Bayesian estimation of DSGE models under adaptive learning: robustness issues

Yuliya Rychalovska*(CERGE–EI) and Sergey Slobodyan†(CERGE–EI)‡

August 2010

Abstract

We evaluate model fit, estimated parameters, and perceived persistence of inflation in several DSGE models of Euro area estimated under adaptive learning and rational expectations (RE). We systematically vary model size, information set available to the learning agents, and the way of forming agents' initial beliefs. We find that assuming adaptive expectations results in better model fit than if RE is used, especially when the agents use very little information to form their beliefs. Initial beliefs which are restricted to be consistent with the estimated RE equilibrium are found to be rather robust, as varying them significantly results in an essentially identical estimation. Pre-sample regression based initial beliefs, while more consistent with the in-sample data on average than REE-consistent initial beliefs, suffer from significant volatility and result in worse model fit. Estimated parameters and the model fit depend significantly on the information set used by the agents, which might explain widely divergent result of previous estimations under AL.

JEL classification: C11,D84,E30,E52
Keywords: DSGE models, Bayesian estimation, adaptive learning

*Email: yuliya.rychalovska@cerge-ei.cz
†Email: sergey.slobodyan@cerge-ei.cz
‡A joint workplace of the Center for Economic Research and Graduate Education, Charles University, Prague, and the Economics Institute of the Academy of Sciences of the Czech Republic. Address: CERGE-EI, P.O. Box 882, Politických vězňů 7, Prague 1, 111 21, Czech Republic

1 Introduction

The recent ability to implement advanced econometric techniques for systematic policy analysis has encouraged a large literature on building and estimating DSGE models. Matching empirically observed features of the data, for example persistence and hump-shaped Impulse Response Funcions (IRF) of key macroeconomic variables such as inflation, output, employment, etc., necessitates inclusion of a variety of rigidities into a standard micro-founded New-Keynesian framework model. These rigidities, both real (habit formation in consumption, investment adjustment costs, variable capital utilization, fixed costs) and nominal (Calvo prices and wages, partial indexation of prices and wages to past inflation), enable models to capture the dynamic properties of observed data, see Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2003, 2007). For example, the inclusion of "mechanical" endogenous persistence mechanisms, such as habit formation and price indexation, can influence the consumption and inflation dynamics and considerably change the overall performance of the model. For that reason, the empirical literature attempts to assess the validity of alternative modelling assumptions and evaluate the ability of various DSGE models to fit macroeconomic data. This issue gains further relevance when considering the growing interest in the application of micro-founded DSGE models to policy making in central banks. In particular, misspecification of the model's microfoundations may affect the welfare criteria and result in an inaccurate ranking of alternative policy regimes.

Some of the rigidities that were used in DSGE models recently, for example partial inflation indexation, have been criticized as being ad-hoc and having no theoretical foundation, see Chari, Kehoe, and McGrattan (2009). Bayesian estimation of the models requires the number of stochastic shocks driving the model to be at least equal to the number of observed variables, but certain shocks included into these models were criticized as lacking structural interpretation; assuming that such shocks are highly persistent could be questioned as well.

Even rigidities-augmented DSGE models which are driven by many shocks could remain misspecified, as evidenced, for example, by recent DSGE-VAR analysis of Del Negro and Schorfheide (2006). One hypothesis regarding the source of the residual misspecification has been the fact that the agents' expectations are rational, meaning that their subjective expectations of forward-looking variables are always consistent with the model and coincide with true mathematical expectations for given parameter values and assumed stochastic structure of the exogenous processes (shocks). Testing this hypothesis has been a major motivation behind a recent literature on estimating New Keynesian DSGE models under assumption that the agents, instead of having Rational Expectations (RE), behave as econometricians and constantly re–estimate the relation between forward–looking and other variables of the model, trying to learn the true functional form of the expectation formation mechanism.

An additional source of support for the less-than-rational beliefs hypothesis can be found in estimates of New Keynesian Phillips curves under assumption of sticky information for price settlers. As Reis (2009) summarizes, often the major source of the lack of fit for these models is an assumption that the agents are making decisions based on expectations that are rational even when based on incomplete information. Gomes (2010) attempts to develop a sticky information model where the agents who cannot act on the latest information use adaptive learning (AL) of the form described by Evans and Honkapohja (2001) to form expectations, instead of expectations constructed rationally on basis of outdated information.

In adaptive learning literature, the agents are assumed to possess imperfect knowledge about the reduced form parameters of the model when forming expectations about the future. Agents forecast future values of the lead variables with a linear function of state and exogenous variables. Agents' beliefs about the dynamics of forward looking variables are updated using the constant-gain Recursive Least Square algorithm. Learning represents an alternative source of endogenous inertia; in addition, it affects the transmission mechanism of the model and makes it time-varying through variation in agents' beliefs.

In the sticky prices DSGE literature, several recent studies made efforts to improve the model fit as well as to address the issue of possible model misspecification by departing from the RE hypothesis and incorporating adaptive learning as an expectation formation mechanism. These studies have documented dramatically different conclusions when comparing the fit and estimated parameters (especially structural rigidities) of the models under rational expectations and with learning. The strongest result in favor of integrating the assumption of bounded rationality into the DSGE models was presented by Milani (2007), who considers a very simple three equation New Keynesian model with inflation indexation, Calvo prices, and habit formation, estimated under assumption that the agents are adaptive learners using constant gain Recursive Least Squares (RLS) learning, and compares the results with estimates under RE. The models are estimated using Bayesian methods. Judged by marginal data density, the model with adaptive learning fits the US data significantly better than the RE model. Under adaptive learning certain structural rigidity parameters reduce significantly, conclusion being that the persistence observed in macroeconomic variables such as inflation might be endogenous and caused by agents' learning. Similar results are reported by Milani (2008).

Murray (2007) estimates a simple New Keynesian model augmented with firm-specific capital and constant gain RLS learning using maximum likelihood method rather than Bayesian estimation. He pays a special attention to selecting initial beliefs that the learning agents hold before the estimation period. Adaptive learning models do not fit the US data better than RE models, and Milani results on unimportance of structural rigidities in presense of adaptive learning are not confirmed. On the contrary, some structural rigidities such as capital adjustment costs become significantly more important under learning.

Vilagi (2007) considers several models, one of them very similar to that studied by Milani (2007). He estimates the models using Bayesian methods and the euro-area data. He concludes that model estimated under adaptive learning fits the data significantly better than the RE model, especially if the agents are assumed to form their expectations using only simple univariate AR(1) processes in observed variables. Some structural rigidities become less pronounced under adaptive learning, but in general Vilagi does not confirm results of Milani (2007, 2008). It is unclear whether the differences are caused by the data (US vs. EU) used for the estimation.

Slobodyan and Wouters (2009) estimate the medium size model of Smets and Wouters (2007) under adaptive learning, where the agents are using constant gain RLS. They pay particular attention to the question of forming initial beliefs of the agents, and to the information set available to the agents forecasting future values of the forward-looking variables. Several of the models with learning fit the data equally well or even better than the RE model. Specific initial beliefs contribute significantly to this result: best performing models are the ones where the initial beliefs are optimized to explain the in-sample data, consistent with previous results. Limiting the set of variables used in the forecasting equations can generate models with improved data fit. Learning models are able to generate a rapid and short lived inflation response to productivity shocks, while the response to monetary shocks is slow but very persistent. These results overcome some of the major shortcomings of the model under RE. Having forecasting equations that differ significantly from those implied by the REE is 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, contradicting claims in Milani (2007, 2008).

Slobodyan and Wouters (2010) study what happens if agents' forecasts are based on very small forecasting models, in particular on a model where expected value of a forward-looking variable depends on a constant and two lags of the variable. This forecasting model is similar to the best method of forecasting in Vilagi (2007). In contrast to other AL papers reviewed, the agents estimate simple forecasting models by Kalman filter. The results indicate that a model in which agents use a simple forecasting model to form expectations does fit the data better than the RE model. Relative to the DSGE model under rational expectations, models with learning are estimated to have lower persistence of the exogenous shocks, especially of price and wage markup shocks; structural rigidity parameters decrease insignificantly. The results are robust to the sample period and precise specification of the forecasting model and initial beliefs.

Finally, Jaaskela and McKibbin (2010) estimate a small open economy DSGE model for Australia under constant gain RLS learning. They find that "mechanical" sources of persistence do not become unimportant under AL, and that the data marginally prefer the model with adaptive learning to the RE model.

In this paper, we contribute to the literature on estimation of DSGE models under adaptive learning. We evaluate empirically the relative importance of several types of "frictions" ("mechanical" rigidities) versus learning. Our major contribution is that we provide the answer to these questions by offering a comprehensive analysis of the factors which could determine a diversity of the estimation results under adaptive learning. In such a way we wish to reconcile contradictory conclusions from the previous studies. We perform Bayesian estimation and compare (in terms of the model fit and structural parameters) DSGE models under RE and different AL schemes and study the robustness of the estimation results in several dimensions: by modifying the model size, way of generating initial beliefs, and the set of variables included into the agents' forecasting models. All models are estimated on Euro Area data set described in Fagan et al. (2001) over the period 1970:Q1–2007:Q4 using from 3 to 7 observable macro variables. Previously, only Vilagi (2007) used this data in Bayesian estimation of a DSGE model under adaptive learning. By treating the models in a unified way, we attempt to shed some light on the general outcomes that could be expected from other DSGE model with adaptively formed expectations, and discuss possible sources of discrepant results observed so far in the literature.

The rest of the paper is organized as follows: in Section 2, we discuss the models used. Adaptive learning set-up, ways of forming intial beliefs, and selection of information sets are taken up in Section 3. Section 4 is devoted to the estimation results, and Section 5 concludes.

2 Models

We estimate and study the effects of AL on the three types of models, denoted in the rest of the paper M1–M3.

M1 is a simple 3 equation New Keynesian (NK) model with Calvo prices, price indexation, and habit formation in consumption. This is a model similar to that studied by Milani (2007, 2008), Vilagi's Model C, and one of the models in Murray (2007). The log-linearized version of the model consists of 3 equations.

The first equation of the model is aggregate demand equation, derived from the optimization problem of households:

$$y_t = \frac{hy_{t-1} + E_t\left[y_{t+1}\right] - \frac{1-h}{\sigma}(r_t - E_t[\pi_{t+1}])}{1+h} + e_t^b, \tag{1}$$

where y_t is real GDP, π_t the inflation rate, and r_t the nominal interest rate. The parameter h represents external habit formation, giving raise to the presense of backward-looking component in the Euler equation. σ is the measure of the elasticity of intertemporal substitution. The exogenous disturbance e_t^b is the measure of the preference shock, and follows the first-order autoregressive process,

$$e_t^b = \rho_b e_{t-1}^b + \upsilon_t^b.$$

The presense of nominal rigidities (Calvo pricing and indexation to lagged inflation) imply the following Phillips curve relation:

$$\pi_t = \frac{1}{1+\beta\iota_p} \times \left\{ \begin{array}{c} \beta E_t[\pi_{t+1}] + \iota_p \pi_{t-1} + \left(\frac{(1-\xi_p\beta)(1-\xi_p)}{\xi_p}\right) \times \\ \left[\left(\frac{\alpha}{1-\alpha} + \frac{\sigma}{(1-h)} + \frac{\eta}{1-\alpha}\right) y_t - \frac{h\sigma}{1-h} y_{t-1}\right] \end{array} \right\} + \upsilon_t^p, \tag{2}$$

where ι_p is the degree of inflation indexation, ξ_p the Calvo parameter,η the labor disutility parameter[1], and α is the capital share parameter of the Cobb-Douglas production function.

[1]The utility fucntion takes the form:

$$U_t = e_t^b \left[\frac{(C_t - H_t)^{1-\sigma}}{1-\sigma} - e_t^l \frac{(l_t)^{1+\eta}}{1+\eta}\right],$$

The innovation $\upsilon_t^p \sim N(0, \sigma_p)$ is an *i.i.d.* process.

Finally, the policy of the central bank in setting the nominal interest rate is described by the following rule:

$$r_t = wr_{t-1} + (1-w)(\varsigma_\pi \pi_t + \varsigma_y y_t) + \upsilon_t^r, \tag{3}$$

where $\upsilon_t^r \sim N(0, \sigma_p)$ is the i.i.d. shock.

M2 is a model *M1* augmented with sticky wages, wage indexation, and sticky employment (similar to Vilagi's Model A). Our model differs from Vilagi's by introducing a shock to the labor supply e_t^l, which enables better capturing of the properties of the wage inflation process. Derivations of the model's log-linearized equations can be found in Vilagi (2007). The model consists of the following equations:

The aggregate demand equation is the same as in the model *M1* and is given by (1).

The price inflation equation is now given as

$$\pi_t = \frac{1}{1+\beta\iota_p} \times \left\{ \beta E_t[\pi_{t+1}] + \iota_p \pi_{t-1} + \left(\frac{(1-\xi_p\beta)(1-\xi_p)}{\xi_p} \right) (wr_t + \alpha l_t - e_t^a) \right\} + \upsilon_t^p, \tag{4}$$

where wr_t is real wage rate and e_t^a the first order autoregressive productivity shock,

$$e_t^a = \rho_a e_{t-1}^a + \upsilon_t^a.$$

The wage inflation equation is given by:

$$\begin{aligned} wr_t \;\; = \;\; & \frac{1}{1+\beta} \times (\beta E_t[wr_{t+1}] + wr_{t-1} + \beta E_t[\pi_{t+1}] - (1+\beta\iota_w)\pi_t + \iota_w \pi_{t-1}) + \\ & + \frac{(1-\xi_w\beta)(1-\xi_w)}{(1+\beta)\xi_w(1+\theta_w\phi)} \left(\frac{\sigma}{(1-h)}(y_t - hy_{t-1}) + \eta l_t - wr_t + e_t^l \right) + \upsilon_t^w, \end{aligned} \tag{5}$$

where l_t are labor hours, ξ_w and ι_w are Calvo wage and wage indexation parameters, and θ_w is the elasticity of substitution between different types of labor. e_t^l is a labor supply shock which follows first order autoregressive process:

$$e_t^l = \rho_l e_{t-1}^l + \upsilon_t^l.$$

The innovation $\upsilon_t^w \sim N(0, \sigma_w)$ is and *i.i.d.* process.

The log-linearized firm's technology process takes the form

$$y_t = (1-\alpha) l_t + e_t^a. \tag{6}$$

Policy rule of the central bank is described by equation (3).

Finally, following Vilagi (2007) and Smets and Wouters (2003), sticky emplyment is modeled

where $H_t = hC_{t-1}$ is external habit.

as follows:

$$em_t - em_{t-1} = \beta\left(E_t[em_{t+1}] - em_t\right) + \frac{(1-\xi_l\beta)(1-\xi_l)}{\xi_l}(l_t - em_t) + \upsilon_t^m, \tag{7}$$

where em_t denotes the number of people employed, ξ_l is the Calvo-type employment parameter, and $\upsilon_t^m \sim N(0, \sigma_m)$ is an *i.i.d.* shock. The inclusion of such an auxiliary equation for employment is motivated by the the absense of the consistent euro area data on aggregate labor hours, whereas the employment variable is available. Since the response of employment to macroeconomic shocks is rather persistent, it is assumed that only a constant fraction ξ_l of firms can adjust employment to its desired total labor input.

The model *M3* is a Smets and Wouters (2003) model. In addition to the frictions included into *M2*, this model has investment adjustment costs and variable capital utilization. Detailed description of the model can be found in the original paper.

To summarize, the model *M1* has 3 endogenous variables (y_t, π_t, r_t), 3 exogenous shocks $(\upsilon_t^p, \upsilon_t^r, e_t^b)$, and 3 observables (y_t, π_t, r_t). The model *M2* is described by 6 model variables $(y_t, \pi_t, r_t, wr_t, l_{t,} em_t)$, 7 shocks $(\upsilon_t^p, \upsilon_t^w, \upsilon_t^m, \upsilon_t^r, e_t^l, e_t^b, e_t^a)$, and 5 observables $(y_t, \pi_t, r_t, wr_t, em_t)$. *M3* contains 11 variables, the *M2* set plus real consumption, real investment, capital stock, its rental rate, and price of capital $(c_{t,} inv_{t,} k_t, rk_t, q_t)$, and 10 exogenous shocks, the *M2* set plus investment shock, shock to government spending and the asset price shock $(e_t^{inv}, e_t^G, e_t^Q)$. Model *M3* is estimated using 7 observable variables $(y_t, c_t, inv_{t,} \pi_t, r_t, wr_t, em_t)$.

3 Variations of adaptive learning

3.1 Constant Gain RLS

We implement the adaptive learning within the DYNARE 3.064 MATLAB toolbox which is used to estimate and simulate DSGE models.[2] The models are driven by the exogenous stocastic processes w_t which are either iid random variables or univariate AR(1) processes:

$$w_t = \Gamma w_{t-1} + \Pi \epsilon_t. \tag{8}$$

Up to the first order of approximation, DYNARE represents our models in the following way:

$$A_0 \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + A_1 \begin{bmatrix} y_t \\ w_t \end{bmatrix} + A_2 E_t y_{t+1} + B_0 \epsilon_t = const, \tag{9}$$

where the vector y_t includes endogenous variables of the model. This representation is exact in our case because we work we log-linearized models. Under RE, the DYNARE solution is

$$\begin{bmatrix} y_t \\ w_t \end{bmatrix} = \mu + T \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + R\epsilon_t. \tag{10}$$

[2] This is the same toolbox that was used in Slobodyan and Wouters (2009, 2010).

Deviating from the RE assumption and following Marcet and Sargent (1989) and Evans and Honkapohja (2001), we assume that the agents forecast future values of the forward-looking variables using a linear function of endogenous variables and exogenous driving processes,

$$y_t^f = \phi_{t-1}^T Z_{t-1}, \tag{11}$$

where the exact set of variables included into Z depends on the information set which the agents are assumed to use in forming their forecast. For more details on information sets, see below.[3] The agents' beliefs about reduced form coefficients ϕ are updated using a constant-gain variant of the Recursive Least Squares (RLS). The constant gain algorithm is one of many adaptive methods that allow operating in a non-stationary environment. Besides an advantage of being widely studied in the adaptive learning literature, this method has a natural interpretation as Weighted Least Squares where the weight of a data point depends geometrically on its vintage, with the most recent point getting the highest weight. The agents thus "forget" information from the distant past which might be desirable if the environment, and in particular the dependence of forward-looking variables on elements of Z, is perceived to be time-varying.

Every period, the agents are updating their beliefs in a constant gain RLS step:

$$\phi_t = \phi_{t-1} + gR_t^{-1} Z_{t-1}(y_t^f - \phi_{t-1}^T Z_{t-1})^T, \tag{12a}$$

$$R_t = R_{t-1} + g(Z_{t-1}Z_{t-1}^T - R_{t-1}). \tag{12b}$$

All endogenous model variables have zero means. Therefore, the beliefs should not include a constant. In some specifications of the learning we do not include the constant to be consistent with the theoretical solution. 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 (12).

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

$$\begin{bmatrix} y_t \\ w_t \end{bmatrix} = \mu_t + T_t \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + R_t \epsilon_t. \tag{13}$$

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.

This procedure makes T_t a complicated function of the data, current parameters, and beliefs which could easily become unstable for one or several periods. As in common in the learning literature, we use a projection facility that skips an updating in such cases (see for instance Evans and Honkapohja 2001).[4]

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

[4] Standard projection facility is invoked when beliefs become unstable. Given that not all information sets

3.2 Initial Beliefs

Equations (12) allow us to track the agents beliefs over time, if both the data and the initial beliefs are known. Following Slobodyan and Wouters (2009), we use three ways of selecting initial beliefs. In the first two ways, initial beliefs are consistent with some REE, while the third is based on regression estimates of the pre–sample data.

The first two ways of selecting the initial beliefs use equation (14) below to calculate ϕ_0 and R_0. At any REE, given for example by (10), one could derive a matrix of second moments of the model variables, Ω. These moments imply a relation between the forward-looking variables y^f and the variables used in forecasting Z, $y_t^f = \phi_0^T Z_{t-1}$, given by a projection of y_t^f onto Z_{t-1}. Initial condition for the second moments matrix R used in equations (12) is taken directly from the corresponding rows and columns of Ω. The formulae for ϕ_0 and R_0 are given by

$$\phi_0 = E\left[Z_{t-1}Z_{t-1}^T\right]^{-1} \cdot E\left[Z_{t-1}\left(y_t^f\right)^T\right], \tag{14a}$$

$$R_0 = E\left[Z_{t-1}Z_{t-1}^T\right], \tag{14b}$$

where the expectations $E[]$ are derived using Ω.

Denote the parameter vector that is used to derive the model equations θ. Denote $\widetilde{\theta}$ a vector of parameters for auxiliary model which generates matrix $\Omega\left(\widetilde{\theta}\right)$ that is then used for calculations in (14). Then, in the first way of deriving initial beliefs, denoted W1, $\theta = \widetilde{\theta}$ at all times. In other words, initial beliefs are consistent with the REE produced by the estimated parameter vector θ. W1 is the closest to the pure rational expectations as only in-sample data variations could break the mapping of REE-implied relations between forward-looking variables y_t^f and predictors Z_{t-1} into the agents' perceptions of these relations; in the beginning of the sample, the two are the same. The way W1 is equivalent to the Case 2 in Murray (2007).

In the second way, W2, $\widetilde{\theta}$ is fixed while θ changes in the posterior optimization or MCMC steps. In principle, $\widetilde{\theta}$ could be selected to be any parameter vector. In this paper, we take several (usually 10 to 20) draws of $\widetilde{\theta}$ from the posterior distribution of parameters, approximated by the multivariate normal distribution, obtained after posterior maximization step under adaptive learning with W1 beliefs. W2 allows for more flexibility than W1, as the initial beliefs could now vary independently from the model itself. On the other hand, thus constructed sets of initial beliefs are consistent with RE equilibria that are rather similar to each other, because the parameter draws $\widetilde{\theta}$ are drawn from the same distribution; therefore, we consider W2 beliefs as a relatively minor disturbance that allows us to check the sensitivity of estimation results to the initial beliefs. In order to take the results of estimation under adaptive learning seriously, the estimation should pass some minimum set of requirements, such as being robust to W2 beliefs. The way W2 is close to, but not equivalent, to the Case 3 of Murray (2007).

Our third initialisation approach, W3, uses regression-based initial beliefs, obtained by running a regression of y_t^f on Z_{t-1} using pre-sample data. We pick the point estimate rather

---

lead to beliefs that could be described by some VAR, we have to resort to imposing projection facility when the transition matrix T_t loses stability.

than a random point from the distribution of regression estimates, the latter being proposed by Giannitsarou and Carceles-Poveda (2007). This way represents a more serious robustness check for AL estimation that W2 for two reasons. First, correlation structure of the variables could change significantly between pre- and in-sample data, in which case pre-sample regression-based initial beliefs are of not much help to the agents in navigating in-sample environment. Second, the model could be so misspecified that even W1 beliefs consistent with REE of the pre-sample estimated model are still very far from those which could have been obtained by any regression. In both cases, our W3 beliefs are likely to be significantly different from the W1 ones, thus allowing us to observe the effect of initial beliefs on the estimation results.

3.3 Information Sets

Most of the theoretical results in the AL literature have been obtained for the case of Minimum State Variable (MSV) learning, where the agents form their expectations using a linear function of endogenous variables and stochastic shocks that is equivalent to the function one would derive as the REE solution of the model. In particular, the set of variables that is assumed to be available to the agents coincides with the variables that determine rational expectations of forward-looking variables. Thus, in MSV learning only the coefficients of the expectation-forming function could differ from their REE counterparts. MSV learning is one of the information sets that we use in this paper, denoted by I1. As is standard in the learning literature, we assume that the agents know exactly the law of motion (8) of the exogenous driving processes.

Assuming that the agents have access to the values of exogenous shocks is theoretically appealing as it enables convergence to the REE as an outcome of certain learning algorithms, namely, RLS with decreasing gain (a recursive analog of standard OLS regression) when the REE is E-stable, see Evans and Honkapohja (2001) for details. However, this assumption could be criticized as unrealistic. Therefore, we employ a second information set, I2, which assumes that the agents use the same endogenous model variables as the ones present in the REE solution, but not the exogenous stochastic processes.

Several papers in the small but growing literature on estimation of DSGE models under AL used an extreme informational assumption, making forecasts of macroeconomic variables depend only on own lag(s) of the variable itself and possibly a constant. Thus, the forecasting equations become a set of univariate AR(1) or AR(2) processes. Despite the fact that this approach denies the agents access to a significant amount of information available in the model, it was shown to lead to a very good model fit, see Vilagi (2007) and Slobodyan and Wouters (2010), among others. For this reason, we include an information set assumption I3 into our study, where all the forward-looking variables are believed by the agents to be simple univariate AR(1) processes.

In contrast to low–dimensional models studied by Milani (2007), Sargent, Williams, and Zha (2006), or Vilagi (2007), combination of some of the information sets and some models leads to necessity of accessing values of endogenous variables that are not observed. In such cases we use output from the Kalman filter to construct the likelihood function for a particular combination of parameters on both sides of the updating equation (12).[5]

4 Estimation Results

4.1 Data and Priors

For our estimations, we use the data set constructed in Fagan et al (2001) over the period 1970:Q1–2007:Q4. The set of observables (varies from 3 to 7 variables depending on the model) includes the key macro-economic variables of the euro area. When constructing the observables, the following time series were used: real GDP (YER), GDP deflator (YED), compensation to employees (WIN), number of employees (LNN), short-term nominal interest rate (STN), real consumption expenditures (PCR), real investment (ITR). The time series of real wages is constructed as WR=(WIN/LNN)/YED. The STN time series was divided by 4 to obtain quarterly data. The natural logarithms of all variables except the STN were taken. The inflation rate is given by $\ln(YED_t) - \ln(YED_{t-1})$. Real variables are linearly detrended using a separate trend for each variable, estimated by OLS; inflation and the nominal interest rate are detrended by the same linear trend in inflation as in Smets and Wouters (2003).

We estimate all the models using Bayesian methods. The table of priors is presented in the Appendix A. We mostly followed the priors chosen in the original papers: Smets and Wouters (2003) for *M3* and Vilagi (2007) for *M1*. For the model *M2* are undertook a combined approach: prior distributions for shocks and some of the structural parameters (like habit and Calvo prices) are based on Smets and Wouters (2003). At the same time we wanted to keep priors on some of the nominal rigidities (which have most controvercial empirical support , like price and wage indexation) as loose as possible. In this approach, we followed Vilagi who assumed uniform prior distributions for indexation. For the same reasons, we choose uniform prior on investment adjustment cost in model *M3*, and thus departed slightly from Smets and Wouters (2003). Overall, since the major task of this paper was to investigate the impact of AL on structural parameters, in particular nominal and real rigidities, we tried to avoid restricting such parameters by strict priors, providing instead for maximum flexibility while attempting to obtain unimodal posterior distributions under both RE and AL. Some of the rigidities such as habit formation appeared to be rather robust to the change of priors.

A number of parameters were calibrated. Similarly to Smets and Wouters (2003), we fix discount factor β at 0.99, capital share $\alpha = 0.30$, the depreciation rate $\tau = 0.025$; the share of steady state consumption in total output and the stady state investment share are assumed to be equal to 0.6 and 0.22 respectively. The labor dsutility parameter η is assumed to be fixed in model *M1* and equals 2.5.

Typically, 200,000 to 500,000 MCMC draws were performed, using two (in some cases three) MCMC chains. For more details on Bayesian estimation of DGSE models, see An and

[5] We use only filtered estimates of endogenous variables, both of right- and left-hand sides of the forecasting equations.

Schorfheide (2007).

4.2 Model Fit

The fit of a model estimated using Bayesian methods can be ascertained using marginal data density, defined as

$$p\left(Y|\mathcal{M}\right) = \int \mathcal{L}\left(\theta|Y\right) p(\theta) d\theta,$$

where $\mathcal{L}\left(\theta|Y\right)$ is the likelihood function of the data Y given parameters of the model θ, and $p(\theta)$ is the prior density. This measure allows a straightforward comparison of two models, say $\mathcal{M}_1$ and $\mathcal{M}_2$ that are estimated on the same data. Posterior odds ratio, a measure of how much more likely a model $\mathcal{M}_1$ is when compared to the model $\mathcal{M}_2$, is given by

$$\frac{\pi\left(\mathcal{M}_1\right)}{\pi\left(\mathcal{M}_2\right)} \cdot \frac{p\left(Y|\mathcal{M}_1\right)}{p\left(Y|\mathcal{M}_2\right)},$$

where $\pi\left(\mathcal{M}_i\right)$ represents prior probability of a model $\mathcal{M}_i$. The first term in the above expression is known as prior odds, and the second as Bayes factor. Usually, the prior probabilities are taken to be equal, and thus a posterior odds ratio equals the corresponding Bayes factor. For more details on model comparison, consult An and Schorfheide (2007).

Logarithms of marginal data densities from the estimations of our models are presented in the Table 1. Initial beliefs are constucted using the RE-consistent method of way W1. Our major result is that the RE hypothesis is indeed restrictive. Relaxing the rationality assumption through introduction of adaptive learning improves the marginal data density of the model for essentially all learning specifications: the only case where RE and AL models have similar fit is model *M2*, information set I1 (MSV learning) with a constant. In all other cases the AL model fit is significantly better than its counterpart under rational expectations. It is hard to compare the Bayes factors across models that have different number of observable variables — three in *M1*, five in *M2*, and seven in *M3*. If the hypothesis of the rational expectations as the main source of misspecification of a DSGE model is correct, then the adaptive learning could correct some of it. One could presume that the resulting improvement in marginal data density of the model under AL relative to the RE model then reflects the degree of mis-specification which could be different in models *M1–M3*. Testing this conjecture is beyond the scope of the current paper.

Another result is that the most restrictive information set I3 is indeed the best for all three models. This result has been observed previously by Vilagi (2007). Slobodyan and Wouters (2010) also suggest that endowing the agents with a minimal set of variables used in forecasting may work well in practice. The evidence on the other two information sets is more mixed: I1, the largest set which is consistent with the rational expectations MSV solution, is marginally better than the resticted MSV set I2 for the smallest model *M1* but is significantly worse in the larger models *M2* and *M3*. Overall, though, we can observe a clear ranking I3–I2–I1, making a very strong case for the statement that the more restrictive is the information set available to

Table 1: Model Comparison in Terms of MArginal Likelihood

Model specification	*M1*	*M2*	*M3*
REE	-134.96	-182.83	-468.83
AL without constant:			
I3: univariate AR(1)	-125.61	-137.66	-421.65
I2: endogenous states	-130.39	-147.71	-436.76
I1: endogenous states and shocks	-129.36	-174.27	-449.11
AL with constant:			
I3: univariate AR(1)	-119.36	-129.2	-419.6
I2: endogenous states	-131.45	-153.22	-442.19
I1: endogenous states and shocks	-123.5	-182.7	-461.68

Log marginal data densities for the three models using different information set assumptions and REE-consistent initial beliefs W1. Bayes factor — a relative probability of one model over another, equals exp of the difference between the corresponding log densities.

the agents for forecasting forward-looking variables, the better is the model fit.

Comparing the AL estimations with and without the constant, we observe a clear separation between the best set of variables I3 and the worse group of I2 and I1. For the very economical forecasting equations implied by assuming I3 (just own lag in every forecasting equation), including the constant improves the model's marginal data density significantly, especially for the smaller models *M1* and *M2*. For sets I1 and I2, including the constant worsens the marginal data density, sometimes by so much that the overall model fit is essentially the same as under the RE (information set I1 with constant, model *M2*). The only exception to this rule is model *M1*, set I1. We believe that large sets of regressors I1 and I2, when used in forecasting equations on the real data over the estimation period, might lead to overfitting. In this case, adding an extra variable — a constant — makes the overfitting problem worse. In the model *M1*, the overfitting problem is not as severe because the total number of right-hand side variables is small (three endogenous variables and one shock). Notice that for the intermediate information set I2, worsening of the fit after including the constant is minimal in model *M1* and moderate in model *M2*, which is consistent with the overfitting of the forecasting equations explanation.

Finally, we analyze the relative fit of alternative model specifications as a function of time. Specifically, we would like to find out whether the gain in the model fit observed under the information set I3 comes from a specific (short-lasted) time period, or whether the superior performance of the model based on univariate forecasting rule holds for the longer time span. Figure 1 shows the cumulative likelihood for model specifications I2 (dashed line) and I3 (dotted line) relative to I1. If a line is trending up in this graph on some time interval, this means that on average the likelihood on this interval is better relative to the I1 model. Figure 1 indicates that I3 model does better than I1 almost all the time, except for 1985-1988 and 1990-1993. Before 1980 and after 1993 we observe a persistent positive trend in the I3 relative cumulative likelihood. This means that for most of the sample the model with I3 set is more appropriate for describing the data generating process than the specification impled by the information set I1; the better model fit is broadly based rather than being due to a particularly favorable performance at a specific time. On another hand, the gain in the model fit under I2 relative to I1 is relatively modest, especially in the second half of the sample. In the first half of 1970es the model with I2 information ser performs worse than the one with the full set of variables and shocks I1.

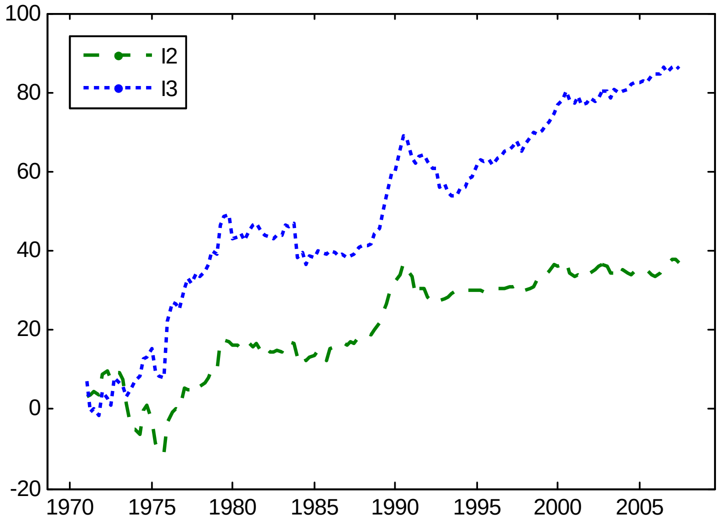

Figure 1: Cumulative likelihood for estimated model specifications I2 and I3 relative to I1.

4.3 Estimated Parameters

Given a large number of treatments in the paper, we compare the effect of AL assumption for the estimates of structural rigidity parameters and persistence of exogenous shocks only for the best information set identified in the previous section, namely I3 with a constant. We will treat this specification as a baseline and consider the outcomes with other sets of variables as a form of sensitivity analysis.

Table 2 presents an overview of the main results of our estimation. As is obvious from the Table, under AL some estimated structural rigidities and persistence of the shocks fall, sometimes significantly. Among the parameters present in all three models, habit persistence parameter h presents the clearest picture: its estimate is lower under AL than under RE. The drop is quite significant: in *M1* and *M2*, posterior mean under adaptive learning lies outside of the 95% Highest Probability Density (HPD) interval of RE estimation, for *M3* it is less expressed because the HPD under RE is very wide. Estimated Calvo pricing and inflation indexation parameters fall in *M1* but stay unchanged or even increase marginally in larger models *M2* and *M3*. Among parameters present in *M2* and *M3* only, Calvo wages parameter falls marginally, wage indexation remains at essentially zero level as under RE, and Calvo parameter for employment falls. Finally, the largest (*M3*) model-specific parameters — elasticities of investment adjustment costs and of capital utilization — both fall, with adjustment costs exhibiting the most dramatic decline among all the parameters studied (from 9.44 under RE to 3.21 under AL).

These results taken in the whole signal that there is indeed an overall drop in structural rigidity parameters when the RE assumption is replaced by the AL one. The parameters that are estimated to be somewhat extreme under the RE fall the most (habit persistence in consumption h, investment adjustment costs φ, Calvo prices ξ_p, Calvo employment ξ_e), with Calvo wage parameter ξ_w being somewhat exceptional. An increase in importance of "mechanical" sources of rigidities is very seldom observed — basically, only price indexation parameter in the model *M2* inches up marginally under AL from a very low level of 0.20 estimated under RE.

We also observe that the overall decrease in importance of "mechanical" frictions is most pronounced in a small model *M1* where there are few rigidities. This outcome is consistent with the view that adaptive learning to a significant degree serves as a tool of remedying misspecifications. For example, both Calvo pricing and price indexation are estimated to be extremely high under the RE in model *M1* ($\xi_p = 0.97$, $\iota_p = 0.71$). These paramaters drop significantly under the AL. Larger model *M2* adds wage rigidities which probably relax the misspecification present in *M1*. As a result, there is not much movement in this group of parameters (Calvo prices and wages, indexation of prices and wages) between RE and AL estimations in *M2*. This result indicates that the ability of adaptive learning to substitute for real and nominal rigidities can be overestimated if one uses very small model. Therefore, Milani's conclusions who obtained his results in the estimated three equation NK model cannot be expected to apply in more complicated models to the same extent.[6] We conclude that learning can substitute for "mechanical" source of rigidities only partially; some of the structural frictions remain quite strong.

Turning attention to the persistence of exogenous processes, we see that there is no clear pattern: productivity shock can become less persistent while employment equation shock persistence goes up. Demand shock becomes less persistent and become more precisely estimated in *M1* and *M3* but remains well within Rational Expectations HDP for this parameter that is very wide in both cases.

Comparing out results to others in the literature, we do not observe significant decline of consumption habits and price indexation parameters to zero in as in Milani (2007) in a simple model *M1*, while in more complex models *M2* and *M3* price indexation is low already under the RE. Murray (2007), comparing Cases 1 and 3 (Case 3 is similar to the information set I2, not presented in the Table 2), observes an essentially unchanged habit formation parameter (0.11 under RE to 0.12), and a dramatic *increase* in capital adjustment friction, a parameter playing a role that is similar to our φ, which is found to drop significantly under the AL in *M3*, the only model where this friction is present. Murray also finds a significant decrease in price indexation from 0.36 to 0.0, which is somewhat consistent with our estimates. Ja̋áskelá and

[6]We have to note that similarly to Murray (2007) and Vilagi (2007), we do not confirm Milani results who found some of the structural rigidity parameters to become insignificant undel learning. Whether this discrepancy is due to the estimation method used (Murray (2007) used Maximum Likelihood estimation) or the data (European in Vilagi (2007) and here vs. US in Milani 2007, 2008) is a subject of further study.

McKibbin (2010) estimate an open economy model of Australia with initial beliefs constructed by a method rather close to our W1 and find that both habit persistence and share of rule-of-thumb consumers (this parameter plays a role somewhat akin to the price indexation parameter) *increase* under AL. Slobodyan and Wouters (2008, 2009) report that in a large model (very similar to *M3*) investment adjustment cost elasticity φ typically drops significantly while habit persistence can grow (in Slobodyan and Wouters 2009; note, however, that they report habits increase from 0.70 under RE to 0.74÷0.79 under AL, while here habits *drop* to 0.73, *i.e.*, to approximately the same value) or decrease slightly (in Slobodyan and Wouters 2008; from 0.77 to 0.68). Finally, Vilagi (2007) finds that habit persistence could either decrease from 0.99 to 0.79 (in a small model equivalent to *M1*) or stay unchanged at 0.99 in a model similar to *M2*. Price indexation drops significantly in both a small model similar to *M1* (from 0.63 to 0.15) and a larger *M2*-like model (from 0.44 to 0.29). Wage indexation also reduces from 0.75 to 0.02. Insignificance of indexation parameters under AL is confirmed by us. Vilagi also finds an unambiguous although small in magnitude drops in Calvo parameters in price- (both *M1*-like and *M2*-like models) and wage-setting (*M2*-like model), which are confirmed by us only for the *M1* model case.

Thus, putting aside results of Milani, one consistent finding of the literature is that capital adjustment cost parameter falls under AL. Another, more tentative, conclusion is that adaptive learning estimation leads to habit persistence parameter's increase if under RE it is less than 0.75, but it decreases if already high under RE as in this paper or Slobodyan and Wouters (2008). It is difficult to establish a common pattern among the parameters describing nominal rigidities because of differences in this modelling block and estimated parameters; overall, there seem to be no uniform movement in nominal rigidities in one or another direction under AL in the papers surveyed, with Vilagi (2007) being an important exception. Our results partially confirm Vilagi (2007) but for the fact that indexation parameters are already extremely low in larger models.

Table 2: Model Comparison in Terms of Estimated Parameters

	h	φ	ι_p	ι_w	ξ_p	ξ_w	ξ_e	ψ	ρ_a	ρ_b	ρ_g	ρ_l	ρ_i	g
M1:RE	0.89		0.71		0.97					0.58				
AL	0.79		0.61		0.91					0.50				0.101
M2:RE	0.90		0.20	0.05	0.91	0.80	0.86		0.96	0.50		0.92		
AL	0.78		0.24	0.04	0.91	0.77	0.75		0.93	0.51		0.97		0.063
M3:RE	0.88	9.44	0.23	0.05	0.91	0.79	0.75	0.21	0.99	0.42	0.93	0.89	0.93	
AL	0.73	3.21	0.24	0.04	0.90	0.73	0.58	0.14	0.97	0.34	0.94	0.97	0.89	0.046

In every cell, the top number is the posterior mean of the estimated parameter distribution under RE, and the bottom one is the result of estimation under the baseline AL specification (only own lag and a constant used to forecast every forward-looking variable). Only structural rigidity parameters and persistence of exogenous shocks are presented.

4.4 Information Sets

As stated previously, information set which the agents use in forming their expectations affects the model fit to a large degree. To access whether a similar effect is observed with respect to the estimated parameters, in the Table 3 below we present the estimates for the middle-of-the-road model *M2* under adaptive learning for all information sets. The most striking result is that parameter values are affected to a significant degree by the information set. For example, some structural rigidity parameters actually increase, rather than decrease, under AL estimation when the agents are allowed access to the MSV solution consistent information set I1. Calvo wages and employment and price indexation are the lowest in the baseline I3 estimation and the highest in MSV I1 estimation, with I2 set being right in the middle. With I1, these parameters are actually higher than under RE. Overall, the REE-consistent set I1 delivers the worst outcome in terms of "mechanical" sources of rigidities, as all parameters either stay the same as under RE or slightly increase, with ρ_l, persistence of the employment shock, being the only exception.

Another interesting feature to note is that information sets with and without the constant deliver the most consistent results for the smallest set I3. The only exception is the estimated gain parameter which is rather high in the baseline estimation with a constant (0.063) but is much lower without it (0.024). As described in the following subsection, high gain implies beliefs that change rather fast in reaction to the data. On the other hand, all other estimations deliver much lower value of the gain in the region of $0.02 \div 0.03$ (and even 0.007 for MSV solution consistent set I1 with constant). If the beliefs are close to being unstable at some point during the estimated sample, the estimation procedure could select a lower gain in which case probability of invoking the projection facility declines.[7] We discuss this issue in more detail in the next subsection where we describe the beliefs about inflation process implied by different information sets.

Considering the whole set of estimations for different information sets, we can say that in accordance with the model fit results, the set of variables that the agents are assumed to be using for forming expectations influences the estimation significantly. Therefore, any comparison of the results of estimation under adaptive learning across different models should take into account the variables used by the agents; in case the assumed information sets differ significantly, the results could not be compared.

4.5 Initial Beliefs and Sensitivity

As mentioned previously, different ways of initializing initial beliefs could be considered as a sensitivity analysis for the AL estimation. Estimation under adaptive learning naturally introduces some free parameters. In W1 way of initializing the learning — REE-consistent initial beliefs — there is only one extra parameter, the gain. However, if the estimation depends too sensitively on the initial beliefs, one could criticize AL estimation with W1 initial beliefs

[7] Estimations with frequent projection facilities tend to fit the data poorly.

Table 3: Information Sets Comparison in Terms of Estimated Parameters

Model M2	h	ι_p	ι_w	ξ_p	ξ_w	ξ_e	ρ_a	ρ_b	ρ_l	g
Rational Expectations	0.90	0.20	0.05	0.91	0.80	0.86	0.96	0.50	0.92	
Adaptive Learning, no constant										
I3: univariate AR(1)	0.80	0.20	0.05	0.90	0.75	0.74	0.94	0.46	0.95	0.024
I2: endogenous states	0.79	0.41	0.18	0.90	0.82	0.78	0.92	0.50	0.92	0.022
I1: endogenous states and shocks	0.92	0.68	0.08	0.89	0.91	0.88	0.97	0.56	0.51	0.028
Adaptive Learning, with constant										
I3: univariate AR(1)	0.78	0.24	0.04	0.91	0.77	0.75	0.93	0.51	0.97	0.063
I2: endogenous states	0.78	0.51	0.14	0.90	0.81	0.79	0.91	0.53	0.92	0.022
I1: endogenous states and shocks	0.88	0.58	0.06	0.90	0.86	0.96	0.96	0.58	0.68	0.007

as using in-sample data for the belief optimization.[8] If, on the other hand, we could establish that the estimation results remain relatively stable when initial beliefs are selected from a distribution centered on the W1 beliefs, we could claim that the estimation is robust to small errors that the agents could make in forming the beliefs. In the latter case, the estimation results are a function of the time variability in expectation-forming function and of changes in the transmission mechanism that are associated with a particular information set, not the initial beliefs that could be somewhat arbitrary.

To generate W2 initial beliefs, we take the multivariate normal distribution of parameters implied by the posterior maximization of a model under adaptive learning with W1 initial beliefs, and select 10÷20 draws from this distribution.[9] For every parameter draw, the corresponding REE is then constructed and initial beliefs are derived that are consistent with the REE in accordance with (14). These initial beliefs are then fixed for the duration of estimation using MCMC. As we argued before, W2 beliefs allow for some flexibility by disentangling the initial beliefs' REE (and, thus, transmission mechanism) and the REE implied by the current parameter draw in the MCMC. On the other hand, W2 beliefs stay constant during the MCMC, which represents a constraint on the estimation. In case the estimation results are very sensitive to the initial beliefs, one would expect either a significant divergence of estimation results in terms of the model fit or estimated parameters, or at least an increase in confidence intervals of the estimated posterior distribution of parameters. Neither of these effects is expected when the estimation is insensitive to the initialization.

Another consequence of W2 initial beliefs is as follows. Consider for a moment posterior

[8] Several papers using estimation under AL have used optimized initial beliefs, cf. some specifications in Milani (2007) and Sargent, Williams, and Zha (2005), effectively using the in-sample data twice. This procedure is hard to justify if one takes the story of agents as econometricians seriously.

On the other hand, REE-consistent W1 initial beliefs in this paper correspond to the agents who know the probability distribution that would be obtained under particular parameter values and use is as a starting point, but allow for non-stationarity and/or structural breaks when forming expectations in real time.

[9] One could also use the MCMC output directly and randomly select points from there, thus drawing from true posterior distribution of parameters, not its multivariate normal approximation that could be significantly incorrect. Beside simplicity, our procedure has an advantage of allowing to draw from scaled up or down distribution easily. We had to restrict our distribution to 50% of the approximate multivariate normal in order to guarantee that most draws for initial beliefs result in a point that satisfies inequality restrictions imposed on parameters during the estimation.

probability maximization step under W1 initial beliefs. In the mode, the REE that is used to construct the initial beliefs (defined by the parameter vector $\widetilde{\theta}$) and the REE implied by the model parameter vector θ are constrained to be equal, $\theta = \widetilde{\theta}$. Denote posterior mode as $\theta_0 = \widetilde{\theta}_0$. Now, fix the initial beliefs' parameter vector at $\widetilde{\theta}_0$ and re-optimize the posterior with respect to θ. It is clear that at the resulting posterior mode, θ_1, we should have a higher value of the posterior, because it is always possible to set $\theta_1 = \widetilde{\theta}_0$ and get the W1 posterior value.[10] Normally, higher posterior is translated into better marginal data density, and thus we expect that W2 estimation with the initial beliefs' parameter vector $\widetilde{\theta}$ fixed at W1 posterior mode value will fit the data better (will have higher marginal data density). If, on the other hand, W2 beliefs parameters $\widetilde{\theta}$ are fixed at values that are close to, but not equivalent, to the W1 parameter vector θ, this "partial optimization" effect is counterbalanced by the fact that the initial point in this procedure is likely to have lower posterior that the W1 estimation. If the W2 initial beliefs are close to W1 ones, we would expect the "partial optimization" effect to be stronger, and thus the model fit to be improved.

We conduct analysis with W2 beliefs for the baseline information set I3 and all models. The results across models are qualitatatively similar, and thus we concentrate on reporting middle-of-the-road model *M2* estimations only.

Consider first the model fit. Marginal data density for the baseline AL estimation for *M2* equals -129.2 (see Table 1). In 17 W2 estimations for the same model with baseline information set, the marginal data density varies between -129.5 and -127.3, with the mean of -128.0. In other words, the changes are tiny, but on average we observe a slight improvement, consistent with the "partial optimization" logic given above.

Now, let turn attention to the estimated parameters. We select three W2 estimations, those with the highest, the lowest, and the median marginal data density among the 17 W2 runs that we have conducted. Table 4 presents, for the parameters that reflect nominal and real rigidities, the posterior mean and the confidence interval for the selected estimation runs. Clearly, there is little or no difference between the three estimations. Only for the gain parameter we do observe somewhat larger changes, but they are well within the estimated HPD intervals. Comparing the parameter estimates with the corresponding row in the Table 3, we see that the parameter estimates under W2 are extremely to those under W1 but sometimes slightly biased, which is consistent with the marginal likelihood being higher on average with W2 initial beliefs than W1 beliefs.

A very similar picture can be observed for other models under baseline AL estimation. Thus, we can conclude this part of the sensitivity analysis by stating that at least for the best information set (only own lag and a constant are used to form expectations), estimation under adaptive learning with REE-consistent initial beliefs is extremely robust to small disturbances in the beliefs.

We next turn to another robustness exercise, where initial beliefs are derived from OLS

[10] This estimation could then be considered as a first step in the joint optimization of model parameters θ and belief parameters $\widetilde{\theta}$, the next step being fixing the model paramters at θ_1 and optimizing with respect to the belief parameters $\widetilde{\theta}$ to get $\widetilde{\theta}_1$, etc.

Table 4: Sensitivity of Estimated Parameters to Initial Beliefs

Marg.lik	h	ι_p	ι_w	ξ_p	ξ_w	ξ_e	ρ_a	ρ_b	ρ_l	g
Lowest	0.79	0.25	0.04	0.91	0.77	0.75	0.93	0.50	0.96	0.058
	0.68 0.91	0.14 0.34	0.0 0.09	0.89 0.93	0.72 0.82	0.72 0.78	0.88 0.97	0.39 0.62	0.92 0.99	0.04 0.07
Median	0.79	0.25	0.04	0.91	0.77	0.75	0.93	0.51	0.95	0.066
	0.68 0.91	0.15 0.36	0.0 0.09	0.89 0.93	0.71 0.82	0.72 0.78	0.88 0.97	0.36 0.62	0.92 0.99	0.05 0.08
Largest	0.78	0.25	0.04	0.91	0.76	0.75	0.93	0.51	0.95	0.064
	0.67 0.90	0.15 0.36	0.0 0.09	0.88 0.93	0.71 0.82	0.72 0.78	0.88 0.98	0.40 0.62	0.93 0.99	0.05 0.08

Model *M2* estimated under the baseline AL specification (only own lag and a constant used to forecast every forward-looking variable), W2 beliefs. Only 3 (out of 17) estimations are presented, with lowest, median, and highest marginal likelihood. In every cell, the top number is the posterior mean of the estimated parameter distribution under AL, and the bottom one is the corresponding confidence interval. Only structural rigidity parameters and persistence of exogenous shocks are presented.

regression using pre-sample data — W3 beliefs. We use 20 quarters of the data to form initial beliefs, and estimate the model using the rest of the sample. Again, we present the results only for the Model *M2*, baseline information set I3 with a constant. Given that the estimated sample is now different, we re-estimate the model using this shorter sample under W1 beliefs as well.

Regarding the model fit, W3 beliefs are doing worse than our baseline adaptive learning specification: marginal likelihood is just -54.8 *vs.* -43.2 for W1 beliefs. Both specifications, though, fit the data significantly better than Rational Expectations estimation at -96.0. The reason is probably that pre-sample based initial beliefs generate forecasting functions that are largely inconsistent with the in-sample data. We discuss the issue in more detail in the following subsection.

Comparison of the estimated parameters shows that estimated parameters differ surprisingly little between the two AL specifications: the only parameter that differs noticeably is price indexation, which increases from 0.25 in the baseline specification to 0.46 under pre-sample beliefs. The gain also differs. As the gain parameter is related to the speed with which the beliefs are updated, this implies that the biggest difference between the beliefs could be their volatility and the speed with which expectation formation function adjusts to the new data. We return to this question again in the next subsection. Finally, for this shorter sample the difference between estimations under RE and baseline AL with W1 beliefs is similar to that presented in the Table 2.

4.6 Beliefs About Inflation and Transmission Mechanism

Adaptive learning could affect model fit in several ways. First, time variation of beliefs allows the model itself to become time varying, cf. (13). This could improve model fit if the process that generates time series of observed variables is itself time-varying. On the other hand, if the beliefs updating process is too volatile, parameter uncertainty could lead to deterioration of the fit. Another channel through which adaptive learning operates is through change in the

Table 5: Initial Beliefs Comparison in Terms of Estimated Parameters

	h	ι_p	ι_w	ξ_p	ξ_w	ξ_e	ρ_a	ρ_b	ρ_l	g
RE	0.87	0.25	0.11	0.88	0.77	0.83	0.934	0.41	0.92	
W1 beliefs	0.81	0.25	0.11	0.83	0.70	0.70	0.96	0.47	0.91	0.058
W3 beliefs	0.79	0.46	0.06	0.81	0.72	0.71	0.95	0.45	0.89	0.025

Posterior means of the estimated parameters under RE, AL with REE-consistent initial beliefs, and AL with pre-sample based beliefs. Baseline AL specification (own lag and a constant), model *M2*. Only structural rigidity parameters and persistence of exogenous shocks are presented.

transmission mechanism. Even when the beliefs are consistent with a REE and are not time varying, if the information set used by the agents to form expectations differs from the MSV set then the transmission mechanism differs from that under the Rational Expectations. The fact that information set affects estimations to a larger degree than initial beliefs (compare the results presented in Tables 1 and 3 with those of Tables 4-5) informs us that the transmission mechanism effect could be more pronounced than that of time variation.

To illustrate the effect of the transmission mechanism, observe Fig. 2 which shows the coefficients in the agents' forecating function for inflation (also called the Perceived Law of Motion, or PLM),

$$\pi_t = a + \rho \pi_{t-1}.$$

We present values of a and ρ for W1 and W3 beliefs, I3 information set, model *M2*, *i.e.*, the same adaptive learning specifications that were described in the Table 5. As beliefs evolve over time, a and ρ are time varying. One difference between the initial beliefs is immediate: pre-sample regression based initial beliefs W3 (blue lines) are much more volatile, despite the fact that estimated gain g for W3 beliefs is much lower than under W1 beliefs (0.025 *vs.* 0.058). According to the updating equations (12), innovations to beliefs are given as $gR_t^{-1} \cdot Z_{t-1}\epsilon_t^T$ where ϵ_t is the forecasting error at time t. Amount of update is then proportional to the effective gain gR_t^{-1}. Under pre-sample beliefs, effective gain is much larger despite the gain g being lower, which implies that R_t, the second moments matrix (essentially, variance of inflation), is much smaller in the pre-sample than the value implied by the REE with which W1 beliefs are consistent. Despite the fact that W3 beliefs seem to be more 'correct' (perceived inflation persistence ρ, blue dotted line, is essentially constant over the estimated time interval, while a systematic decline is observed for W1 beliefs), their volatility leads to a significantly worse model fit.

'Initial beliefs' in the Fig. 2 are given as the very first point of the graph. In the Rational Expectations Equilibrium implied by the model parameters in W1 estimation, the agents believe in a very persistent inflation. This feature is shared by the models *M2* and *M3* but, significantly, not *M1* (initial inflation persistence in model *M1* is closer to 0.5). Notice that if the initial persistence is close to unity, the effective gain parameter is restricted to low numbers; otherwise, large forecasting errors could result in update of persistence to values above one, which invokes projection facility. Estimation with large numbers of projection facilities tend to result in a

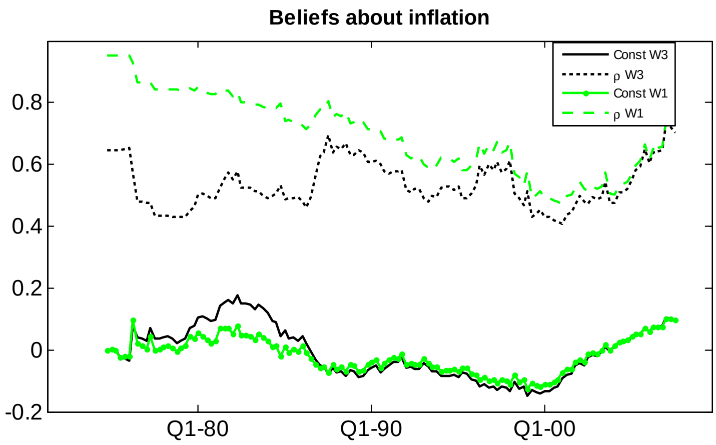

Figure 2: PLM beliefs about inflation (constant and persistence) under REE-consistent initial beliefs (W1) and pre-sample based initial beliefs (W3). The agents perceive the following inflation process:

$$\pi_t = a + \rho\pi_{t-1}.$$

very bad model fit.

Thus, there is a basic tension in initial beliefs that are REE-consistent. The beliefs contain point estimates of the parameters in the forecasting functions (ϕ in Eq. 12), and perceived volatility of variables used to forecast forward-looking variables. If the perceived volatility is inconsistent with the data-generating process of observed variables, the estimation procedure whould attempt to adjust the gain parameter to counteract effect of 'wrong' R_0. However, the gain parameter that is too large could then lead to frequent projection facility hits and very volatile beliefs ϕ, with deteriorating model fit as a consequence. A way of overcoming this problem might be to introduce two gain parameters, one for updating point forecasting function parameters ϕ and another for updating the second moments matrix R, or adding a scale parameter for the matrix R that could be either estimated or calibrated.

To illustrate the effect of the information set used for forecasting on the transmission mechanism, we perform the following. We take a model *M2* estimated under Rational Expectations and fix the parameters at their posterior mode values. Then, taking these parameter values as given, we construct initial beliefs under the information sets I1, I2, and I3 that are compatible with this REE.[11] Finally, using thus derived transmission mechanism (matrices T and R in Eq. 10), we calculate impulse response function of inflation, interest rate, and output to productivity, price mark-up, consumption preference, and monetary policy shocks. The results are presented in Fig. 3. The solid, dashed, and dotted lines correspond to impulse responses under I1, I2, and I3 information set, respectively. As we see from the figure, even in cases when the impact effect of a particular shock on a variable is similar under the three information sets, the impulse responses show very disparate transitions towards the long-run steady state (*i.e.*, the response of nominal interest rate to consumption preference shock). On another hand, there

[11] At the REE, the constants in forecasting functions must equal zero, therefore, information sets with and without the constant generate equivalent initial beliefs.

are impulse responses that diverge already on impact. Some of them also evolve in different directions during the adjustment (for example, the response of output to price mark-up shock). Therefore, for the same model parameters, the three alternative information sets used by learning agents for forecasting imply different transmission mechansims. Our results also indicate that there is a tendency for impulse responses with I3 information set to be more persistent but less pronounced in magnitude, at least at the REE consistent initial beliefs.

Another exercise useful for studying the properties of the transmission mechanism under alternative learning is the analysis of inflation persistence implied by different information sets. In order to perform this task, we take estimation results under learning for the model *M2* with three information sets. In every case, we fix the parameters at the corresponding posterior mode values and compute inflation persistence implied by the Actual Law of Motion, given by (13).[12] The results are presented in Fig.4. In accordance with the previous analysis of Impulse Responses, we see that initial I3 REE-consistent beliefs imply very persistent inflation. Another important observation is that implied inflation persistence under I1 and I2 information sets is rather stable over time, wherheas it exhibits dramatic changes under I3. This means that if "true" inflation persistence based on real data was changing fast, the I3 learning model would be better positioned to capture this data generating process. Given that I3 learning rule outperforms I2 and I1 in terms of the model fit (see Table 1), we suggest that the data generating process was indeed changing with the time, and I3 model was successful in capturing such a dynamic adjustment. We believe that our results are in line with some of the previous studies on inflation persistence (especially those that estimatied the persistence with time-variation in the mean of inflation or over short time sample), which suggest that inflation in the euro area might have been only moderately persistent in 1995-2002. For a detailed survey of such results see Table 3.1 in Altissimo et al (2006).

To provide somewhat more intuiton about the dynamics of implied inflation peristence, we consider the components that contribute to the update of the beliefs coefficients. Equation (12a) illustrates that, given the same forecasting error, the update of belief coefficients ϕ will depend on the value of the "effective" gain gR^{-1}, where R^{-1} is inverse of the matrix of the second moments. Figure 5 shows one of the terms of the second moments matrix corresponding to expected squared inflation for 3 models I1, I2, and I3. The graph indicates that squared inflation is expected to be very high under I1 information set, thus making the effective gain very small and resulting in relatilvely minor updates to the belief coefficients in response to the forecasting errors. This, in turn, leads to very smooth time series for implied inflation persistence. In a model with I2 information set, agents beleive that the second moment of inflation is much smaller and rather stable, which implies that beliefs are updated stronger than in I1 case, but the implied inflation persistence still does not vary much. Finally, agents using I3 information set beleive that, initially, expected squared inflation is rather high but drops fast, which is consistent with a noticeable fall in implied inflation persistence. This

[12] Note that we cannot compare the beliefs about inflation directly as in the Fig. 2, because the forecasting equations are different under the information sets I1-I3. In terms of the adaptive learning literature, Fig. 4 presents the Actual Law of Motion, or ALM, for inflation.

downward adjustment of perceived second moment of inflation combined with high estimated gain coefficient (about 2 times higher than for I2 or I1) result in a high value of "effective" gain. High gain means significant updates of belief coefficients which is reflected in a significant fall of implied inflation persistence under I3 beliefs before 2000.[13]

5 Conclusions

As far as the sensitivity of the estimation results to the chosen learning rule is concerned, we find that the more restrictive information set available to the agents is, the better is the model fit. In particular, the greatest improvement in marginal data density and the most significant change in the estimated parameters relative to the RE estimation are observed when the forecasts are made using univariate AR(1) processes. In general, the estimated parameters associated with different information sets used in forecasting rules vary significantly. We demonstrate that for the full information set consisting of all endogenous states and shocks — the same set of variables as under rational expectations, the estimated structural rigidity parameters as well as model marginal data density are closest to those obtained under RE. This conclusion is in line with some of the earlier research which documented little difference between the estimation results under RE and adaptive learning. Very importantly, we have established that this conclusion holds independently of the model complexity: In all three models considered here, ranging from a three equation New Keynesian model to a Smets and Wouters type model, forecasting forward-looking variables using univariate AR(1) processes brings the best results.

We believe that the reason for significant differences among the estimation results with alternative information sets is the effect of the set on transmission mechanism of the model: even for identical parameter values, the impulse responses implied by the REE-consistent initial beliefs (i.e., before any belief updating has taken place) show very disparate dynamic behavior of macroeconomic variables. Magnitude and persistence of the responses depend on the assumed information set to a significant degree, leading to divergence in the overall model fit and estimated parameters.

We also find that different ways of forming the initial beliefs influence the dynamics of the model under learning. REE-consistent initial beliefs produce more persistent and less volatile evolution of inflation expectations than pre-sample regression based initial beliefs. High volatility of pre-sample beliefs is a probable reason for the fact that they generally lead to a worse model fit than REE-consistent initial beliefs in our estimation. On the other hand, pre-sample beliefs could be very sensitive to short-term but pronounced developments in the data, which might lead to the agents' forecasting functions changing dramatically over the in-sample. This sensitivity is likely to lead to a large variability of estimated results, making them less creadible.

On the contrary, the REE-consistent initial beliefs are found to be very robust in adaptive

[13]Subsequent increase in implied persistence is explained by the fact that during this period the simple AR1 forecasting model was mostly underpredicting inflation, which led the agents to update perceived persistence of inflation upwards.

learning estimations, in the sense that moderate deviations in the REE employed to construct such initial beliefs lead to essentially identical marginal data density and estimated parameters. The fact that REE-consistent initial beliefs produce rather robust outcomes means that the estimation results are mainly driven by the time-varying transmission mechanism introduced by adaptive learning rather than by the specific initial conditions. The robustness of the REE-consistent initial beliefs can also motivate the preference for such "model-driven" approach to form the initial conditions when estimating the DSGE models for policy analysis and forecasting.

References:

Adolfson, M., Laseen, S., Linde, J, and M. Villani, (2006): "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through," Sweriges Riksbank WP18.

Altissimo, F., Ehrmann, M., and F.Smets, (2006): "Inflation persistence and price-setting behaviour in the euro area : a summary of the Inflation Persistence Network evidence," National Bank of Belgium WP95.

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

Chari, V.V., P. J. Kehoe, and E. R. McGrattan, (2009): "New Keynesian Models: Not Yet Useful for Policy analysis," *American Economic Journal: Macroeconomic*s, **1**(1), 242-266.

Del Negro, M., and F. Schorfheide, (2006): "How Good is What You've Got? DSGE-VAR as a Toolkit for Evaluating DSGE Models," *Economic Review*, Federal Reserve Bank of Atlanta, Q2, 21-37.

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

Gomes, O. (2010): "The Sticky Information Macro Model: Beyond Perfect Foresight," *Studies in nonlinear Dynamics and Econometrics*, **14**(1), Article 1.

Jǎǎskelǎ, J., and R. McKibbin (2010): "Learning in an Estimated Small Open Economy Model," Reserve Bank of Australi Research Discussion Paper 2010-02.

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

Milani, F. (2007): "Expectations, Learning and Macroeconomic Persistence", *Journal of Monetary Economics*, **54**, 2065–2082.

Milani, F., (2008): "Learning, Monetary Policy rules, and Macroeconomic Stability," *Journal of Economic Dynamics and Control*, **32**(10), 3148-3165.

Murray, J. (2007): "Empirical Significance of Learning in a New Keynesian Model with Firm-Specific Capital", CAEPR Working Paper 027.

Orphanides A., and J. Williams, "Inflation targeting under imperfect knowledge," *Economic Review*, Federal Reserve Bank of San Francisco, pages 1-23, 2007.

Primiceri, G.E. (2005): “Why Inflation Rose and Fell: Policymakers' Beliefs and U.S. Postwar Stabilization Policy," mimeo.

Reis, R. (2009): “A Sticky-Information General–Equilibrium Model for Policy Analysis,” NBER Working Paper 14732.

Sargent, T.J., N. Williams, and T. Zha (2005): “Shock and Government Beliefs: The Rise and Fall of American Inflation," *American Economic Review*, forthcoming.

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

Smets, F., and R. Wouters, “Forecasting with a Bayesian DSGE Model: An Application to the Euro Area," Journal of Common Market Studies, 42(4), 841-67, 2004.

Slobodyan,S., and R. Wouters (2009): “Learning in an Estimated Medium-Scale DSGE Model,” CERGE-EI WP 396.

Slobodyan,S., and R. Wouters (2010): “Estimating a Medium-Scale DSGE Model with Expectations Based on Small Forecasting Models,” mimeo.

Vilagi, B., “Adaptive learning and macroeconomic persistence: comparing DSGE models of the euro area," mimeo, 2007.

Woodford, M. (2003): Interest and Prices: Foundations of a Theory of Monetary Policy, Princeton University Press, Princeton, NJ.

Appendix. Tables and Figures

Table of priors

Parameters		Prior distribution, M1			Prior distribution, M2			Prior distribution, M3		
		Type	Mean	St.err	Type	Mean	St.err	Type	Mean	St.err
St. err, shocks:										
prefer. consum.	σ_b	*I.Gam.*	0.1	2	*I.Gam*	0.2	2	*I.Gam*	0.2	2
policy rule	σ_r	*I.Gam.*	0.1	2	*I.Gam*	0.1	2	*I.Gam*	0.1	2
price markup	σ_p	*I.Gam.*	0.1	2	*I.Gam*	0.15	2	*I.Gam*	0.15	2
wage markup	σ_w				*I.Gam*	0.25	2	*I.Gam*	0.25	2
productivity	σ_a				*I.Gam*	0.4	2	*I.Gam*	0.4	2
labor supply	σ_l				*I.Gam*	0.1	2	*I.Gam*	0.1	2
employment	σ_m				*I.Gam*	0.1	2	*I.Gam*	0.1	2
investment	σ_{inv}							*I.Gam*	0.1	2
price of capital	σ_q							*I.Gam*	0.4	2
gover.spend.	σ_g							*I.Gam*	0.3	2
AR coeff-ts:										
prefer. consum.	ρ_b	$U(0,1)$	0.5	0.29	*Beta*	0.85	0.1	*Beta*	0.85	0.1
productivity	ρ_a				*Beta*	0.85	0.1	*Beta*	0.85	0.1
labor supply	ρ_l				*Beta*	0.85	0.1	*Beta*	0.85	0.1
investment	ρ_{inv}							*Beta*	0.85	0.1
govern. spend.	ρ_g							*Beta*	0.85	0.1
Struct. params:										
Util. fun. cons.	σ	*Norm.*	1	0.375	*Norm.*	1	0.375	*Norm.*	1	0.375
Util. fun. labor	η				*Norm.*	2.5	0.25	*Norm.*	2	0.5
Habit	h	*Beta*	0.7	0.1	*Beta*	0.7	0.1	*Beta*	0.7	0.1
Index. prices	ι_p	$U(0,1)$	0.5	0.29	$U(0,1)$	0.5	0.29	$U(0,1)$	0.5	0.29
Index. wages	ι_w				$U(0,1)$	0.5	0.29	$U(0,1)$	0.5	0.29
Calo prices	ξ_p	*Beta*	0.75	0.1	*Beta*	0.75	0.05	*Beta*	0.75	0.05
Calvo wages	ξ_w				*Beta*	0.75	0.05	*Beta*	0.75	0.05
Calvo empl-nt	ξ_m				*Beta*	0.75	0.15	*Beta*	0.5	0.15
Int.rate smooth	w	$U(0,1)$	0.5	0.29	$U(0,1)$	0.5	0.29	*Beta*	0.8	0.1
pol. rule, inflat.	ζ_π	*Norm.*	1.5	0.1	*Norm.*	1.5	0.1	*Norm.*	1.7	0.1
pol. rule, out.	ζ_y	*Norm.*	0.5	0.1	*Norm.*	0.5	0.1	*Norm.*	0.125	0.05
Inv.adjust.cost	φ							$U(1,11)$	6	2.5
Var.cap. utiliz.	ψ							*Norm.*	0.2	0.075
Gain (learning)	g	*Gam.*	0.035	0.03	*Gam.*	0.035	0.03	*Gam.*	0.035	0.03

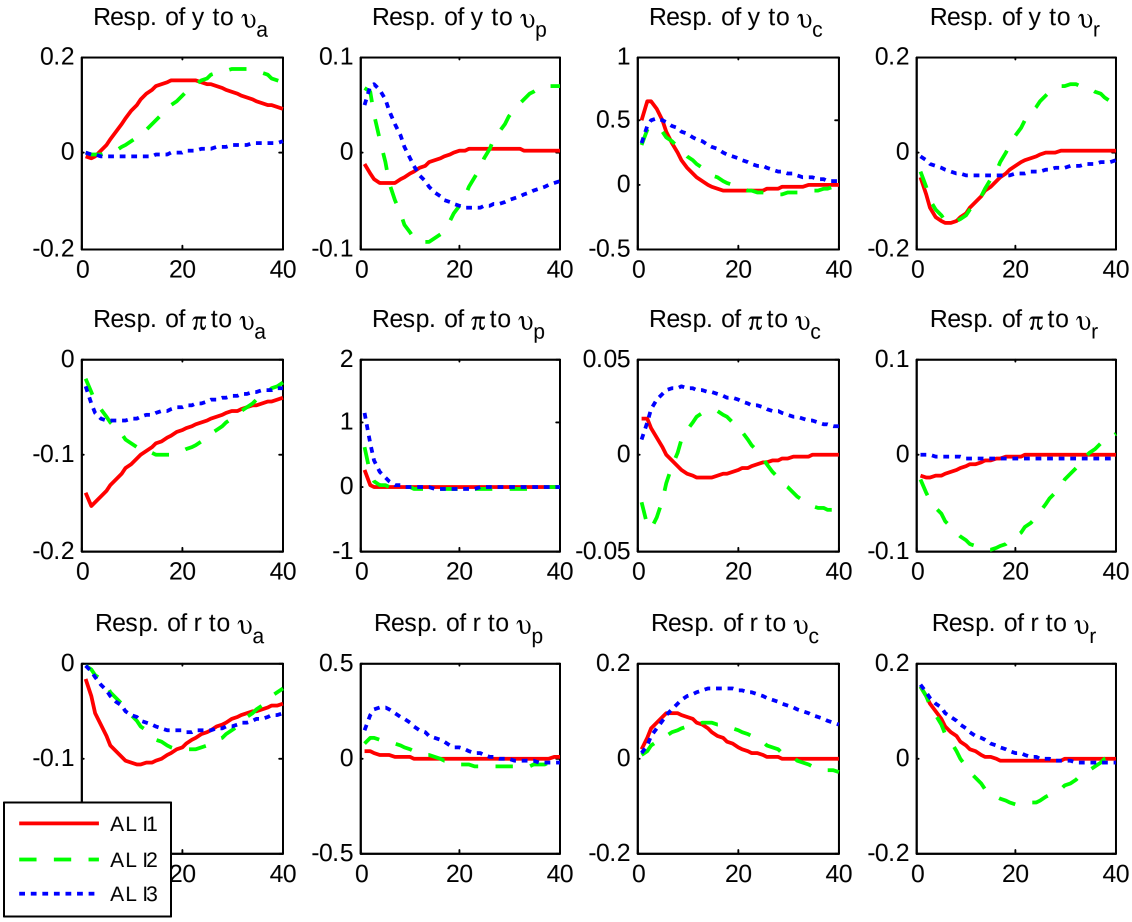

Figure 3: Impulse response functions under different information sets

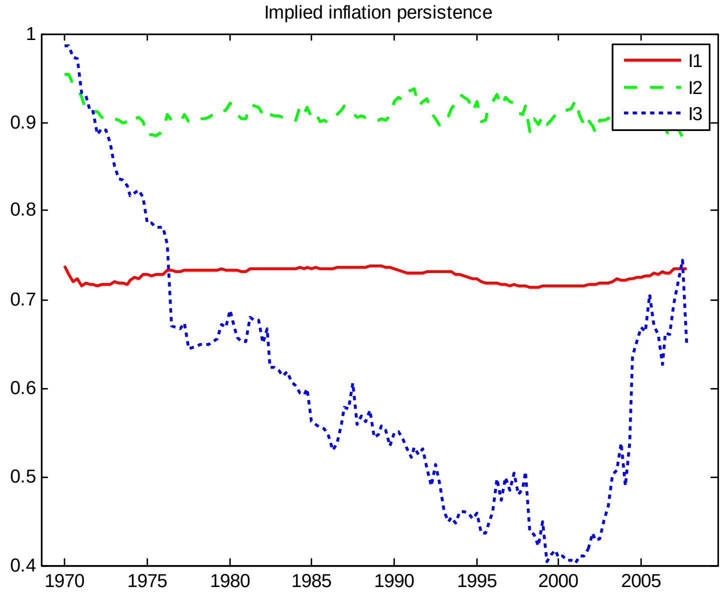

Figure 4: Implied inflation persistence under different information sets

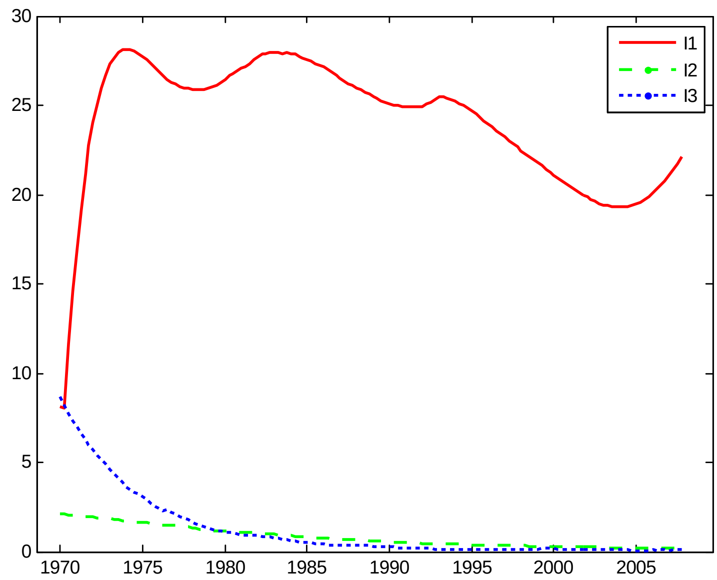

Figure 5: Implied second moments of inflation under different information sets