywords: constant gain adaptive learning, medium–scale DSGE model, DSGE–VAR
JEL Classification: C11, D84, E30, E52

Abstrakt

V tomto článku odhadujeme empirickou relevanci učení soukromých agentů v odhadovaném středně velkém DSGE modelu. Nahrazujeme standardní předpoklad racionálního očekávání z modelu Smets a Wouters (2007) mechanismem konstantního učení. Pokud agenti znají správnou strukturu modelu a učí se pouze parametry, pak oba odhadovací mechanism ústí v podobný fit, pouze přechodová dynamika, která je jimi generovaná, může tyto dva přístupy odlišit. Pokud navíc agenti používají omezený informační set při formování vnímaného zákona pohybu, implikovaná dynamika modelu se změní a marginální likelihood funkce modelu je vylepšená. Avšak odhadovaný zisk v parametrech je malý, což naznačuje větší roli pro počáteční odhady při vylepšování fitu modelu. Učící modely s nejvyššími posteriori pravděpodobnostmi přidávají určitou setrvačnost do DSGE modelu, která redukuje rozdíl mezi odezvovými funkcemi DSGE modelu a DSGE-VAR modelu, který je více řízen daty. Na druhou stranu, dynamika přidaná učením systematicky nezmění odhadované strukturální parametry modelu spjaté s reálnými a nominálními nedokonalostmi v DSGE modelu.

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, 2005, 2007) it was shown that these models, when equiped 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. (2008) shows that these models are still misspecified along various dimensions. Some of the responses to shocks in the DSGE–VAR are characterised by a high persistence, which could be hard to obtain under restrictions imposed by a DSGE model with real and nominal frictions. One potential dimension 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 be hardly considered realistic; therefore, it is important to check the consequences of relaxing this assumption.

In this paper, 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) recursive least squares algorithm. The beliefs about the relationship between expectations and current and past variables adapt to the patterns recently observed in the data. Our approach is similar to the exercise of Milani (2004), who in contrast to us uses a smaller model which under rational expectations does not fit the data as well, and to the work of Orphanides and Williams (2003-2007) who concentrate more on the monetary policy implications of imperfect information and learning.

Several authors have suggested that adaptive learning can enhance the propagation mechanism of the DSGE models and generate the persistence that is otherwise caused by these models' frictions or the driving stochastic processes. For instance, Orphanides and Williams (2005) illustrate how adaptive learning can lead to in-

flation scares or to increased inflation persistence. 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.

We believe that the role of learning is probably smaller in a medium–scale model that fits the data well, with the residuals close to white noise. When driven by thin tail Gaussian innovations, dynamics under learning will not tend to deviate too much from the RE outcomes at least if the initial beliefs are close to the RE beliefs: the potential benefits of adaptive learning are related to the induced time varying beliefs, but this variation is very much limited.

Learning constitues part of a model transmission mechanism. If it can indeed generate persistence that substitutes for the structural inertia sources present in the DSGE models, it is interesting to understand the exact nature of such persistence. The modified persistence can be observed after the agents' beliefs have converged to the limiting invariant distribution, in which case we can talk about learning's permanent effect. Another — transitional — effect is the result of movement from specific initial beliefs towards the beliefs compatible with the invariant distribution; this effect disappears as soon as the transition is over. Given short samples available for estimating DSGE models, distinguishing between permanent and transitory effects can be hard.

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 presample 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. By disconnecting the initial beliefs completely from the pre–sample observations, the last approach may be driven by spurious correlations in the sample data.

The dynamics generated by the learning process is crucially influenced by the assumptions about the information set that is used in forming the beliefs. An

extreme 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 case, this assumption impies 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, the traditional adaptive learning approach as applied in Evans and Honkapohja (2001) also assumes that agents know exactly the exogenous driving processes. This case is referred to as the MSV learning model, or MSV_L for short. We can further differentiate between a situation in which the agents additionally learn about the constants (the inflation target, real interest rate, deterministic trend growth rate) and the one where we assume that the 'correct' values of these parameters are known. It may be more realistic to assume that agents use only a limited information set to form their expectations. In our setup, the most natural assumption would be that agents use only the observed data in their belief regressions. This learning specification is referred to as VAR learning, or VAR_L, in the sequel.

If learning increases the persistence of the model, the roots of the reduced form model increase, implying that the probablity of roots exceeding one in absolute value (and generating unstable outcomes) increases. Such events are supressed by imposing a projection facility while updating the beliefs. These corrections complicate the estimation procedure by introducing discontinuties in the likelihood surface. In the empirical applications, we also estimate the gain parameter that determines the speed of updating of expectation functions' coefficients. The gain is also related to the time horizon that agents take into account when updating their beliefs. Higher gains increases the probability that the dynamics of the system hits the stability boundary during the learning process or that the moment matrices take on unrealistic values. Such events further complicate our estimations.

The learning is introduced at the level of the linearised Euler equations. The implications of learning for the agents' budget constraints are therefore neglected. See Preston (2005) for a more consistent treatment of learning in DSGE models.

The structure of the paper is as follows. First we briefly review the medium size

DSGE model based on the work of CEE (2005) and estimated on US data in Smets and Wouters (2007). The DSGE–VAR approach is used to indicate some of the potential misspecification issues. Then we present the assumptions about the learning process that are considered in this paper. In section four, 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 five and six, we turn to estimating the model under MSV and VAR learning, respectively.

2 The model estimates under rational expectations

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[1]. 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

---

[1]We refer to Smets and Wouters (2007) for the formal presentation of the model.

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 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 estimation 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 match with the number of observables that are 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, when equiped with a rich set of 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. However, Del Negro et al. (2008) show that these models are still misspecified along various dimensions. A combined DSGE–VAR model, where the restrictions of the DSGE model are relaxed and treated as a prior restriction on a VAR model, produces a higher marginal likelihood and performs even better in terms of out–of–sample

---

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

forecasts. In addition, the estimated impulse–response functions of the DSGE–VAR model give some indication of where the misspecification is situated. Here we repeat this DSGE–VAR exercise on the model estimated in Smets and Wouters (2007). The observed misspecification will serve as a benchmark to evaluate the performance of the models under learning.

Table 1 summarise the estimation results for both the DSGE and the DSGE–VAR approach. A fourth order VAR 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 (the hyperparameter $\lambda = 1$). The marginal likelihood of the DSGE–VAR model is considerably higher than for the DSGE model[3]. The mode, the mean and the 5 and 95 percentiles of the posterior distribution of the parameters as obtained by the Metropolis-Hastings algorithm are reported.

{Insert Table 1}

In the DSGE model, 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. In the DSGE–VAR approach, the data are less informative about these constants: the estimated mode for the growth rate is 0.38, for the inflation rate it is 2.6%, and for the annual nominal rate it is it 5.3%.

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 in the DSGE model and 0.95, 0.82 and 0.95 in the DSGE–VAR approach . The high persistence of the productivity and wage mark–up processes implies that at long

---

[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.

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 in the DSGE model and 0.44 and 0.15 in the DSGE–VAR approach). The estimated standard errors of the shocks are systematically lower in the DSGE–VAR approach.

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. The DSGE–VAR parameters in general are close to the DSGE parameters, but the different demand components are estimated to be more sensitive to the interest rate as both the investment adjustment cost and inverse of the intertemporal rate of substitution are lower. The indexation of wages is somewhat lower, but the indexation of prices is higher.

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 in the DSGE and 1.77 in the DSGE–VAR). 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.

The resulting IRF's indicate some significant deviations between the DSGE model and the more data–driven DSGE–VAR approach. Figure 1 shows one interesting dimension on which the standard REE–DSGE tend to be misspecified if we accept the DSGE–VAR model as the benchmark. Following a monetary shock, the REE–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 one of output. In the DSGE–VAR model, inflation typically responds more gradually with only a very weak, if any, response in the first quarters, followed by a more persistent decline in inflation afterwards. This type of inflation response is more standard for many of the SVAR experiments. This gradual and persistent reaction of inflation following a monetary shock contrasts with the immediate and very short–lived effect of the productivity shock on inflation. This contrasting reaction of inflation present in the DSGE–VAR responses is of course quite difficult to achieve in the DSGE model with rational expectations. They will serve as the benchmark when discussing the estimation results for the models with learning.

{Insert Figure 1}

3   Learning setup

3.1   Updating of beliefs

We implement the adaptive learning within the DYNARE 3.064 MATLAB toolbox which is used to estimate and simulate DSGE models. Our model is driven by the exogenous driving processes x_t,

$$x_t = \rho x_{t-1} + \epsilon_t + \theta \epsilon_{t-1}.$$

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

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

DYNARE represents our model in the following way:

$$A_0 \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + A_1 \begin{bmatrix} y_t \\ w_t \end{bmatrix} + A_2 E_t y_{t+1} + B_0 \epsilon_t = 0, \tag{2}$$

where the vector y_t includes endogenous variables of the model.[4] The solution of the model is provided by DYNARE as

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

The equations (1) form part of the system (3). The solution (3) notationally differs from the Minimum State Variable (MSV) solution, which for a system consisting of (2) and (1) is usually written as

$$y_t = a + b y_{t-1} + c w_t.$$

The vector y contains a subset of state variables y^s and variables that appear with a lead in the model, y^f.[5] 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.[6],[7] Finally, the agents' beliefs about reduced form coefficients α and β are updated using a constant–gain variant of the Recursive Least Squares (RLS). In our model, there are 12 forward–looking

---

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

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

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

[7]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.

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 $Z_t = \left(1, \left(y_{t-1}^s\right)^T, w_t^T\right)^T$, and collected the beliefs into a single matrix $\phi^T = \left(\alpha, \beta^T\right)$.[8]

3.2 Initial Beliefs

Equations (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 4 ways of determining the initial beliefs: three that are consistent with some REE, and one that is based on regression estimates of the pre–sample data.

Any REE, given for example by (3), implies an equilibrium relation between the forward–looking variables y_t^f and the vector Z_t. This relation is given by

$$E\left[Z_t Z_t^T\right]^{-1} \cdot E\left[y_t^f Z_t^T\right],$$

where the expectations $E[]$ are also derived using the distribution implied by the REE solution (see Del Negro et al 2008). This relation is used to initialise ϕ_0. We take $E\left[Z_t Z_t^T\right]$ as initial condition for the second moments matrix R_0. One could select the initial beliefs to always correspond to the REE which is implied by the estimated parameter vector. We think that this way 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. The second REE–consistent way of selecting the initial beliefs makes them consistent with some REE, not necessarily the REE that corresponds to the optimized parameter vector. Another difference between the two methods is that in the first one,

---

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

initial beliefs always correspond to the current parameters, while in the second they remain fixed during the optimization step. Finally, similarly to Milani (2005), we could optimize the initial beliefs, together with other parameters of the model. This is the third way of generating the initial beliefs.[9]

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 Giannitsarou and Carseles–Poveda (2007).

In all estimations reported, we assume that the agents know both the law of motion (1) of the exogenous driving processes and the standard deviations of the innovations ϵ_t.[10] 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 (1) in the updating step (5). We do not pursue this route here.

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. All variables in the MSV REE solution are model variables which are not, in fact, observed. 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 (5).[11]

---

[9]The difference between the second and the third way is that the REE in the 2nd way is based on the data. We estimate the model under RE for a pre–sample time period to generate the REE which is then used to form the initial beliefs used in the second method.

In contrast, the third way of forming initial beliefs is based on in–sample optimization of both initial beliefs and the parameter values. Given the large dimensionality of our beliefs matrix (12×21 for beliefs about reduced form parameters alone, plus beliefs about initial R which is a 21×21 symmetric positively definite matrix), we form an auxilliary model which is exactly the same as the estimated one but for parameter values; the parameter values of the auxilliary model are then estimated together with the the main model parameters. The REE implied by the auxilliary model is then used to construct the initial beliefs.

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

[11]In terms of Hamilton (1994), we use $\widehat{y}_{t-1|t-1}^s$ and $\widehat{w}_{t|t}$ 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

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 $\widetilde{Z}_t = \left(\left(y_{t-1}^s \right)^T, w_t^T \right)^T$ as the data vector. However, if we assume that the agents are also (implicitly) learning the values of the growth rates or inflation target, we include the constant into (5). In addition, we also report the results of estimations where the agents are using a PLM which does not coincide with the MSV solution. In particular, they might include only the observed non–zero mean variables into the PLM, in which case, inclusion of a constant into (5) is warranted as well.

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 equation (2) for $\left(y_t^T, w_t^T \right)^T$ and derive a time–varying VAR representation of the model:

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

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 set–up 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[12]. Given that DYNARE provides equations (3) and (2) 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 (VAR_L) case.

---

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.

[12]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.

3.4 Projection facility and computational issues

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. Often cited motivation for adaptive learning is an attempt to describe the real world behavior of the agents; in practice, forecasters do not use explosive models. The simplest method of dealing with explosive behavior of T_t is to ignore the updating step which has resulted in undesirable roots, and use previous μ_t, T_t, and R_t instead. This mechanism is very similar to the theoretical construct of a projection facility described in Evans and Honakpohja (2001) which was used, for example, by Orphanides & Williams (2007).[13] We have to note, however, that thus implemented projection facility results in discontinuous jumps in the likelihood function: even for two very close parameter vectors the likelihood function could differ significantly if one results in a sequence of $\{T_t\}_{t=1}^{N}$ which is never explosive, while the other generates a sequense with at least one t such that T_t is explosive. Such a non–continuous likelihood function results in significant computational difficulties which are described later. Moreover, such discontinuities arise also in the case when the projection facility is not implemented: allowing T_t to be explosive even for very few periods leads to sharp deterioration in precision of predicting the future values of y^f, and thus to a much worse likelihood.

Another issue related to using the updating equation (5) is the sensitivity to the value of "effective gain", given as gR_t^{-1}. Even if the gain parameter g is small, the beliefs ϕ might be very sensitive to the data if the matrix R is "small" in the sense that its smallest eigenvalues are very close to zero. If this is the case, instances of Z_t which are not orthogonal to the eigenvectors of R corresponding to the smallest eigenvalues would lead to very large changes of beliefs ϕ. This extreme sensitivity of the beliefs on the data leads to further complications in the likelihood function, as close parameter vectors might result in very different behavior of $\{T_t\}_{t=1}^{N}$. We

---

[13]Usage of projection facility amounts to restricting beliefs ϕ to a small neighborhood of their REE values. Discarding explosive T_t is equivalent to imposing a restriction on a highly non–linear function of ϕ instead.

Here, we cannot impose restrictions on ϕ directly, because when y^f and y^s include different variables, the forecasting rule is not a simple VAR. To transform the rule into VAR, one needs information contained in the REE, specifically, the matrix T.

discuss observable consequences of this sensitivity in later sections; in order to mitigate somewhat the negative effect of it on the likelihood function, we use a Ridge correction mechanism.[14]

As demonstrated by McCallum (2006), for models with time t information available (time–t dating of Evans and Honlapohja 2001), determinacy is a sufficient condition for E–stability (of the MSV solution). Moreover, if more than one solution is stable, then the unique solution derived using decreasing order of the system eigenvalues, is E–stable. Our model belongs to the class considered by McCallum (2006), DYNARE uses the solution which corresponds to eigenvalues ordered in decreasing order, and the points in the parameter space which result in indeterminate equilibria are heavily penalized by adding a large constant to the likelihood function (and thus cannot be a result of the optimization). Thus, we are guaranteed that the MSV solution is always stable. Note, however, that the result does not necessarily hold if the agents are using a non–MSV PLM. Stability conditions in this case have to be checked numerically for every constellation of parameter values.

4 Simulation exercises

4.1 Simulated second moments

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 some of the stochastic shocks.

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

---

[14]If at time t the smallest eigenvalue of R_t is less than some small λ, $(R + \lambda I)^{-1}$ is used instead of R^{-1} in the equation (5a). Here I is the identity matrix. In our estimations, λ is usually set to $1 \cdot 10^{-5}$.

1. the RE model as the benchmark model;

2. MSV learning with perfect information about the constants;

3. MSV learning where agents learn the constants as well; and

4. VAR–learning with beliefs about observed variables and a constant only.

In all these experiments the structural parameters of the model, including 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.[15] In order to understand the influence of the initial beliefs on the simulation outcomes, we run 1000 simulations for each model; each simulations starts from a different initial PLM (initial beliefs).[16] 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. As is often observed in simulated data under adaptive learning, occasionally time series can deviate from their long–run averages to a significant degree. In section 4.3, we discuss the impact of these events or "exits" on the simulation results.

In Table 2, we report the volatility and persistence of the generated data. The first two columns describe the output for the RE model. We present standard deviations (for both the growth rate and levels) as well as autocorrelations for

---

[15]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.

[16]We draw 1000 parameter vectors from the posterior distribution of the model estimated under RE, and use the resulting REE to construct initial beliefs.

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

the simulated counterparts of observed variables. Note that the volatilities for the levels are systematically lower during the transition period because the state variables are initialised at their steady state values.

{Insert Table 2}

The outcomes for the MSV learning model (with or without a constant) and with a small (0.01) gain parameter remain almost identical to the outcomes under RE. The additional variation `that is` related to the beliefs updating does not significantly increase the volatility or the persistence of the observable variables. This is especially true for the permanent dynamics. For the transition dynamics, there are very small changes: the standard deviation and the correlations tend to be smaller during the transition period for the model with learning. For this small gain parameter, 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 stronger in the case where agents are also learning about the constants in the model. These effects remain rather 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 volatilities of the inflation and the interest rate that experience the strongest increase. In the simulations with a higher gain, the percentages of observations that are disregarded by the projection facility increases quickly: 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

18

to grow quickly in these cases. An additional source of increased variability for larger gains is the so–called "exits" (see section 4.3).

Turning to the VAR learning model, we observe significant deviations even for a small gain of 0.01. During the transition period, the standard deviations for most of the variables increase, although for the permanent dynamics some of the standard deviations tend to decline. The persistence in the series increases considerably (with the exceptions of the wage changes). With higher gain, all standard deviations and correlations tend to increase and this increase is more pronounced than under the MSV beliefs, reflecting again the fact that the unit root 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 Figure 2a and 2b, 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 which quickly revert towards the neighborhood of the mean value. These jumps are the "exits" which will be discussed later. As Figure 2b makes clear, under VAR_L the exits could already be observed for a much smaller gain of 0.02.

{Insert Figure 2a and 2b}

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. Figures 3a and 3b show the IRF of output and inflation to monetary policy and productivity shocks. The graphs contains 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.[18]

---

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

These graphs, therefore, show 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. IRFs to a monetary policy shock converge very fast, which is not surprising given the very low estimated persistence of the monetary policy shock and tendency of the constant gain learning to "forget" past data. 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.

{Insert Figure 3a and 3b}

The impulse responses under the VAR learning differ from the IRFs under RE considerably more than under MSV_L, and the decile dispersion is also wider. As in the MSV_L, decile dispersion for impulse responses to the monetary policy shock decreases faster than for the productivity shock. Under VAR_L, the agents' Perceived Law of Motion (and therefore, their expectations) differs substantially from the RE reduced form, which explains this discrepancy.

To better understand the reason VAR learning generates second moments and IRFs that are significantly different from those under RE, we can look at the agents' expectations of future variables. In a linearised model these expectations are linear functions of past values of the state variables. Under rational expectations, coefficients of these linear functions are fixed; adaptive learning allows them to vary with time. Agents use the PLM to form expectations of the future variables. In Figures 4, we plot one–step ahead expectations for the RE model and the three learning specifications used in this section. The expectations are taken from one simulation run, choosen randomly out of the 1000 generated as described above. As is obvious from Figure 4, the differential effect of MSV and VAR learning comes mainly from the way these types of learning affect expectations. Under the MSV_L, impulse responses of both actual and expected inflation are very close

---

themselves change.

20

to those under the RE; the same result holds for other model variables.[19] In contrast, with VAR learning, inflation expectations react differently. For a monetary policy shock, inflation expectations are affected much less initially, but the effect is longer lived[20]. The reason is that the RE–model consistent beliefs under VAR_L imply that the agents assume the inflation process to be much more persistent than under the RE: coefficient on past inflation in the inflation forecasting equation is close to 0.8 rather than 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.

{Insert Figure 4}

The impact of subsequent learning on these IR is very minor: for the MSV learning and the gain values we considered, 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_L, in fact, are not correct even on average: equilibrium beliefs are given by the Restrictive Perceptions Equilibrium (RPE) discussed in section 6. 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 Figure 5, where we track behavior 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.

{Insert Figure 5}

---

[19]The impulse responses under MSV learning differ from those under the RE because we used a random draw from the posterior distribution to construct agents' beliefs in the learning model.

[20]The same dynamic profile applies for the productivity shock but the overall effect on inflation is reinforced.

4.3 Simulations and large deviations

As mentioned previously, occasionally a simulated time series exhibits a very large deviation from its long–run average, which is typically reversed after a small number of periods. Such behavior was also observed in simulated data under learning by Orphanides and Williams (2007) and Giannitsarou (2007).

We believe that such events could be described as exits or "large deviations". If the dynamics under learning is stable (E–stability of the equilibrium is a necessary condition for stability of the learning dynamics), on average, beliefs should equal their equilibrium values; they cannot deviate too far from the equilibrium for long periods of time. However, constant gain (perpetual) learning introduces permanent stochasticity into the model, as beliefs can never be exactly equal to their equilibrium values.

Suppose that a sequence of shocks hits a simulated model in such a way that the beliefs move in (approximately) the same direction period after period and leave some neighborhood of the equilibrium beliefs. Agents' expectations will shift as a result, which might also be manifested in observed variables moving far from their equilibrium values. Clearly, such a situation cannot continue indefinitely: observing such sequence of shocks is unlikely. Theory of large deviations [Dembo and Zetouni (1998), Freidlin and Wentsel (1998)] studies properties of such "rare events"; their probabilities "have asymptotics of the form $\exp\left\{-C\varepsilon^{-2}\right\}$ as $\varepsilon \to 0$ (rough asymptotics, i.e., not up to equivalence but logarithmic equivalence)", see Freidlin and Wentsel (1998). For an application of the large deviations theory to models of adaptive learning with constant gain see Cho, Williams, and Sargent (2002) and Kolyuzhnov, Bogomolova, and Slobodyan (2006), among others.

One of the basic predictions of the large deviations theory is that if the system is started near the equilibrium point, number of periods necessary to observe the first crossing of the boundary of a given neighborhood of the equilibrium (called first exit time τ) is distributed exponentially. As a result, probability of the first exit happening after N periods is given as

$$\Pr(\tau \geq N) \sim e^{-\lambda N}.$$

Therefore, a very simple check to see if the extreme events in the simulation could be "rare" large deviations is to plot the logarithm of empirical probability that the first escape time is greater than N (logarithm of the share of simulation runs that escaped after N) versus N. In these coordinates one should see a downward sloping straight line. Figure 6 illustrates this hypothesis based on simulation results from a MSV learning model with a constant and a gain of 0.03. A large deviation is defined as a realisation where a state vector variable takes on a value that exceeds four times the standard deviations (observed in the REE) away from the steady state. The results show that indeed, we do observe the straight line for a large range of N. The fit is especially good if we exclude simulation runs where projection facility was used too often.[21] Theoretical calculation of the parameter of exponential distribution η is a very complicated procedure that will not be attempted in this paper. Similar results are obtained for the VAR_L model with a constant, if the model is started at the Restricted Perceptions Equilibrium (RPE) beliefs; for a discussion of the RPE under VAR_L , see section 6.

{Insert Figure 6}

The exits affect measured standard deviations and autocorrelation of the simulated time series, as is illustrated in Table 3. For example, output growth simulated under MSV learning with gain equal to 0.02 has 'transitional' standard deviation of 0.941 and a 'permanent' one equal to 0.962; when runs with exits are removed, these numbers reduce to 0.938 and 0.946, respectively. Effect of exits on autocorrelations of the simulated data is minor. Exits have larger effect on the 'permanent' standard deviations; as a result, after taking them into account, the difference between 'transitional' and 'permanent' numbers becomes smaller for all the simulated variables. We believe this is due to the fact that even if state variables (approximately) revert back to their sample averages after an exit, the beliefs do not necessarily do so, and the dynamics after the exit is, in fact, different than the one before it. Eliminating simulated runs with exits reduces this effect.

---

[21]Using projection facility changes dynamics of the model. Therefore, if it is used too often, we deal with a significantly different system; such runs cannot be compared directly to the runs where projection facility was used seldom, if at all.

For simulated time series under VAR learning with gain of 0.02, excluding exits does not lead to convergence of measured standard deviations for the 'transitional' and 'permanent' dynamics; the opposite effect is often observed. In addition, measured autocorrelations often reduce after the exited runs are removed. We attribute this behavior to the fact that RE model–based beliefs do not constitute an equilibrium under VAR_L; therefore, when the gain is sufficiently large, 'transitional' and 'permanent' dynamics are, in fact, different, as the beliefs drift towards their RPE values.

{Insert Table 3}

5   Estimation under MSV learning

In this section, we present the estimation results for the case of MSV learning, when the 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 parameters. 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 (3.3).

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.

24

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 are reported in Table 4 and are very similar to the results for the RE model in Table 1. Both the posterior distribution of the parameters and the marginal likelihood of the model are extremely close to the REE estimates.

{Insert Table 4}

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. 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 10% and 90% bounds of the posterior distribution for the gain are 0.0 and 0.034 with the mean around 0.018.

As can be seen from Table 4, 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.

The marginal likelihoods of the learning and RE models are very similar, meaning that the data is not able to distinguish the two models.[22] 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, 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 marginal likelihood will serve as the benchmark to evaluate the alternative specifications of the learning process.

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 necessary 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.[23]

{Insert Table 5}

As one could expect, the model with optimised initial beliefs outperforms all other MSV–L models and the RE model in terms of marginal likelihood.[24] Note that this

---

[22]The mode of the posterior distribution under learning is significantly higher than the one under the RE, but this is mainly due to the fact that there is an additional parameter, the learning gain, in the parameter set of the learning model.

[23]For MSV learning without constant, dimensionality of the beliefs space is $20 \times 12 = 240$ coefficients in the forecasting functions plus $20 \times 19/2 + 20 = 210$ elements of the second moments matrix. Estimating a separate model for the beliefs adds only 20 extra parameters.

[24]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

difference in the marginal likelihood is much smaller than the difference between the posterior modes of these models. The likelihood function of the model with optimised initial beliefs is characterised by an irregular surface that creates severe computational problems : 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 not from the time variation induced by adaptive learning.

Turning attention to the estimated parameters we can say that the interest elasticity of investment is high (investment adjustment cost is low), risk aversion σ_c is low but the habit persistence coefficient η is relatively high. At the same time, the investment shock is more volatile and less persistent, while the price mark–up shock is estimated to be very persistent. Comparing the estimated parameters in the 'initial belief' and the 'real' models, we note that the price and wage stickiness and the price indexation coefficients are all lower in the 'real' model than in the model estimated under RE, while the opposite holds for price stickiness and indexation in the 'initial belief' model. Monetary policy rule in the 'real' model is characterised by a stronger reaction to the inflation, output gap, and past interest rate (interest rate smoothing) than in the 'initial belief' one.

_____

make the marginal likelihood of this model directly comparable to the other models estimated in this paper.

The implications of the estimated parameters and beliefs for the IRF of the productivity and monetary policy shock are illustrated in Figure 7. 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.

{Insert Figure 7}

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 contemporaneous reaction of inflation is almost zero and the peak effect occurs more than one year later (and definitely after the peak response in output).[25]

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. But it is mainly the updating of these beliefs over time that enforces the gradual reaction of inflation to the policy shocks. This is remarkable given that

---

[25]The impact of the optimised beliefs under MSV learning are much stronger than the one discussed in the simulation session (see Figure 3a), where the beliefs were drawn from the posterior distribution of the RE model. The estimated "initial belief" parameters deviate indeed considerably from the RE model.

the monetary policy shock is assumed to be known by the agents and appear explicitly in their belief equations under MSV learning. 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 with optimised initial beliefs model, the indexation parameters are more or less unchanged and only the habit persistence coefficient is larger with adaptive learning. The persistence of the exogenous shocks stays also largely the same with the exception of the persistence of the investment shock which becomes much lower.

5.3   MSV beliefs with presample model initialisation

An alternative way to derive the initial coefficients of the MSV beliefs is to estimate a RE model over a sample that preceedes 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 6). 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 severe computational problems. The posterior mode of this model is higher than in the benchmark case, suggesting that the pre–sample REE contains some potentially useful information about the in–sample initial beliefs. This information is however very sensitive to specific parameter constellations and this parameter uncertainty reduces the marginal likelihood.[26]

---

[26]MSV learning models tend to generate second moment matrices with tiny smallest eigenvalues, necessitating usage of the ridge correction mechanism. In such cases, beliefs are usually

{Insert Table 6}

In Figure 8 we plot the time–varying IRFs for the productivity and the monetary shock. Note the following features. First, the initial beliefs are somewhat out of line with what the model prefers, resulting in large adjustments of the belief coefficients early in the sample; this results in significant time variation of the IRFs, especially for inflation. The estimated pre–sample model differs in many directions from the in–sample one. For instance, the persistence of the price and wage shocks is very different from that observed in other models, price stickiness is very low and indexation modest. These features of the initial beliefs result in strong reaction of inflation to the shocks in the beginning of the sample. Second, output responds in this model stronger to the productivity shock than in the model with optimised beliefs or RE model. Third, the reaction of inflation to productivity shocks increases over time, but contrary to the model with optimised beliefs this responce gradually becomes more persistent and hump shaped. Fourth, the maximum responce of both output and inflation to the monetary shock decreases in magnitude over time, the impact effect also decreases, but the hump shaped reaction increases over time. At the end of the sample the IRF tends to converge towards the responses generated by the RE model rather than to these of the MSV learning model with optimised beliefs.

{Insert Figure 8}

5.4 MSV beliefs with presample 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.

---

strongly adjusted in the early in–sample periods, and overfitting of the initial data might become an issue.

Under this approach, the initial beliefs are such that beliefs adjust even stronger early in the sample than with the pre–sample REE model initialisation of section 5.3. Depending on the length of the pre–sample, the estimated gain parameter is extremely high, varying between 0.04 to 0.06, and this leads again to various computational difficulties. For example, optimisation routines are finding local rather than global optima, MCMC 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 considered models in which private agents know the correct specification of the model, but learn about the values of the belief parameters. Here 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. We assume that agents use the same list of seven observable variables as we do in our estimation of the overall model. This form of misspecified belief equations is probably a more realistic approximation of the actual information set available to economic agents. 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. 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 four 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. We refer to this setup of the beliefs as VAR learning. Similarly to the case of MSV learning, we also consider different ways of constructing the initial beliefs at the start of the sample.

VAR learning has several implications in the adaptive learning set–up. First, E–stability of the equilibrium is not guaranteed any longer, as the result of MacCallum applies only to the MSV case. However, for parameter values close to the REE, E–stability is obtained. Second, using a different information set for forming forecasts may influence the equilibrium laws of motions themselves, and makes the RE model consistent beliefs different from those that would obtain on average in an infinitely long simulation. In the appendix, we provide a more detailed discussion of these issues.

6.1   VAR learning with model consistent initial beliefs

Table 7 summarizes the estimated parameters, while Figure 9 shows the implied IRF functions.

{Insert Table 7}

{Insert Figure 9}

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 mis–specified 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.

32

Numerical simulations of the E–stability ODE under VAR learning with constant (derived in the Appendix) show that for parameter values close to the REE and the initial beliefs close to the RE–model consistent ones, dynamics under learning is expected to remain stable. However, divergence is often observed if the simulations are started with beliefs about constants which are not equal to zero as in the RE–model consistent case. Simulations show that beliefs about constants are especially volatile. Therefore, if some data point makes the agents to assume a particularly large constant value in some forecasting equations, further evolution of the beliefs might become unstable. During the estimation procedure, such situations are most probably associated with 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.

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 8). 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.

{Insert Table 8}

{Insert Figure 10}

The IRFs are again relatively stable and close to the ones of VAR learning with model consistent beliefs.The inflation response to the monetary policy shock is

33

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 becomes again much more gradual and persistent.

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. The conclusion that the gain in the fit comes mainly from the initial beliefs, is also confirmed by the marginal likelihood of this model where the initial optimised beliefs are kept constant over the complete sample. The small gain also suggests that the beliefs are not materially changed according to the constant gain learning schedule. Therefore, it might be useful to test whether alternative and more efficient learning procedures (such as Kalman filter) are performing better in this context. 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 shock. Although there is considerable updating in this model, no computational problems were encountered, which is probably related to the fact that the 2nd moments matrix doesn not have very small eigenvalues, and thus $\left\| R_t^{-1} \right\|$ is never too large .

6.3   VAR learning with presample 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 vs 0.001 for both RE–model consistent and optimised VAR beliefs) in order to allow the beliefs to adjust and to move away from the imposed initial beliefs.

{Insert Table 9}

34

{Insert Figure 11}

The IRF shows a reaction in output that is still in line with the other models, although the impact of the monetary shock on output is very high in this model, probably reflecting initially high belief coefficient on inflation in the output forecasting equation. The reaction of inflation fluctuates a lot over time. The response of inflation to the productivity shock in the beginning of the sample is strongly negative but at the end inflation does not seem to react at all on impact. The reaction of inflation to the policy shock is already gradual in the beginning of the sample and over time the impact effect becomes even positive, but with an extremely persistent negative effect after several quarters. This response is similar to the often observed price puzzle in SVAR models. Again, we attribute this significant time variation to the fast adjustment of beliefs from the values imposed by the pre–sample RE model. We also checked whether the choice of the presample 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 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 into 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. Specific initial beliefs contribute significantly to this result, which proves that the model–consistent expectation imposed by the rational expectations hypothesis is too restrictive. The best performing models are the ones where the initial beliefs are optimised to explain the in–sample data, consistent with previous results in the literature. Limiting the set of variables used in the forecasting

equations to a list of observed macro–variables can generate 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 rather low estimated gain, which suggests that it is the initial beliefs different from the RE–model consistent ones that bring about improvement in the marginal likelihood, rather than the beliefs updating of the constant gain learning type per se. On the other hand, choosing the initial beliefs that are too far away 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 Kalman filter, which can converge faster than constant gain learning, see Sargent and Williams (2005), could modify the tentative conclusion about higher importance of initial beliefs than of the learning process itself.

In terms of IRFs our discussion was limited to the implications of the productivity and the monetary policy shock. The implications for the other shocks still need to be documented. The implications for 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 results also overcome some of the major shortcomings of the REE–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, but is again in line with the results from the DSGE–VAR methodology which indicate misspecification but no systematic bias in the structural parameters.

8 References

Cho, I.-K., N. Williams and T.J. Sargent (2002), "Escaping Nash Inflation", *Review of Economic Studies*, Vol. 69(1), p. 1-40.

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.

Del Negro, M. and F. Schorfheide (2008). "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)," *Journal of Monetary Economics*, Vol. 55(7), p. 1191-1208 .

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.

Dembo, A. and O. Zeitouni (1998), "Large Deviations Techniques and Applications". Springer.

Eichenbaum, M. and J. Fisher (2007), "Estimating the Frequency of Re–optimization in Calvo–Style Models", *Journal of Monetary Economics*, Vol 54 (7), p. 2032-2047.

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

Freidlin, M.I. and A.D. Wentzell (1998), "Random Perturbation of Dynamical Systems". Springer-Verlag.

Giannitsarou, C., and E. Carceles–Poveda (2007), "Adaptive Learning in Practice", *Journal of Economic Dynamics and Control,* Vol. 31(8), p. 2659-2697.

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*, Vol. 27(4), p. 1241–77.

Marcet, A. and T. J. Sargent (1989), "Convergence of least squares learning mechanisms in self-referential linear stochastic models," *Journal of Economic Theory*, Vol. 48(2), p. 337-368.

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", *Journal of Economic Dynamics and Control*, Vol. 32(10), p. 3148-3165.

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 (2003), "Imperfect Knowledge, Inflation Expectations, and Monetary Policy", NBER Working Paper No. W9884.

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

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.

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.

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 VAR Learning: E–Stability and Mean Dynamics

For simplicity, we present the results for the case without a constant. The logic will carry through if we were to include it.

As stated previously, VAR learning implies that the agents use the following PLM:

$$y_t = bz_{t-1}, \tag{6}$$

where z_t is a subset of the model endogenous variables. z_t selects specific endogenous variables, $z_t = Hy_t$

The agents' PLM (6), inserted into the model equations[27]

$$
\begin{aligned}
y_t &= \beta E_t y_{t+1} + \delta y_{t-1} + \kappa w_t, \\
w_t &= \rho w_{t-1} + \epsilon_t,
\end{aligned}
$$

produces the Actual Law of Motion (ALM)

$$
\begin{aligned}
y_t &= (I - \beta bH)^{-1} \delta y_{t-1} + (I - \beta bH)^{-1} \kappa w_t, \tag{7} \\
y_t &= (\widetilde{T}_y, \widetilde{T}_w) \cdot \left(y_{t-1}^T, w_t^T\right)^T. \tag{8}
\end{aligned}
$$

For comparison, if the agents started from the PLM that corresponds to the MSV,

$$y_t = by_{t-1} + cw_t,$$

the resulting ALM would be

$$y_t = (I - \beta b)^{-1} \delta y_{t-1} + (I - \beta b)^{-1} (\beta c\rho + \kappa) w_t. \tag{9}$$

The ALM under VAR learning differs from the MSV solution; moreover, there is only one parameter vector b, which could be used to set parameters of the PLM

---

[27]These equations are, in fact, the same as (3): the structure of the model is such that the terms at w_{t-1} and ϵ_t in (3) combine to produce exactly κw_t.

and ALM equal to each other. This implies that the usual method of undetermined coefficients could not be used to derive the stationary point of the map from PLM to ALM. To derive this point, we need to calculate the E–stability ODE first and then look at its stationary point(s). Restricting the PLM to a subset of the variables is associated with Restrictive Perceptions Equilibrium (RPE), see Evans and Honkapohja (2001, Ch. 13.1.2). Moreover, one of the variables in z is not present in the MSV set of variables and it is likely to contain some information from the variables excluded from the RPE set; thus, we do expect that in equilibrium this variable will be used in the forecasting equations. Therefore, the equilibrium in question is, in fact, a mixure of the under– and over–parameterisation. For lack of a better term, we shall continue calling it RPE.

Let's consider the problem of E–stability. The updating equations could be written as

$$b_t = b_{t-1} + gR_t^{-1}z_{t-1}(y_t^f - b_{t-1}^T Z_{t-1})^T = \tag{10a}$$

$$b_{t-1} + gR_t^{-1}z_{t-1}\left((\widetilde{T}_y, \widetilde{T}_w) \cdot \left(y_{t-1}^T, w_t^T \right)^T - b_{t-1}^T z_{t-1} \right)^T, \tag{10b}$$

$$R_t = R_{t-1} + g(z_{t-1}z_{t-1}^T - R_{t-1}). \tag{10c}$$

Taking limits and expectations on the right hand side, we obtain that the E–stability differential equations are

$$\frac{db}{d\tau} = \widetilde{T}_y E\left[y_{t-1}, z_{t-1}^T \right] R^{-1} + \widetilde{T}_w E\left[w_t, z_{t-1}^T \right] R^{-1} - b, \tag{11a}$$

$$\frac{dR}{d\tau} = E\left[Z_{t-1}Z_{t-1}^T \right] - R. \tag{11b}$$

The equation for R is globally stable around the equilibrium point $\overline{R} = M = E\left[z_{t-1}z_{t-1}^T \right] = HE\left[y_{t-1}y_{t-1}^T \right] H^T$. The variance–covariance matrices $E\left[y_{t-1}, z_{t-1}^T \right]$ and $E\left[w_t, z_{t-1}^T \right]$ are themselves complicated functions of b, as they are determined from (7) as a solution to matrix Lyapunov equation. Thus, the right–hand side of the E–stability ODE is a highly nonlinear function of $\widetilde{T}_y$ and $\widetilde{T}_w$, which are, in turn, nonlinear functions of b. It is very likely that the resulting RPE differs significantly from the RE–model consistent beliefs. The ODE is highly dimensional: in case without the constant, its dimension equals 112. Linearization of (11) and calculation of eigenvalues needed to determine local E–stability is likely to result

in an extremely complicated expression that will not contain any intuitive results. Therefore, we turn to numerical simulations to find the equilibrium and to determine whether it is E–stable.

Numerical simulations of the mean dynamics ODE (11) show that if one starts from the RE–model consistent beliefs, the resulting RPE is stable, whether the constant is included into the vector z or not. The RPE is very different from the RE–model consistent beliefs. In particular, the variance–covariance matrix implies smaller variances, and the response of some forward–looking variables to the observables is much stronger than at the REE. However, when we simulate the ODE for the case with constant and initialise the beliefs about the constants with some non–zero vector, a sort of "quasi–stability" is observed: for a while, the beliefs seem to converge to the equilibrium values, but then fast divergence starts. Given that the mean dynamics ODE is nonlinear with more than one hundred dimensions, we are not ready to claim that this behavior indicates small region of attraction of the strongly E–stable equilibrium, especially in the direction of beliefs about the constants. A purely numerical divergence might become an issue. However, if the system is so close to the instability boundary that small numerical errors might push MATLAB ODE solver into divergence, small numerical errors made by the agents operating in real time may lead to the same result (divergence).

We advance two hypotheses for the estimated gain value under VAR learning being so small. The first hypothesis is that the central tendency of beliefs evolution (the mean dynamics) is not compatible with the data under VAR learning, and the mean dynamics path from the RE–model consistent beliefs to the RPE generates beliefs that are not conductive to good model fit. The second hypothesis points to learning about the constants as the main culprit, because numerical simulations of the mean dynamics indicate that it might lead to instability.

To test the first hypothesis, we simulate the mean dynamics ODE (11) starting from the RE–model consistent beliefs and sample the resulting path at $t = 0.5$, 1, 2, and 3. We then estimate our model with gain set to zero and setting initial (and fixed) beliefs equal to these sampled values. The resulting posterior mode is -843, -863, -907, and -951, respectively, *vs.* -839 achieved if the beliefs are fixed at their

RE–model consistent values.[28] We consider this result a partial confirmation of the hypothesis: evolution of beliefs towards RPE makes them incompatible with the data we use to estimate the model.

To understand whether learning about the constants is problematic, we estimate the model starting from the RE–model consistent beliefs but fixing the gain at 0.01. After the estimation, the beliefs change significantly, moving towards the RPE values; in particular, the diagonal elements of the matrix of second moments are much closer to the RPE than to the RE–model values. At the same time, the constants in the forecasting equations are believed to be large; in numerical simulations, such values could lead to the "quasi–stability" described above. If, indeed, the data forces beliefs outside of some region of attraction, the only way to prevent this is to make the gain parameter small. Hovewer, an argument against the second hypothesis is that estimation without constants (where instability related to learning the constants is absent) produces very similar results, with the estimated gain equal to 0.001 and the posterior mode about -839.

Evaluating the body of evidence from numerical simulations and estimation exercises described above, we believe that the mean dynamics trajectory connecting RE–model consistent beliefs with the RPE beliefs is incompatible with the data, which leads to the gain being estimated at the lower end of the allowed interval.

The very presense of VAR learning changes the resulting equilibrium. An evolution of beliefs from their initial values towards the RPE makes the beliefs, and as a result the transmission mechanism of the model, incompatible with the observed data. An interesting question is whether some other set of variables used in forecasting equations might generate a RPE better aligned with the data, but this is beyond the scope of this paper.

B   Data appendix

The model is estimated using seven key macro-economic time series: real GDP,

---

[28]The relationship between continuous time units of the mean dynamics and discrete time periods of the real time learning is given as $t = Ng$, where g is the gain. For the gain of 0.01, which is slightly less than the estimates for MSV learning and on the lower end of the range used in other studies, $t =$0.5, 1, 2, and 3 correspond to 50, 100, 200, and 300 periods.

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.

Table 1: Marginal likelihood of DSGE versus DSGE-VAR model as indication of misspecification

	Prior distribution			Benchmark REE-DSGE model Posterior distribution				DSGE-VAR model (λ=1) Posterior distribution			
	Distr.	Mean	St.Dev.	Mode	Mean	5 percent	95 percent	Mode	Mean	5 percent	95 percent
ρ_a	Beta	0.50	0.20	0.96	0.96	0.94	0.98	0.95	0.93	0.89	0.99
ρ_b	Beta	0.50	0.20	0.18	0.22	0.08	0.36	0.44	0.50	0.20	0.76
ρ_g	Beta	0.50	0.20	0.98	0.98	0.96	0.99	0.82	0.77	0.61	0.96
ρ_q	Beta	0.50	0.20	0.71	0.71	0.62	0.81	0.51	0.50	0.32	0.68
ρ_r	Beta	0.50	0.20	0.13	0.15	0.04	0.24	0.15	0.16	0.04	0.25
ρ_p	Beta	0.50	0.20	0.90	0.89	0.81	0.97	0.69	0.56	0.21	0.85
ρ_w	Beta	0.50	0.20	0.97	0.97	0.95	0.99	0.95	0.74	0.56	0.98
μ_p	Beta	0.50	0.20	0.74	0.70	0.55	0.86	0.59	0.54	0.26	0.83
μ_w	Beta	0.50	0.20	0.89	0.85	0.76	0.94	0.81	0.53	0.18	0.85
ρ_{ag}	Normal	0.50	0.25	0.53	0.52	0.38	0.67	0.52	0.52	0.35	0.69
φ	Normal	4.00	1.50	5.49	5.74	3.97	7.42	3.84	3.89	2.37	5.69
σ_c	Normal	1.50	0.38	1.40	1.38	1.17	1.59	1.16	1.20	0.86	1.54
η	Beta	0.70	0.10	0.71	0.71	0.64	0.78	0.66	0.63	0.51	0.73
ξ_w	Beta	0.50	0.10	0.74	0.71	0.60	0.81	0.69	0.73	0.62	0.85
σ_l	Normal	2.00	0.75	1.92	1.84	0.92	2.79	1.89	1.77	0.81	2.77
ξ_p	Beta	0.50	0.10	0.66	0.65	0.56	0.74	0.65	0.64	0.57	0.72
ι_w	Beta	0.50	0.15	0.59	0.59	0.39	0.78	0.51	0.52	0.29	0.76
ι_p	Beta	0.50	0.15	0.23	0.24	0.10	0.38	0.39	0.46	0.20	0.75
ψ	Beta	0.50	0.15	0.55	0.55	0.36	0.72	0.51	0.53	0.35	0.74
Φ	Normal	1.25	0.13	1.62	1.61	1.48	1.74	1.53	1.55	1.41	1.69
r_π	Normal	1.50	0.25	2.03	2.04	1.75	2.33	1.77	1.76	1.33	2.06
ρ	Beta	0.75	0.10	0.82	0.81	0.77	0.85	0.78	0.78	0.74	0.85
r_y	Normal	0.13	0.05	0.08	0.09	0.05	0.13	0.08	0.11	0.04	0.20
$r_{\Delta y}$	Normal	0.13	0.05	0.22	0.22	0.18	0.27	0.21	0.21	0.17	0.27
π	Gamma	0.63	0.10	0.82	0.79	0.61	0.96	0.65	0.68	0.50	0.85
β	Gamma	0.25	0.10	0.16	0.17	0.08	0.26	0.20	0.23	0.10	0.36
L	Normal	0.00	2.00	-0.10	0.53	-1.30	2.33	0.01	-0.11	-1.50	1.26
γ	Normal	0.40	0.10	0.43	0.43	0.41	0.45	0.38	0.39	0.26	0.53
α	Normal	0.30	0.05	0.19	0.19	0.16	0.22	0.18	0.19	0.15	0.22
Gain	Gamma	0.035	0.03								
σ_a	Invgamma	0.10	2.00	0.45	0.46	0.41	0.51	0.41	0.40	0.35	0.47
σ_b	Invgamma	0.10	2.00	0.24	0.24	0.20	0.28	0.16	0.16	0.10	0.21
σ_g	Invgamma	0.10	2.00	0.52	0.53	0.48	0.58	0.37	0.38	0.32	0.43
σ_q	Invgamma	0.10	2.00	0.45	0.45	0.37	0.53	0.46	0.48	0.36	0.62
σ_r	Invgamma	0.10	2.00	0.24	0.24	0.22	0.27	0.18	0.18	0.15	0.21
σ_p	Invgamma	0.10	2.00	0.14	0.14	0.11	0.17	0.14	0.15	0.12	0.20
σ_w	Invgamma	0.10	2.00	0.24	0.24	0.21	0.28	0.20	0.19	0.13	0.23

Posterior Mode	-841.46		-834.17
Log Data Density (Laplace)	-923.05		-897.77
Log Data Density (Mod.Harm.Mean)	-922.15		-897.78

Table 2: Simulation results for the different learning models and different gain parameters

| | REE model | | beliefs based on MSV regression | | | | | | beliefs based on MSV regression + cte | | | | | | beliefs based on VAR(1) regression + cte | | | | | |
| | | | γ = 0.01 | | γ = 0.02 | | γ = 0.05 | | γ = 0.01 | | γ = 0.02 | | γ = 0.05 | | γ = 0.01 | | γ = 0.02 | | γ = 0.05 | |
	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable
standard deviations																				
Δy	0.94	0.94	0.94	0.94	0.94	0.95	0.98	1.89	0.94	0.95	0.94	0.96	0.99	2.56	0.98	0.88	1.24	1.12	1.66	2.93
Δc	0.69	0.69	0.68	0.69	0.68	0.72	0.72	2.68	0.68	0.69	0.68	0.74	0.73	3.94	0.77	0.75	1.05	1.19	1.64	3.48
Δi	2.37	2.38	2.38	2.40	2.40	2.46	2.56	9.84	2.38	2.40	2.40	2.52	2.59	13.69	1.78	1.53	2.38	3.35	3.48	7.56
Δw	0.55	0.56	0.55	0.56	0.56	0.57	0.62	1.34	0.55	0.56	0.56	0.57	0.63	1.93	0.61	0.51	0.78	0.69	1.11	1.87
L	2.33	2.43	2.20	2.44	2.26	2.46	2.43	5.89	2.20	2.44	2.26	2.50	2.47	9.89	3.00	2.43	3.44	3.41	4.11	10.87
π	0.50	0.51	0.48	0.51	0.50	0.55	0.60	2.40	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	0.55	0.54	0.55	0.54	0.57	0.59	2.50	0.54	0.56	0.54	0.59	0.60	5.16	0.69	0.60	0.89	1.51	1.45	9.39
y	4.30	4.53	4.14	4.54	4.22	4.58	4.46	11.31	4.14	4.54	4.22	4.64	4.52	19.69	5.27	4.27	5.82	5.93	6.61	17.96
c	3.85	4.19	3.59	4.18	3.68	4.27	3.93	12.89	3.59	4.20	3.68	4.38	3.99	23.49	5.16	4.44	6.00	6.05	7.48	21.05
i	10.78	11.25	10.79	11.48	10.95	11.70	11.61	54.21	10.79	11.50	10.97	12.09	11.83	87.24	12.46	12.30	14.32	25.67	16.50	61.23
w	2.37	2.48	2.32	2.48	2.35	2.49	2.51	7.70	2.32	2.47	2.35	2.54	2.54	16.11	2.77	2.28	3.02	3.67	3.70	13.65
autocorrelations																				
Δy	0.27	0.28	0.26	0.28	0.26	0.28	0.28	0.46	0.26	0.28	0.26	0.29	0.31	0.52	0.39	0.30	0.51	0.39	0.59	0.55
Δc	0.31	0.32	0.29	0.32	0.29	0.32	0.32	0.50	0.29	0.32	0.29	0.33	0.35	0.56	0.54	0.53	0.65	0.58	0.72	0.65
Δi	0.59	0.59	0.59	0.59	0.59	0.59	0.60	0.68	0.59	0.59	0.59	0.60	0.61	0.70	0.77	0.73	0.80	0.78	0.77	0.71
Δw	0.16	0.17	0.14	0.17	0.15	0.17	0.17	0.33	0.14	0.17	0.15	0.17	0.21	0.40	0.20	0.12	0.33	0.29	0.43	0.51
L	0.97	0.97	0.96	0.97	0.96	0.97	0.97	0.97	0.96	0.97	0.96	0.97	0.97	0.97	0.98	0.96	0.97	0.97	0.97	0.97
π	0.82	0.82	0.80	0.83	0.81	0.83	0.83	0.81	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	0.89	0.89	0.90	0.89	0.89	0.89	0.91	0.89	0.90	0.89	0.90	0.93	0.93	0.92	0.90	0.93	0.94	0.96	0.97

45

Table 3: Simulation results controlling for escape dynamics

	beliefs based on MSV regression + cte								beliefs based on VAR(1) regression + cte							
	$\gamma = 0.01$				$\gamma = 0.02$				$\gamma = 0.01$				$\gamma = 0.02$			
	including escapes		excluding escapes		including escapes		excluding escapes		including escapes		excluding escapes		including escapes		excluding escapes	
	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable	transition	stable
standard deviations																
Δy	0.94	0.95	0.94	0.94	0.94	0.96	0.94	0.95	0.98	0.88	0.96	0.88	1.24	1.12	1.07	0.88
Δc	0.68	0.69	0.68	0.69	0.68	0.74	0.68	0.70	0.77	0.75	0.75	0.74	1.05	1.19	0.88	0.76
Δi	2.38	2.40	2.38	2.40	2.40	2.52	2.39	2.41	1.78	1.53	1.73	1.52	2.38	3.35	2.08	1.86
Δw	0.55	0.56	0.55	0.56	0.56	0.57	0.55	0.56	0.61	0.51	0.58	0.51	0.78	0.69	0.64	0.56
L	2.20	2.44	2.20	2.44	2.26	2.50	2.25	2.45	3.00	2.43	2.92	2.41	3.44	3.41	3.16	2.38
π	0.48	0.52	0.48	0.52	0.50	0.57	0.49	0.52	0.76	0.56	0.71	0.55	1.20	1.35	0.91	0.66
R	0.54	0.56	0.54	0.56	0.54	0.59	0.54	0.56	0.69	0.60	0.67	0.59	0.89	1.51	0.77	0.79
y	4.14	4.54	4.13	4.55	4.22	4.64	4.22	4.57	5.27	4.27	5.15	4.24	5.82	5.93	5.42	4.16
c	3.59	4.20	3.58	4.20	3.68	4.38	3.68	4.24	5.16	4.44	5.05	4.42	6.00	6.05	5.45	4.04
i	10.79	11.50	10.78	11.50	10.97	12.09	10.93	11.54	12.46	12.30	12.19	12.15	14.32	25.67	13.93	14.08
w	2.32	2.47	2.32	2.47	2.35	2.54	2.35	2.47	2.77	2.28	2.71	2.29	3.02	3.67	2.84	2.67
autocorrelations																
Δy	0.26	0.28	0.26	0.28	0.26	0.29	0.26	0.28	0.39	0.30	0.39	0.30	0.51	0.39	0.48	0.30
Δc	0.29	0.32	0.29	0.32	0.29	0.33	0.29	0.32	0.54	0.53	0.53	0.53	0.65	0.58	0.63	0.53
Δi	0.59	0.59	0.59	0.59	0.59	0.60	0.59	0.59	0.77	0.73	0.77	0.72	0.80	0.78	0.81	0.76
Δw	0.14	0.17	0.14	0.17	0.15	0.17	0.15	0.17	0.20	0.12	0.19	0.13	0.33	0.29	0.30	0.24
L	0.96	0.97	0.96	0.97	0.96	0.97	0.96	0.97	0.98	0.96	0.98	0.96	0.97	0.97	0.98	0.96
π	0.80	0.83	0.80	0.83	0.81	0.83	0.81	0.83	0.85	0.84	0.85	0.84	0.86	0.88	0.87	0.87
R	0.89	0.90	0.89	0.90	0.89	0.90	0.89	0.90	0.92	0.90	0.92	0.90	0.93	0.94	0.93	0.93

Table 4: Regression results for MSV learning with model-consistent initialisation of beliefs

	Prior distribution			Learning with MSV beliefs (excl. constants) Posterior distribution				Learning with MSV beliefs (incl. constants) Posterior distribution			
	Distr.	Mean	St.Dev.	Mode	Mean	5 percent	95 percent	Mode	Mean	5 percent	95 percent
ρ_a	Beta	0.50	0.20	0.96	0.96	0.94	0.98	0.96	0.96	0.94	0.98
ρ_b	Beta	0.50	0.20	0.20	0.23	0.07	0.36	0.20	0.23	0.08	0.37
ρ_g	Beta	0.50	0.20	0.97	0.97	0.96	0.99	0.97	0.97	0.96	0.99
ρ_l	Beta	0.50	0.20	0.71	0.72	0.62	0.82	0.72	0.72	0.61	0.82
ρ_r	Beta	0.50	0.20	0.13	0.16	0.05	0.26	0.13	0.16	0.05	0.26
ρ_p	Beta	0.50	0.20	0.90	0.89	0.81	0.97	0.90	0.89	0.81	0.97
ρ_w	Beta	0.50	0.20	0.97	0.97	0.95	0.99	0.97	0.97	0.95	0.99
μ_p	Beta	0.50	0.20	0.74	0.68	0.50	0.85	0.74	0.68	0.52	0.85
μ_w	Beta	0.50	0.20	0.89	0.84	0.74	0.94	0.89	0.85	0.76	0.94
ρ_{ag}	Normal	0.50	0.25	0.53	0.53	0.39	0.68	0.53	0.53	0.38	0.67
φ	Normal	4.00	1.50	5.30	5.67	3.90	7.35	5.30	5.65	3.94	7.33
σ_c	Normal	1.50	0.38	1.40	1.34	1.12	1.55	1.40	1.35	1.13	1.57
λ	Beta	0.70	0.10	0.71	0.72	0.65	0.79	0.71	0.72	0.66	0.79
ξ_w	Beta	0.50	0.10	0.74	0.70	0.60	0.82	0.74	0.71	0.61	0.81
σ_l	Normal	2.00	0.75	2.03	1.92	0.96	2.88	2.03	1.97	0.98	2.93
ξ_p	Beta	0.50	0.10	0.67	0.65	0.56	0.75	0.67	0.65	0.56	0.74
ι_w	Beta	0.50	0.15	0.62	0.59	0.38	0.80	0.62	0.59	0.39	0.80
ι_p	Beta	0.50	0.15	0.20	0.22	0.09	0.34	0.20	0.22	0.08	0.34
ψ	Beta	0.50	0.15	0.53	0.55	0.37	0.73	0.53	0.55	0.37	0.74
Φ	Normal	1.25	0.13	1.61	1.61	1.48	1.75	1.61	1.61	1.48	1.74
r_π	Normal	1.50	0.25	2.00	2.02	1.74	2.32	2.00	2.02	1.72	2.30
ρ	Beta	0.75	0.10	0.82	0.81	0.77	0.85	0.82	0.81	0.77	0.85
r_y	Normal	0.13	0.05	0.09	0.08	0.05	0.12	0.09	0.09	0.05	0.12
$r_{\Delta y}$	Normal	0.13	0.05	0.22	0.22	0.17	0.26	0.22	0.22	0.17	0.26
π	Gamma	0.63	0.10	0.78	0.81	0.63	0.97	0.78	0.80	0.63	0.97
β	Gamma	0.25	0.10	0.15	0.17	0.07	0.27	0.15	0.17	0.07	0.26
L	Normal	0.00	2.00	0.34	-0.02	-1.67	1.56	0.33	0.02	-1.69	1.61
γ	Normal	0.40	0.10	0.44	0.43	0.41	0.45	0.44	0.43	0.41	0.45
α	Normal	0.30	0.05	0.19	0.19	0.16	0.22	0.19	0.19	0.16	0.22
Gain	Gamma	0.035	0.03	0.012	0.018	0.001	0.034	0.012	0.019	0.002	0.036
σ_a	Invgamma	0.10	2.00	0.45	0.45	0.41	0.50	0.45	0.45	0.41	0.50
σ_b	Invgamma	0.10	2.00	0.24	0.24	0.20	0.28	0.24	0.24	0.20	0.28
σ_g	Invgamma	0.10	2.00	0.52	0.53	0.48	0.58	0.52	0.53	0.48	0.58
σ_l	Invgamma	0.10	2.00	0.45	0.45	0.37	0.53	0.45	0.45	0.37	0.54
σ_r	Invgamma	0.10	2.00	0.24	0.24	0.22	0.27	0.24	0.24	0.22	0.27
σ_p	Invgamma	0.10	2.00	0.14	0.14	0.11	0.17	0.14	0.14	0.11	0.17
σ_w	Invgamma	0.10	2.00	0.24	0.24	0.20	0.28	0.24	0.24	0.20	0.28
Posterior Mode				-837.52				-837.51			
Log Data Density (Laplace)				-922.61				-922.48			
Log Data Density (Mod.Harm.Mean)				-922.56				-922.65			

Table 5: Regression results for MSV learning with optimised initial beliefs

	Prior distribution			MSV beliefs - Optimised initialisation Posterior distribution				Initial Belief Model Posterior distribution
	Distr.	Mean	St.Dev.	Mode	Mean	5 percent	95 percent	Mode
ρ_a	Beta	0.50	0.20	0.98	0.96	0.94	0.99	
ρ_b	Beta	0.50	0.20	0.16	0.23	0.13	0.32	
ρ_g	Beta	0.50	0.20	0.97	0.98	0.96	0.99	
ρ_l	Beta	0.50	0.20	0.32	0.45	0.33	0.56	
ρ_r	Beta	0.50	0.20	0.11	0.15	0.05	0.26	
ρ_p	Beta	0.50	0.20	0.96	0.93	0.88	0.97	
ρ_w	Beta	0.50	0.20	0.97	0.97	0.95	0.99	
μ_p	Beta	0.50	0.20	0.87	0.84	0.78	0.90	
μ_w	Beta	0.50	0.20	0.79	0.72	0.64	0.81	
ρ_{ag}	Normal	0.50	0.25	0.54	0.53	0.41	0.66	
φ	Normal	4.00	1.50	2.83	4.33	3.06	5.54	3.81
σ_c	Normal	1.50	0.38	0.93	0.99	0.85	1.12	1.30
λ	Beta	0.70	0.10	0.79	0.80	0.75	0.84	0.69
ξ_w	Beta	0.50	0.10	0.64	0.63	0.57	0.69	0.68
σ_l	Normal	2.00	0.75	1.46	1.61	0.80	2.35	1.97
ξ_p	Beta	0.50	0.10	0.60	0.63	0.59	0.67	0.77
ι_w	Beta	0.50	0.15	0.59	0.53	0.33	0.71	0.59
ι_p	Beta	0.50	0.15	0.18	0.22	0.10	0.34	0.51
ψ	Beta	0.50	0.15	0.46	0.38	0.27	0.49	0.62
Φ	Normal	1.25	0.13	1.56	1.60	1.50	1.71	1.25
r_π	Normal	1.50	0.25	1.95	1.91	1.58	2.22	1.70
ρ	Beta	0.75	0.10	0.84	0.84	0.80	0.88	0.64
r_y	Normal	0.13	0.05	0.14	0.13	0.07	0.18	0.03
$r_{\Delta y}$	Normal	0.13	0.05	0.21	0.19	0.15	0.24	0.15
π	Gamma	0.63	0.10	0.59	0.61	0.50	0.73	0.57
β	Gamma	0.25	0.10	0.27	0.26	0.14	0.38	0.25
L	Normal	0.00	2.00	1.20	0.92	-0.14	2.00	1.26
γ	Normal	0.40	0.10	0.45	0.43	0.41	0.45	0.45
α	Normal	0.30	0.05	0.18	0.18	0.16	0.19	0.19
Gain	Gamma	0.04	0.03	0.02	0.01	0.01	0.02	
σ_a	Invgamma	0.10	2.00	0.46	0.47	0.42	0.52	
σ_b	Invgamma	0.10	2.00	0.26	0.25	0.22	0.28	
σ_g	Invgamma	0.10	2.00	0.51	0.53	0.48	0.58	
σ_l	Invgamma	0.10	2.00	0.66	0.61	0.53	0.68	
σ_r	Invgamma	0.10	2.00	0.23	0.24	0.21	0.26	
σ_p	Invgamma	0.10	2.00	0.14	0.14	0.12	0.16	
σ_w	Invgamma	0.10	2.00	0.24	0.23	0.20	0.26	

Posterior Mode	804.01
Log Data Density (Laplace)	-888.99
Log Data Density (Mod.Harm.Mean)	-910.97

Tabel 6: Regression results for MSV learning with initial beliefs based on a pre-sample model

	Prior distribution			MSV beliefs - presample initialisation Posterior distribution				Pre-Sample Belief Model Posterior distribution
	Distr.	Mean	St.Dev.	Mode	Mean	5 percent	95 percent	Mode
ρ_a	Beta	0.50	0.20	0.97	0.97	0.95	0.99	0.87
ρ_b	Beta	0.50	0.20	0.93	0.94	0.91	0.97	0.97
ρ_g	Beta	0.50	0.20	0.99	0.99	0.98	1.00	0.88
ρ_l	Beta	0.50	0.20	0.40	0.44	0.35	0.52	0.47
ρ_r	Beta	0.50	0.20	0.05	0.07	0.01	0.12	0.31
ρ_p	Beta	0.50	0.20	0.96	0.96	0.93	0.99	0.52
ρ_w	Beta	0.50	0.20	0.98	0.97	0.95	0.98	0.46
μ_p	Beta	0.50	0.20	0.73	0.65	0.55	0.75	0.42
μ_w	Beta	0.50	0.20	0.77	0.71	0.66	0.78	0.51
ρ_{ag}	Normal	0.50	0.25	0.57	0.56	0.43	0.70	0.64
φ	Normal	4.00	1.50	3.11	4.61	3.20	6.05	3.62
σ_c	Normal	1.50	0.38	1.13	1.04	0.85	1.22	0.94
λ	Beta	0.70	0.10	0.62	0.68	0.64	0.73	0.52
ξ_w	Beta	0.50	0.10	0.58	0.60	0.54	0.66	0.83
σ_l	Normal	2.00	0.75	1.73	1.86	1.14	2.60	1.69
ξ_p	Beta	0.50	0.10	0.53	0.53	0.46	0.61	0.41
ι_w	Beta	0.50	0.15	0.49	0.45	0.29	0.63	0.43
ι_p	Beta	0.50	0.15	0.52	0.52	0.34	0.69	0.37
ψ	Beta	0.50	0.15	0.45	0.42	0.35	0.48	0.50
Φ	Normal	1.25	0.13	1.51	1.48	1.39	1.57	1.58
r_π	Normal	1.50	0.25	1.97	1.97	1.71	2.26	1.30
ρ	Beta	0.75	0.10	0.79	0.78	0.74	0.81	0.97
r_y	Normal	0.13	0.05	0.16	0.16	0.12	0.19	0.15
$r_{\Delta y}$	Normal	0.13	0.05	0.26	0.26	0.22	0.30	0.06
π	Gamma	0.63	0.10	0.46	0.47	0.40	0.55	0.57
β	Gamma	0.25	0.10	0.20	0.27	0.12	0.39	0.21
L	Normal	0.00	2.00	2.59	2.61	2.11	3.13	0.90
γ	Normal	0.40	0.10	0.45	0.45	0.42	0.49	0.58
α	Normal	0.30	0.05	0.21	0.20	0.19	0.21	0.20
Gain	Gamma	0.04	0.03	0.02	0.02	0.02	0.03	
σ_a	Invgamma	0.10	2.00	0.46	0.48	0.43	0.53	0.59
σ_b	Invgamma	0.10	2.00	0.10	0.11	0.09	0.12	0.05
σ_g	Invgamma	0.10	2.00	0.52	0.53	0.48	0.58	0.91
σ_l	Invgamma	0.10	2.00	0.58	0.56	0.50	0.61	0.81
σ_r	Invgamma	0.10	2.00	0.23	0.24	0.21	0.27	0.09
σ_p	Invgamma	0.10	2.00	0.20	0.19	0.17	0.21	0.43
σ_w	Invgamma	0.10	2.00	0.28	0.27	0.23	0.31	0.35

Posterior Mode	830.57
Log Data Density (Laplace)	-915.54
Log Data Density (Mod.Harm.Mean)	-944.37

Table 7: Regression results for VAR learning with model-consistent initialisation of beliefs

	Prior distribution			VAR beliefs - REE initialisation Posterior distribution			
	Distr.	Mean	St.Dev.	Mode	Mean	5 percent	95 percent
ρ_a	Beta	0.50	0.20	0.95	0.95	0.93	0.98
ρ_b	Beta	0.50	0.20	0.19	0.22	0.08	0.35
ρ_g	Beta	0.50	0.20	0.99	0.98	0.97	1.00
ρ_l	Beta	0.50	0.20	0.54	0.54	0.45	0.64
ρ_r	Beta	0.50	0.20	0.12	0.15	0.05	0.25
ρ_p	Beta	0.50	0.20	0.91	0.88	0.82	0.95
ρ_w	Beta	0.50	0.20	0.97	0.95	0.90	0.99
μ_p	Beta	0.50	0.20	0.57	0.59	0.41	0.76
μ_w	Beta	0.50	0.20	0.88	0.81	0.72	0.91
ρ_{ag}	Normal	0.50	0.25	0.51	0.51	0.36	0.67
φ	Normal	4.00	1.50	3.89	4.18	2.63	5.79
σ_c	Normal	1.50	0.38	1.37	1.34	1.10	1.59
λ	Beta	0.70	0.10	0.72	0.73	0.65	0.81
ξ_w	Beta	0.50	0.10	0.74	0.72	0.62	0.82
σ_l	Normal	2.00	0.75	2.14	2.03	1.03	3.01
ξ_p	Beta	0.50	0.10	0.68	0.68	0.60	0.76
ι_w	Beta	0.50	0.15	0.61	0.55	0.36	0.75
ι_p	Beta	0.50	0.15	0.07	0.14	0.05	0.23
ψ	Beta	0.50	0.15	0.46	0.49	0.30	0.66
Φ	Normal	1.25	0.13	1.64	1.63	1.50	1.76
r_π	Normal	1.50	0.25	1.95	1.98	1.68	2.29
ρ	Beta	0.75	0.10	0.82	0.81	0.77	0.86
r_y	Normal	0.13	0.05	0.09	0.09	0.05	0.12
$r_{\Delta y}$	Normal	0.13	0.05	0.22	0.22	0.17	0.25
π	Gamma	0.63	0.10	0.76	0.79	0.60	0.95
β	Gamma	0.25	0.10	0.15	0.16	0.07	0.25
L	Normal	0.00	2.00	0.47	0.33	-1.40	2.23
γ	Normal	0.40	0.10	0.42	0.42	0.39	0.45
α	Normal	0.30	0.05	0.18	0.19	0.16	0.22
Gain	Gamma	0.035	0.030	0.001	0.002	0.000	0.004
σ_a	Invgamma	0.10	2.00	0.45	0.46	0.41	0.50
σ_b	Invgamma	0.10	2.00	0.28	0.29	0.26	0.32
σ_g	Invgamma	0.10	2.00	0.52	0.53	0.48	0.58
σ_l	Invgamma	0.10	2.00	0.90	0.92	0.83	1.00
σ_r	Invgamma	0.10	2.00	0.24	0.25	0.22	0.27
σ_p	Invgamma	0.10	2.00	0.06	0.08	0.06	0.11
σ_w	Invgamma	0.10	2.00	0.28	0.28	0.25	0.31

|		
Posterior Mode	834.61	
Log Data Density (Laplace)	-921.74	
Log Data Density (Mod.Harm.Mean)	-921.65	

Table 8: Regression results for VAR learning with optimised initial beliefs

| | Prior distribution | | | VAR beliefs - Optimised initialisation | | | | Initial Belief Model |
| | | | | Posterior distribution | | | | Posterior distribution |
	Distr.	Mean	St.Dev.	Mode	Mean	5 percent	95 percent	Mode
ρ_a	Beta	0.50	0.20	0.96	0.95	0.92	0.99	
ρ_b	Beta	0.50	0.20	0.12	0.14	0.04	0.25	
ρ_g	Beta	0.50	0.20	0.97	0.97	0.95	0.99	
ρ_l	Beta	0.50	0.20	0.56	0.55	0.44	0.65	
ρ_r	Beta	0.50	0.20	0.12	0.16	0.04	0.26	
ρ_p	Beta	0.50	0.20	0.95	0.93	0.88	0.98	
ρ_w	Beta	0.50	0.20	0.88	0.80	0.62	0.97	
μ_p	Beta	0.50	0.20	0.57	0.59	0.45	0.75	
μ_w	Beta	0.50	0.20	0.67	0.58	0.38	0.80	
ρ_{ag}	Normal	0.50	0.25	0.52	0.52	0.36	0.68	
φ	Normal	4.00	1.50	3.39	3.74	2.37	5.09	3.02
σ_c	Normal	1.50	0.38	1.01	1.06	0.77	1.33	1.14
λ	Beta	0.70	0.10	0.74	0.74	0.64	0.83	0.73
ξ_w	Beta	0.50	0.10	0.71	0.73	0.65	0.81	0.57
σ_l	Normal	2.00	0.75	2.33	2.31	1.26	3.30	1.44
ξ_p	Beta	0.50	0.10	0.62	0.62	0.55	0.69	0.62
ι_w	Beta	0.50	0.15	0.57	0.53	0.35	0.73	0.43
ι_p	Beta	0.50	0.15	0.07	0.10	0.03	0.16	0.26
ψ	Beta	0.50	0.15	0.50	0.49	0.27	0.69	0.40
Φ	Normal	1.25	0.13	1.58	1.59	1.45	1.72	1.37
r_π	Normal	1.50	0.25	1.85	1.84	1.53	2.15	1.70
ρ	Beta	0.75	0.10	0.85	0.86	0.82	0.90	0.57
r_y	Normal	0.13	0.05	0.12	0.12	0.06	0.17	0.04
$r_{\Delta y}$	Normal	0.13	0.05	0.20	0.20	0.16	0.25	0.17
π	Gamma	0.63	0.10	0.74	0.75	0.58	0.94	0.74
β	Gamma	0.25	0.10	0.19	0.21	0.10	0.32	0.21
L	Normal	0.00	2.00	-0.60	-0.54	-2.05	0.81	-0.49
γ	Normal	0.40	0.10	0.42	0.42	0.40	0.44	0.42
α	Normal	0.30	0.05	0.17	0.18	0.15	0.21	0.26
Gain	Gamma	0.035	0.030	0.001	0.002	0.001	0.003	
σ_a	Invgamma	0.10	2.00	0.46	0.47	0.42	0.51	
σ_b	Invgamma	0.10	2.00	0.27	0.27	0.24	0.30	
σ_g	Invgamma	0.10	2.00	0.52	0.53	0.48	0.57	
σ_l	Invgamma	0.10	2.00	0.91	0.93	0.84	1.02	
σ_r	Invgamma	0.10	2.00	0.23	0.23	0.21	0.26	
σ_p	Invgamma	0.10	2.00	0.06	0.07	0.05	0.08	
σ_w	Invgamma	0.10	2.00	0.29	0.29	0.26	0.32	

Posterior Mode	821.12
Log Data Density (Laplace)	-903.91
Log Data Density (Mod.Harm.Mean)	-904.29

Table 9: Regression results for VAR learning with initial beliefs based on a pre-sample model

| | Prior distribution | | | VAR beliefs - Pre-sample initialisation | | | | Initial Belief Model |
| | | | | Posterior distribution | | | | Posterior distribution |
	Distr.	Mean	St.Dev.	Mode	Mean	5 percent	95 percent	Mode
ρ_a	Beta	0.50	0.20	0.97	0.96	0.93	0.99	0.87
ρ_b	Beta	0.50	0.20	0.21	0.22	0.09	0.34	0.97
ρ_g	Beta	0.50	0.20	0.99	0.99	0.99	1.00	0.88
ρ_l	Beta	0.50	0.20	0.65	0.66	0.56	0.77	0.47
ρ_r	Beta	0.50	0.20	0.11	0.14	0.04	0.24	0.31
ρ_p	Beta	0.50	0.20	0.93	0.89	0.80	0.97	0.52
ρ_w	Beta	0.50	0.20	0.68	0.70	0.52	0.91	0.46
μ_p	Beta	0.50	0.20	0.68	0.64	0.50	0.79	0.42
μ_w	Beta	0.50	0.20	0.49	0.49	0.24	0.72	0.51
ρ_{ag}	Normal	0.50	0.25	0.54	0.53	0.38	0.68	0.64
φ	Normal	4.00	1.50	4.25	4.64	3.13	6.18	3.62
σ_c	Normal	1.50	0.38	1.06	1.11	0.88	1.34	0.94
λ	Beta	0.70	0.10	0.79	0.78	0.72	0.84	0.52
ξ_w	Beta	0.50	0.10	0.75	0.75	0.69	0.82	0.83
σ_l	Normal	2.00	0.75	2.62	2.55	1.58	3.50	1.69
ξ_p	Beta	0.50	0.10	0.61	0.63	0.51	0.74	0.41
ι_w	Beta	0.50	0.15	0.47	0.47	0.29	0.66	0.43
ι_p	Beta	0.50	0.15	0.57	0.59	0.40	0.80	0.37
ψ	Beta	0.50	0.15	0.51	0.49	0.29	0.69	0.50
Φ	Normal	1.25	0.13	1.57	1.57	1.44	1.71	1.58
r_π	Normal	1.50	0.25	1.78	1.75	1.43	2.08	1.30
ρ	Beta	0.75	0.10	0.87	0.87	0.84	0.91	0.97
r_y	Normal	0.13	0.05	0.13	0.13	0.07	0.19	0.15
$r_{\Delta y}$	Normal	0.13	0.05	0.20	0.20	0.16	0.24	0.06
π	Gamma	0.63	0.10	0.70	0.75	0.56	0.93	0.57
β	Gamma	0.25	0.10	0.18	0.21	0.09	0.34	0.21
L	Normal	0.00	2.00	0.44	0.69	-0.70	2.14	0.90
γ	Normal	0.40	0.10	0.44	0.43	0.40	0.47	0.58
α	Normal	0.30	0.05	0.18	0.18	0.15	0.21	0.20
Gain	Gamma	0.035	0.030	0.016	0.017	0.012	0.022	
σ_a	Invgamma	0.10	2.00	0.46	0.46	0.41	0.51	0.59
σ_b	Invgamma	0.10	2.00	0.29	0.29	0.26	0.32	0.05
σ_g	Invgamma	0.10	2.00	0.52	0.53	0.48	0.58	0.91
σ_l	Invgamma	0.10	2.00	0.95	0.97	0.88	1.07	0.81
σ_r	Invgamma	0.10	2.00	0.23	0.24	0.21	0.26	0.09
σ_p	Invgamma	0.10	2.00	0.20	0.20	0.18	0.23	0.43
σ_w	Invgamma	0.10	2.00	0.29	0.30	0.27	0.33	0.35

Posterior Mode	859.07	
Log Data Density (Laplace)	-946.19	
Log Data Density (Mod.Harm.Mean)	-938.17	

Figure 1: IRF of a monetary policy and a productivity shock on inflation: DSGE versus DSGE-VAR
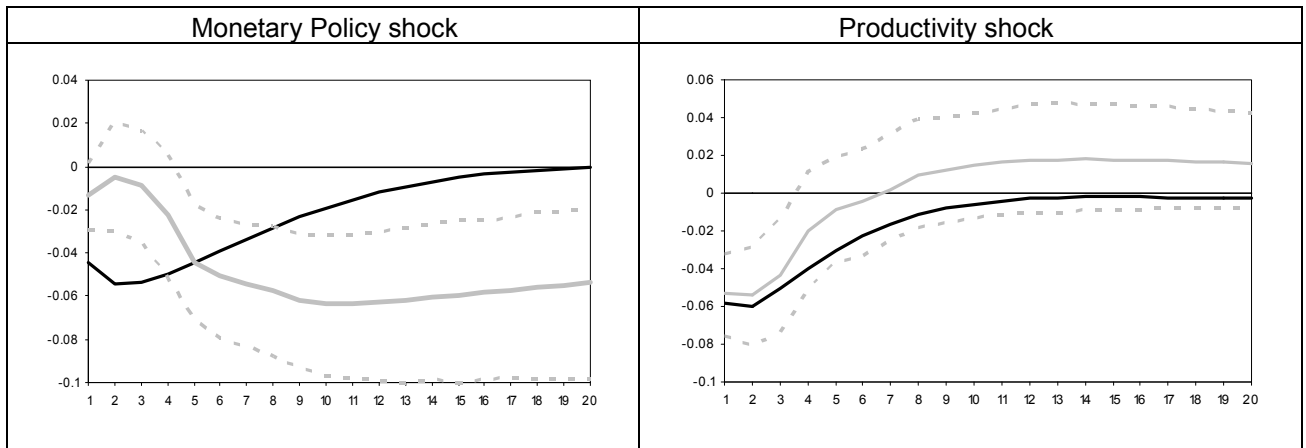
Grey line: the benchmark DSGE-VAR IRF (mode in bold and 90% interval).

Black line: the REE-DSGE IRF.

Figure 2a: Simulation profile for MSV learning with different gains:
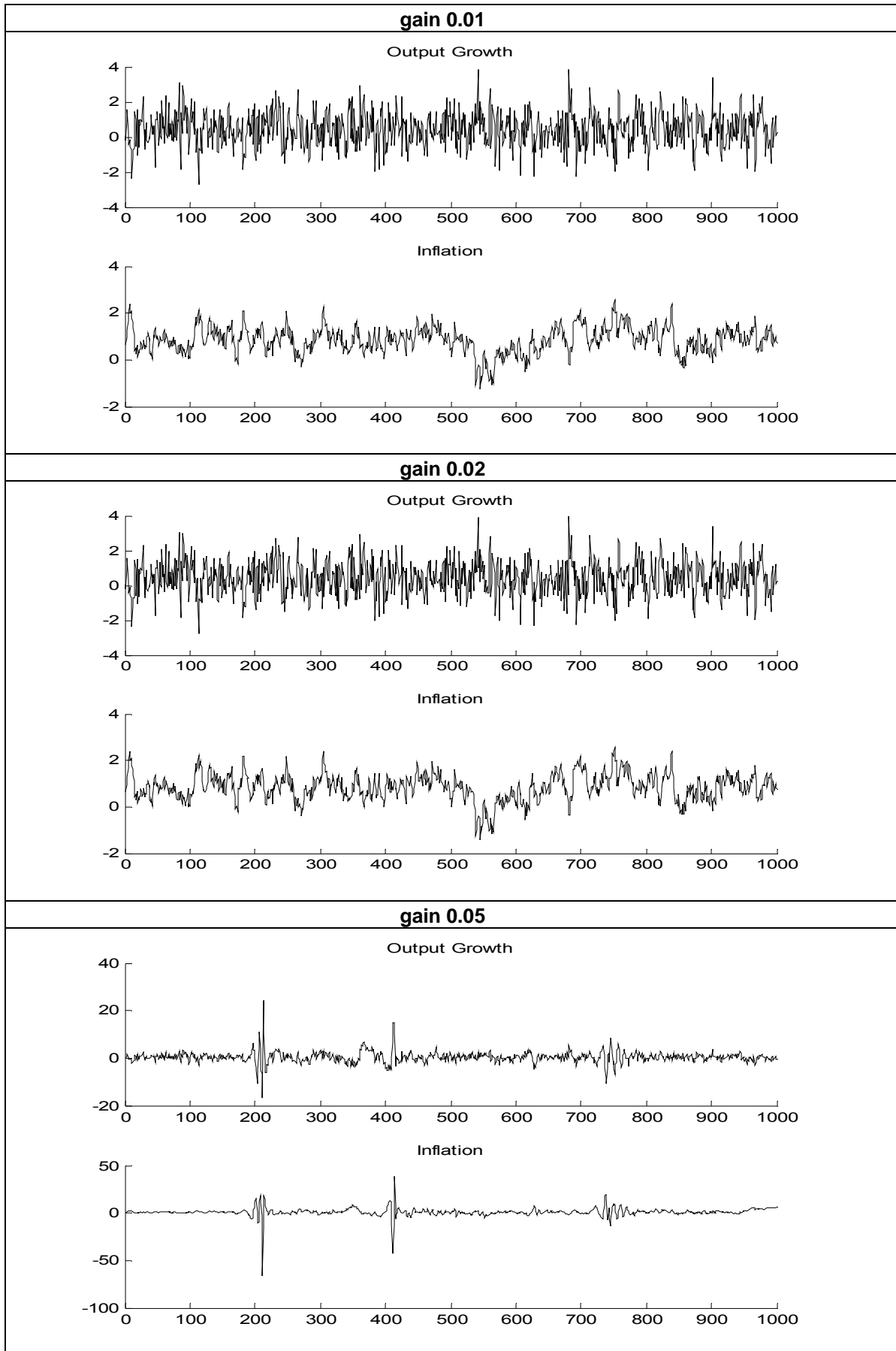
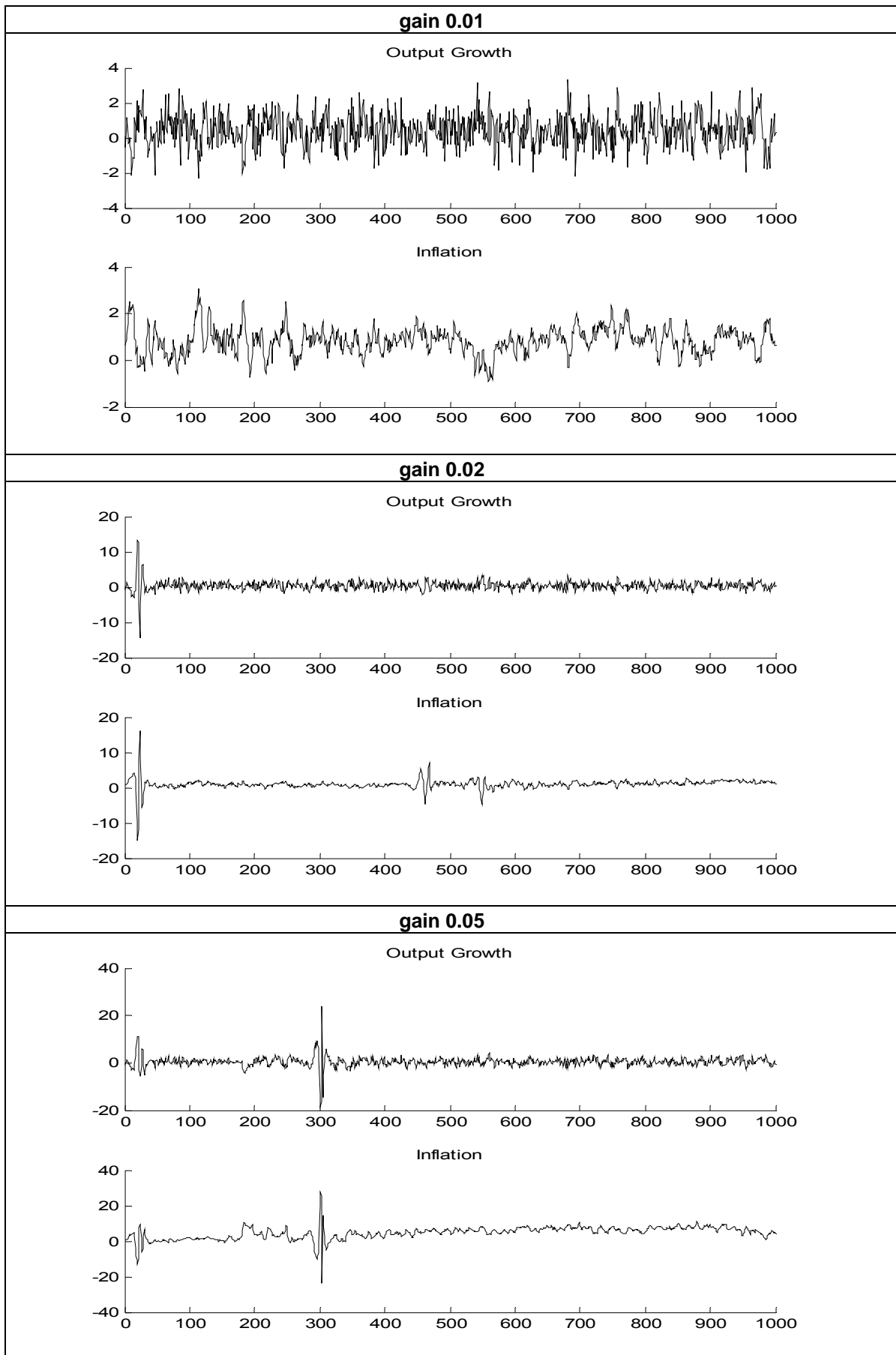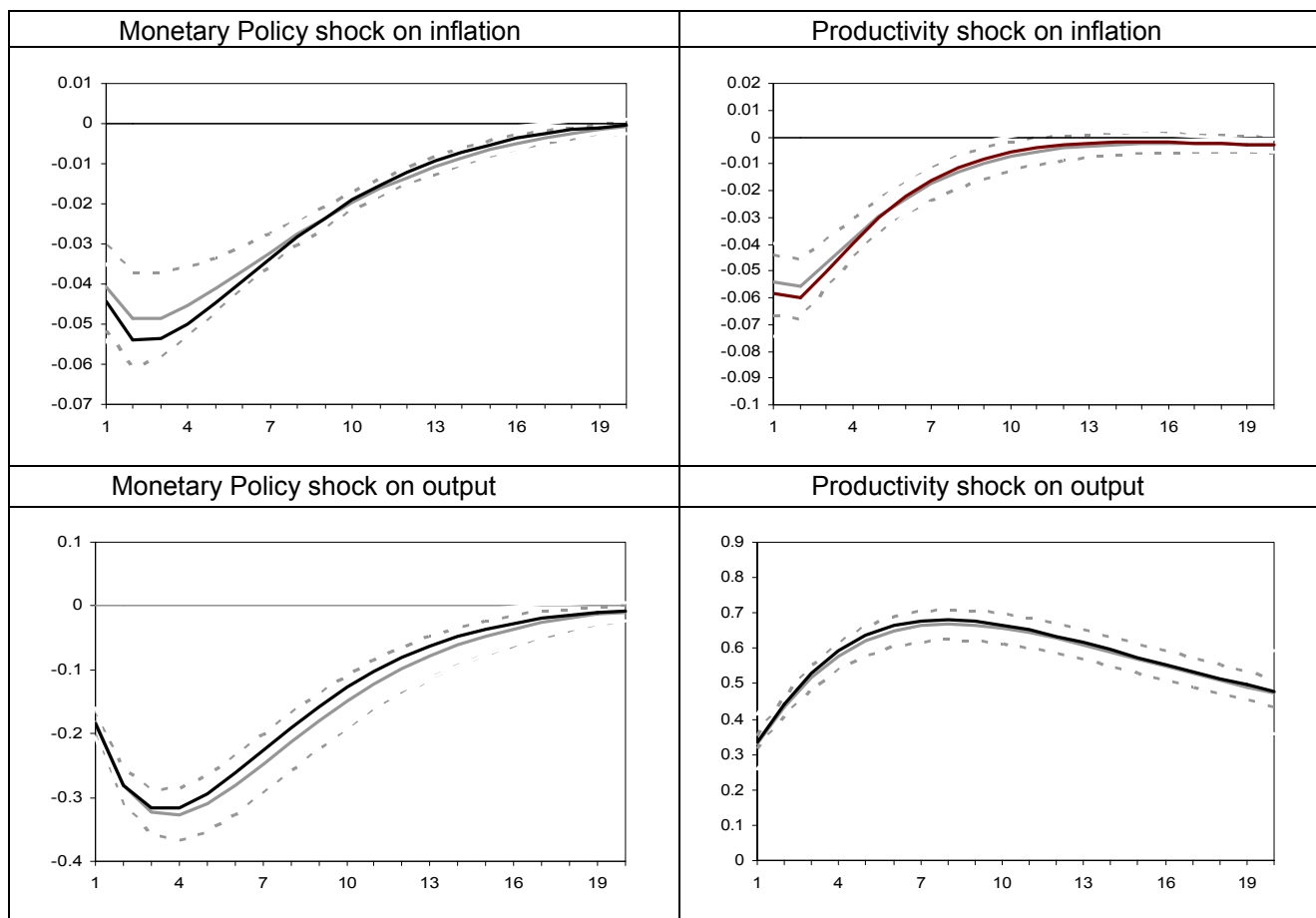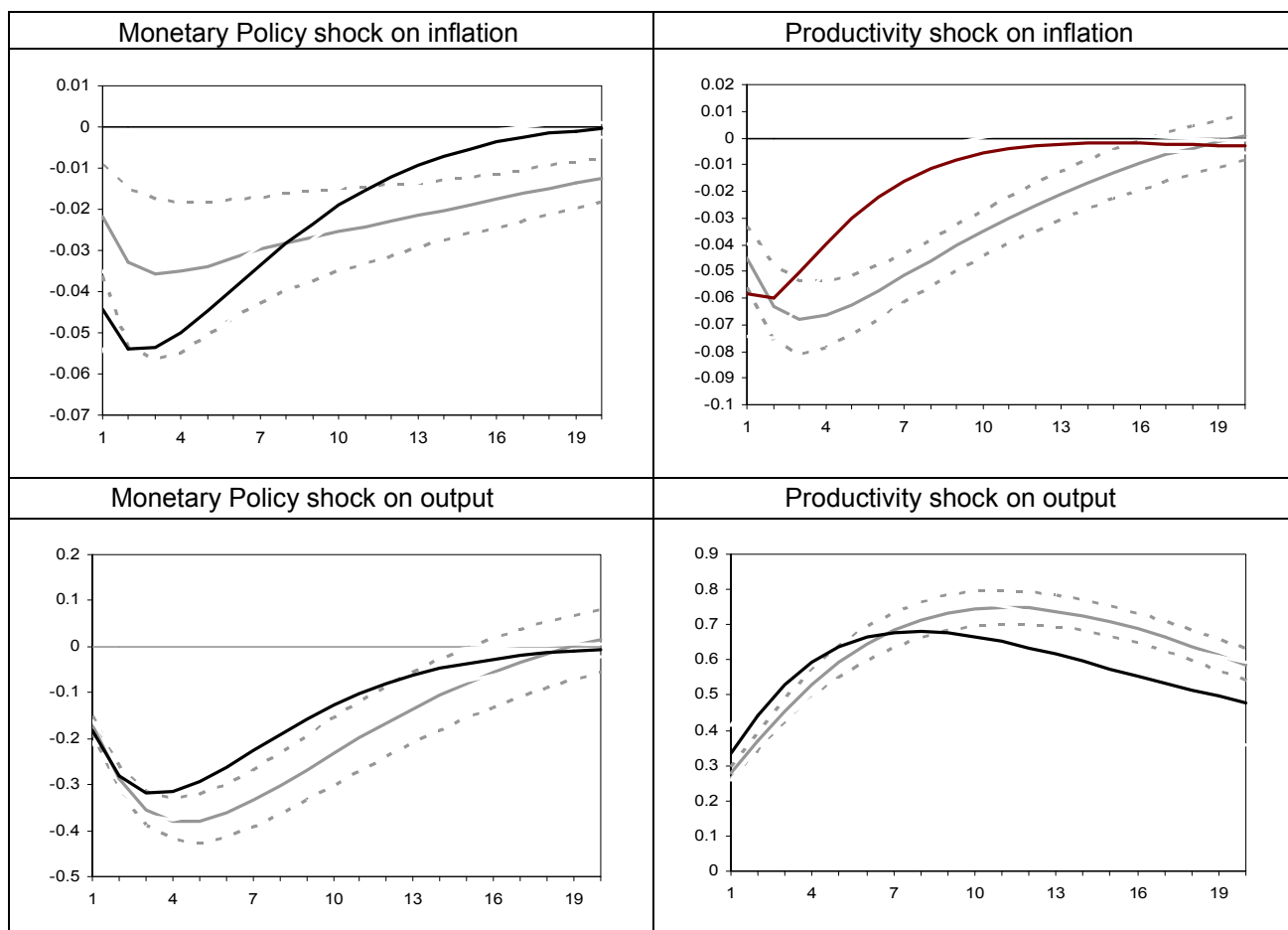**Table 2b: Simulation profile for VAR learning and different gains:**

gain 0.01

Output Growth

Inflation

gain 0.02

Output Growth

Inflation

gain 0.05

Output Growth

Inflation

Figure 3a : IRF for a monetary policy and a productivity shock on output and inflation: MSV learning for different initial beliefs (gain=.02)

Gray line : IRF for different initial beliefs based on draws from the estimated posterior distribution of the REE model (median and 10-90% deciles)

Black line: median for the DSGE-REE model.

Figure 3b : IRF for a monetary policy and a productivity shock on output and inflation: VAR learning for different initial beliefs (gain=.02)

| | Monetary Policy shock on inflation | Productivity shock on inflation |
| Monetary Policy shock on output | Productivity shock on output |

Gray line : IRF for different initial beliefs based on draws from the estimated posterior distribution of the REE model (median and 10-90% deciles)

Black line: median for the DSGE-REE model.

Figure 4: Differences in the reaction of realised and expected inflation following a monetary policy shock under RE and learning
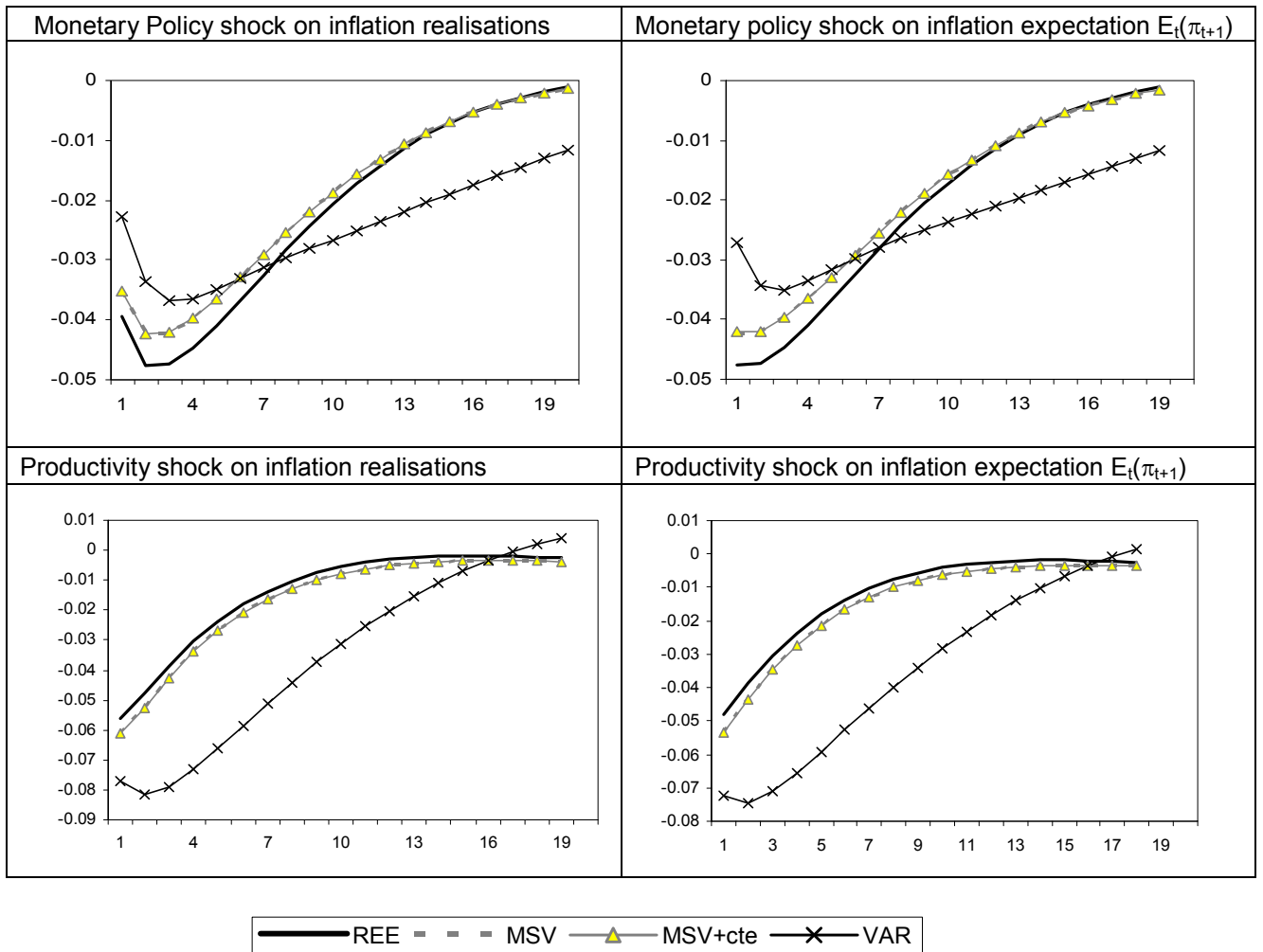
IRF evaluated at the mode for the REE model, and with a random initial belief for the learning models

Figure 5: 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)
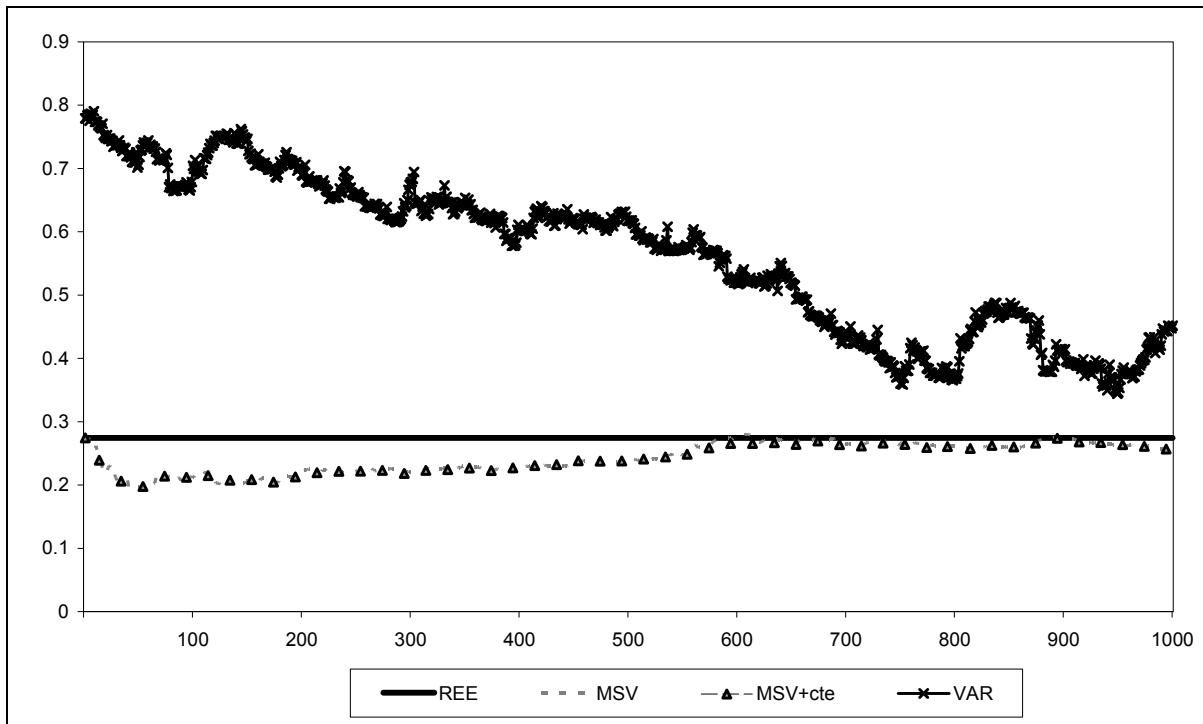
Figure 6: Relation between logarithm of non-escape probability and first escape time
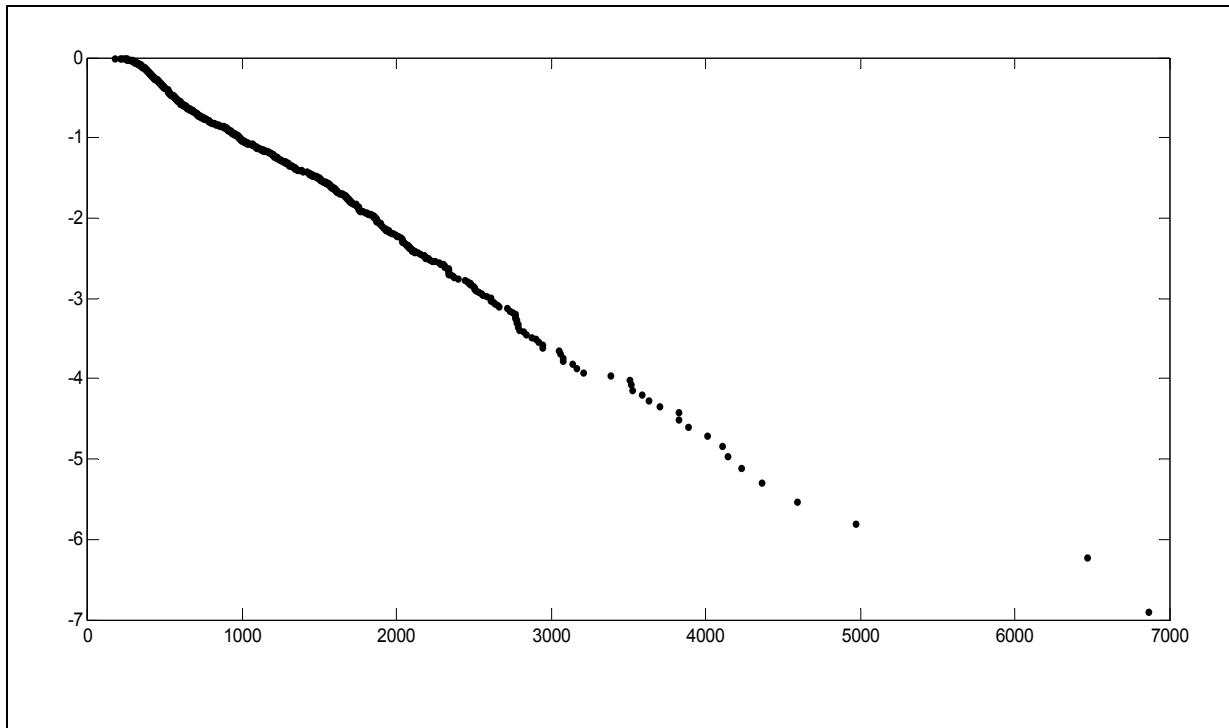
Simulations with MSV+cte learning and gain = 0.03. A large deviation is defined as a realisation where a state vector variable takes on a value that exceeds four times the standard deviations (observed in the REE) away from the steady state.

Figure 7: Impulse Response Functions for the MSV model with optimised initial beliefs
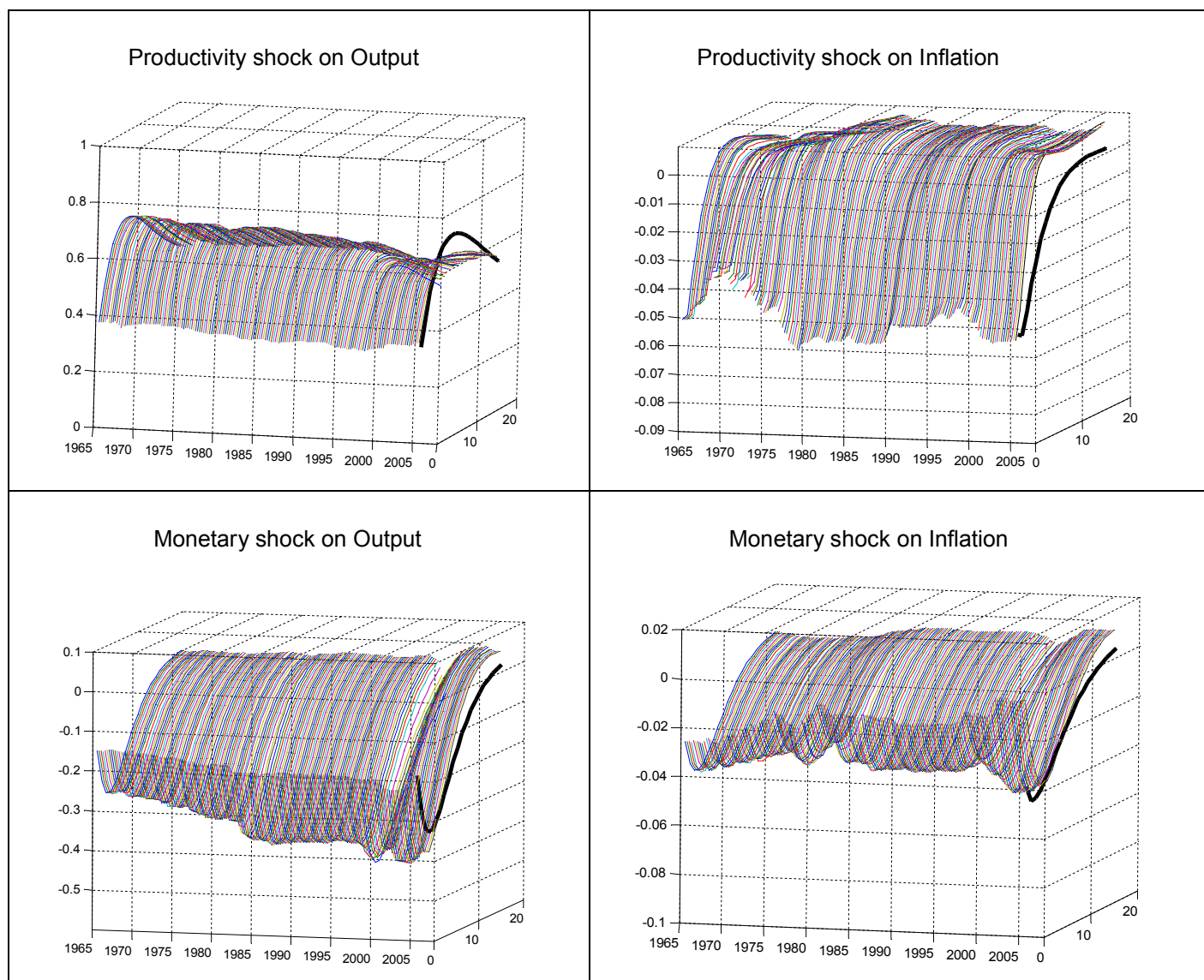
Figure 8: Impulse Response Functions for the MSV model with pre-sample based initial beliefs
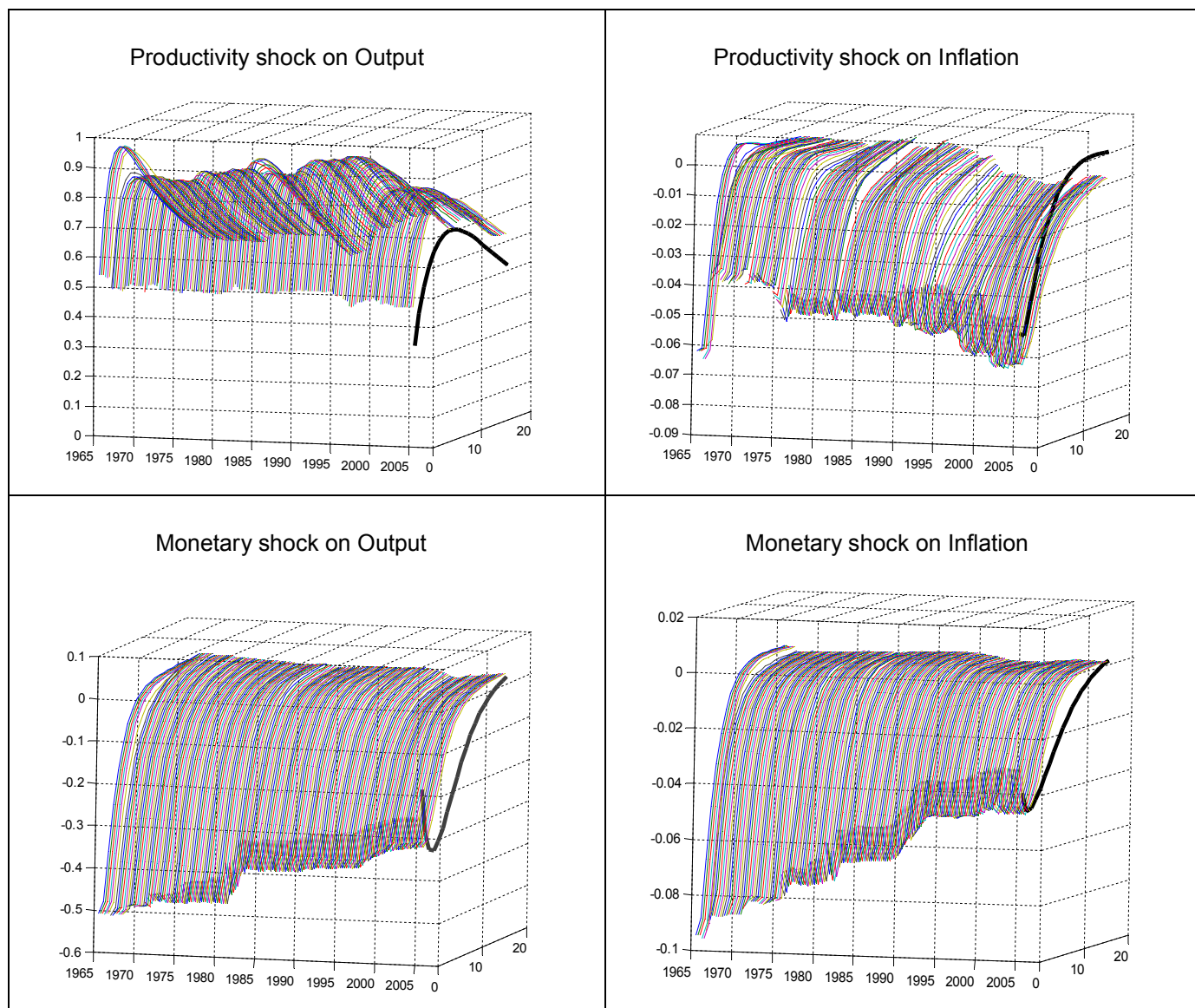
Figure 9: Impulse Response Functions for the VAR beliefs with model consistent initial beliefs
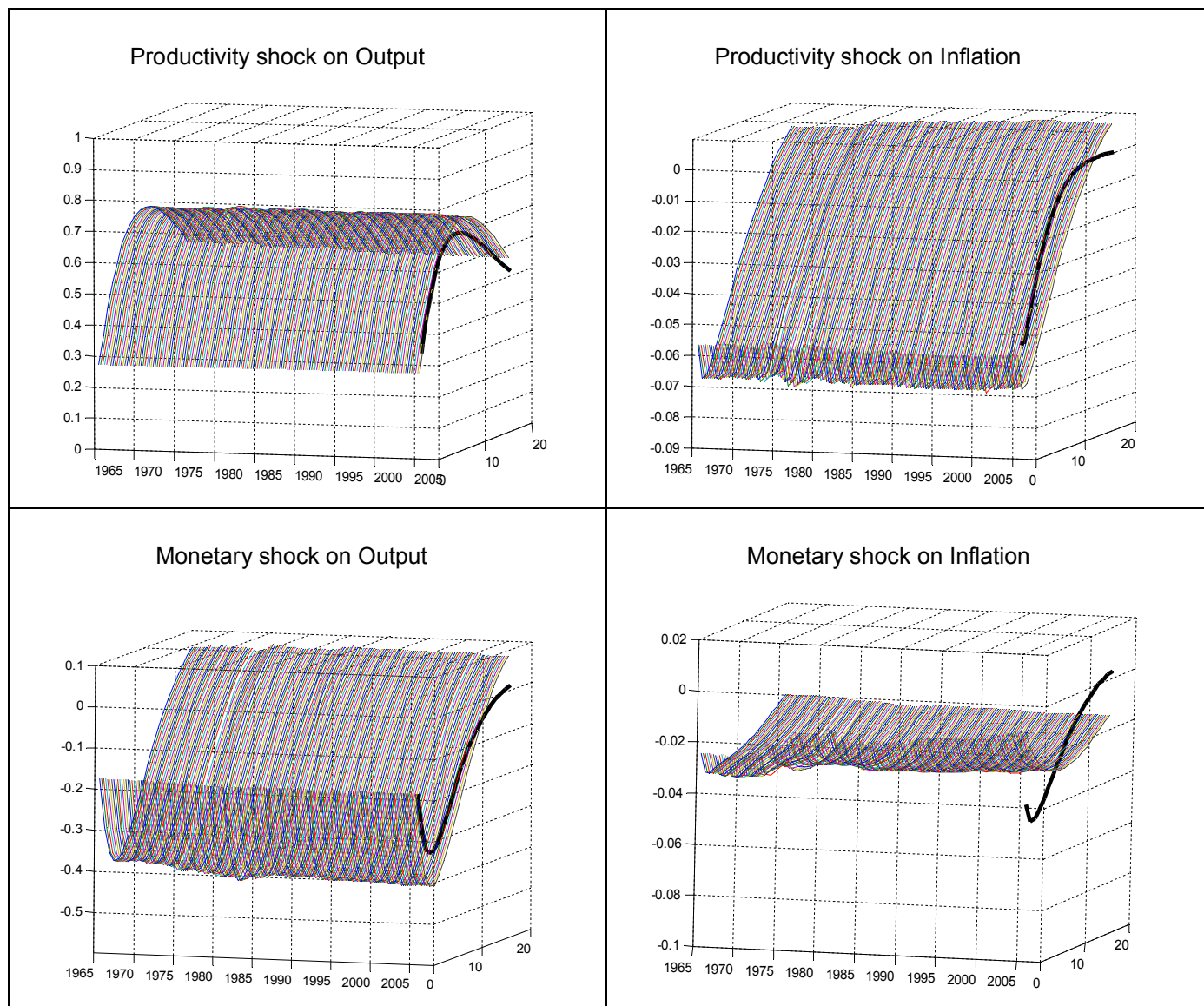
Figure 10: Impulse Response Functions for the VAR beliefs with optimised initial beliefs
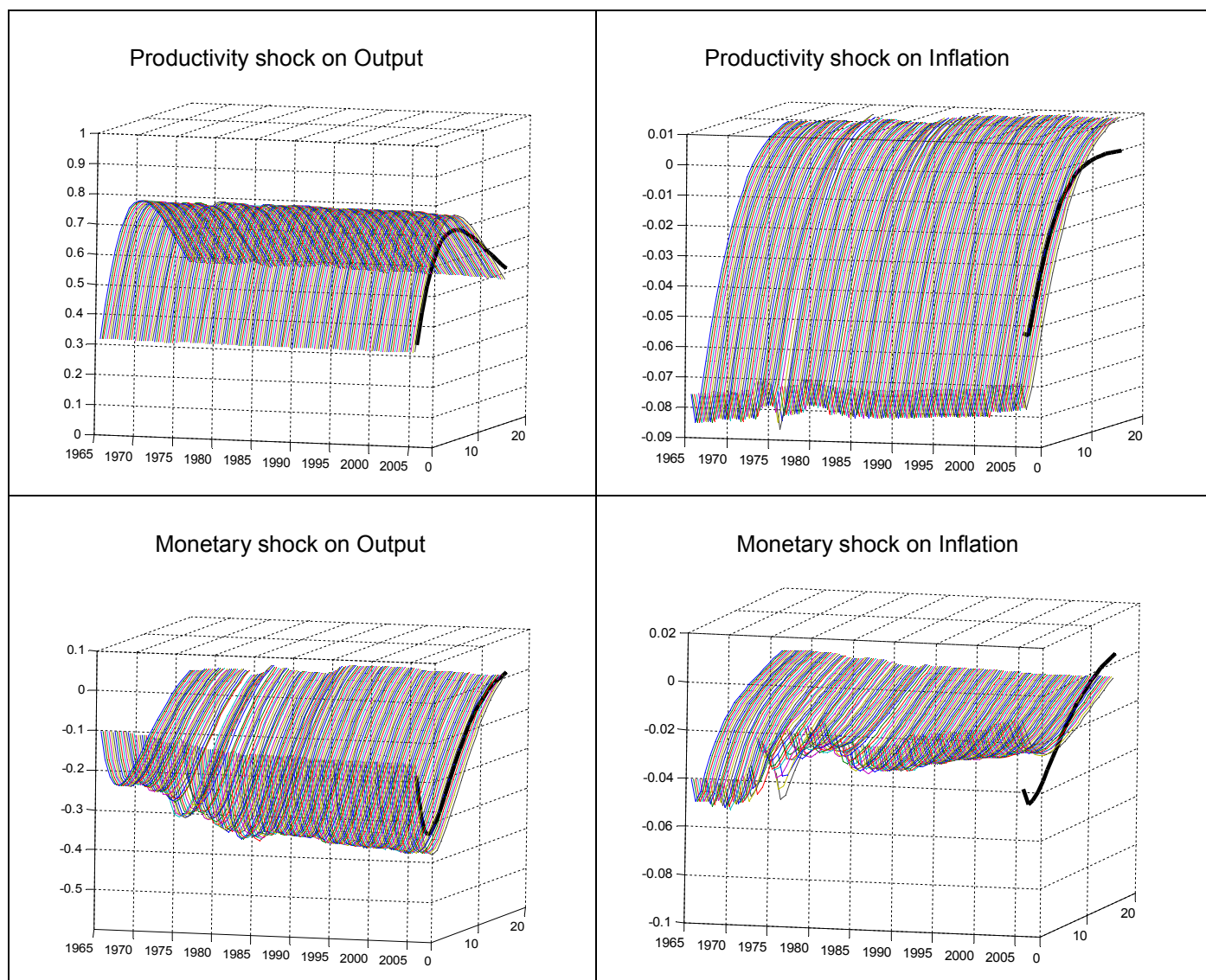
Figure 11: Impulse Response Functions for the VAR beliefs with pre-sample based initial beliefs

CERGE-EI
P.O.BOX 882
Politických vězňů 7
111 21 Praha 1
Czech Republic
http://www.cerge-ei.cz