

On the Initialization of Adaptive Learning in Macroeconomic Models

MICHELE BERARDI

University of Manchester

JAQUESON K. GALIMBERTI

ETH Zurich

This version: November 6, 2015

Abstract

We review and evaluate methods previously adopted in the applied literature of adaptive learning in order to initialize agents' beliefs. Previous methods are classified into three broad classes: equilibrium-related, training sample-based, and estimation-based. We conduct several simulations comparing the accuracy of the initial estimates provided by these methods and how they affect the accuracy of other estimated model parameters. We find strong evidence against their joint estimation, instead recommending the adoption of pre-determined initials based on a training sample. We also demonstrate the empirical relevance of our results by estimating a New Keynesian Phillips curve with learning.

Keywords: expectations, adaptive learning, initialization, algorithms.

JEL codes: C63, D84, E03, E37.

1 Introduction

Adaptive learning algorithms have been proposed to provide a procedural rationality view on agents' process of expectations formation. Reopening a long standing debate on how should expectations be modeled in macroeconomic models, the heuristics provided by learning algorithms come at the cost of introducing new degrees of freedom into the analysis. One open node relates to how these recursive mechanisms should be initialized in order to be representative of agents' learning-to-forecast behavior.

The main characteristic of the adaptive learning approach is its reliance on recursive algorithms in order to represent how agents update their beliefs as new observations about the economic relationship of interest becomes available. Such recursions naturally demand an initial starting point, and it is the numerical specification of these conditions that we denote as the initialization

problem. Clearly, the uncertainties affecting the initialization of the learning process will propagate recursively into the predictions obtained with the model, and it seems crucial that the researcher understands the magnitude of these distortions and for how long they may be expected to last.

In this paper we investigate this issue with particular attention at the applied literature of learning in macroeconomics. Here applied is taken to encompass both theoretical simulations as well as exercises of empirical estimation and calibration. Examples can be found in Sargent (1999); Marcet and Nicolini (2003), or more recently in Eusepi and Preston (2011); Milani (2011), between many others cited throughout the paper. The main distinctive feature of these works consists in the replacement of the rational expectations (RE) assumption of an instantaneous adjustment of agents expectations, with a characterization of agents as adaptive learners of their own environment. More generally, our study will be relevant for scholars interested in the methods needed to uncover the initial beliefs of economic agents in models where such beliefs actually matter for economic dynamics.

The relevance of the initialization issue can be illustrated considering the long debated causes of the period of Great Inflation during the 1970s in the US. One of the main explanations for that episode comes from Sargent's (1999) hypothesis that the evolution US inflation rates over the period can be attributed to the evolution of the monetary authority's beliefs about the trade-off between inflation and unemployment, the so-called Phillips curve. Subsequent studies advanced on this issue attributing the rise in inflation rates to delayed policy responses to ongoing structural changes in the economy of that period (see, e.g., Bullard and Eusepi, 2005; Orphanides and Williams, 2005a). Importantly, as evidenced in Primiceri (2006); Sargent et al. (2006), the point of departure in policymaker's beliefs, which is what we refer to as the learning initials in this particular context, is a crucial feature in such explanation.

Assumptions about initial beliefs also matter for the fit of models that introduce adaptive learning on the side of other market participants, such as households and firms. Examples are given by Carceles-Poveda and Giannitsarou (2008) for asset pricing models, Huang et al. (2009) in a standard growth model, and Slobodyan and Wouters (2012) in a medium-scale dynamic stochastic general equilibrium (DSGE) model. Overall, these studies present results showing that whereas the introduction of learning has interesting effects on the dynamics and the fit of models to the data, a great portion of the improvements may be associated to transition dynamics from specific initial beliefs. Hence, it is important to have a systematic evaluation of the different alternatives available as initialization methods, and we attempt to fill that gap in this paper.

We review the literature in order to pool together the pre-existing initialization methods into an archetypal classification that can be broadly defined in three major classes: equilibrium-related methods, training sample-based methods, and estimation-based methods. The equilibrium-related initializations are generally obtained taking RE equilibrium (REE) as a reference, and exploring

distributional deviations from that assumption. The training sample-based initializations, as the name suggests, are obtained with the application of the learning algorithm (or variations) over a pre-sample of observations that is left aside from the original sample of data available. Finally, the more recent estimation-based methods consist of approaches involving the joint estimation of the initials with other model parameters, hence allowing the use of the same data that is used for inferences about structural features of the model to guide the specification of learning initials.

In spite of the obvious relevance of the initialization issue, surprisingly, we did not find many other attempts to systematically assess such methods. One exception is provided by Carceles-Poveda and Giannitsarou (2007), although their analysis is mainly methodological, and not empirical; in the asset pricing context, Carceles-Poveda and Giannitsarou (2008) show how to apply some of the methods they have developed earlier with quantitative exercises. Slobodyan and Wouters (2012) also conduct extensive analysis on the effects of different specifications of the initials, but focusing mainly on approaches for their joint estimation with other model parameters. Therefore, such applications cover only specialized versions of the broad classes of methods we consider in this paper.

With the freedom to develop our own assessment framework, we compare the initialization methods on the basis of the accuracy of their delivered initial estimates and their effects over the accuracy of other estimated model parameters. For that purpose we conduct several simulation exercises based on a model of the New Keynesian Phillips curve (NKPC) under learning. To evaluate accuracy we calculate measures of the Mean Squared Deviation (MSD) between true parameter values used for the model simulations and their corresponding estimates obtained according to the different initialization methods. We relate the MSD measures to two main principles to judge the quality of an initialization. First, we look at the coherence of the initialization estimates to the dynamics implied by the learning process; second, we consider how susceptible the method is to biases that push up the model's explanatory power over the initial portion of observations in the sample.

Our simulation results point to some interesting findings. First, the training sample-based initialization methods are in general favored in terms of the coherence criterion, since these methods are found to provide more accurate estimates of the learning initials. Second, the equilibrium-related methods tend to be less accurate than those based on a training sample, a result mainly due to the fact that learning dynamics are not taken into account in initials based on equilibrium assumptions, or on estimates of the model under equilibrium conditions. Finally, the worst outcomes are observed under the estimation-based initialization method, where initials are jointly estimated with model's structural and learning parameters. Particularly concerning, we find that biases introduced by the joint estimation of initials can spillover to the estimates of other model parameters, and these distortions can persist in large samples too.

This last finding is of particular relevance for empirical analysis, where interest is usually in uncovering the underlying values of structural parameters that may validate the model’s consistency with data evidence. In order to further enhance our understanding on the relevance of these different initialization methods for applied macroeconomics, we also present an empirical application on the determination of US inflation rates under the Phillips curve framework. Adopting a generalized method of moments (GMM) estimation approach, we find that initials can, indeed, affect the estimates of structural parameters. Particularly, we show that the distortions associated with the approach of joint estimation of initials can overturn evidence that is generally in favor of the introduction of a backward-looking rule of thumb in the way firms set their prices within the NKPC framework; in contrast, this hybrid NKPC specification can not be statistically ruled out when the learning recursions are initialized using training sample-based methods.

The remainder of this paper proceeds as follows. In section §2 we provide a brief introduction to the use of adaptive learning in macroeconomic models, and establish the initialization problem. A review of the initialization methods previously adopted in the literature is presented in section §3. We then proceed to present our simulation analysis, in section §4, and an empirical application, in section §5, both aiming at a comparative evaluation between the different methods of initialization. Finally, we conclude this paper with our main recommendations in section §6.

2 Adaptive Learning and the Initialization Problem

2.1 A brief primer on adaptive learning

Adaptive learning is introduced in macroeconomic models as an alternative to the assumption that agents hold rational expectations. One implication of the rational expectations assumption is that agents’ beliefs are always consistent with the true model of the economy. Hence, under RE the economy instantaneously adjusts itself towards an equilibrium after any kind of shock that may have realistically affected agents’ beliefs. In contrast, adaptive learning introduces some degree of persistence in the process through which agents update their beliefs, which allows such beliefs to deviate from RE in the short run, while keeping up with the idea of consistency in the long run.

To help fix ideas consider an univariate linear forward-looking model, where the determination of the current value of a variable of interest, y_t , depends on the value expected for that same variable in the next period, y_{t+1}^e plus a mean zero random shock, u_t , i.e.,

$$y_t = \beta y_{t+1}^e + u_t. \tag{1}$$

Simple as it stands, this specification may represent the reduced form of the equilibrium equations of an economic model which could potentially be non linear; it also corresponds, e.g., to simplified versions of two well known models: the Cagan model of inflation, letting y_t stand for the price level and u_t for a mean zero random supply of money; and, the standard model of asset pricing under risk neutrality, letting y_t stand for the asset price and u_t for a mean zero random sequence of dividends.

Model (1) admits multiple RE solutions depending on agents' perceived law of motion (PLM), which specifies how agents form expectations. Particularly, if agents condition their forecasts on a constant, $y_{t+1}^e = 0$ solves the model for any β , and when $\beta = 1$, any value of y_{t+1}^e would represent an RE equilibrium as well. Hence, the stochastic process followed by the economy, also known as the actual law of motion (ALM), is directly determined by the specification of agents' PLM.

Under learning the corresponding ALM is given by

$$y_t = \beta \phi_{t-1} + u_t, \quad (2)$$

where ϕ_{t-1} denotes agents' estimates of the constant in their PLM based on observations available up to the previous period. Different recursive algorithms have been proposed in the literature to represent how agents update such estimates. Due to its widespread popularity between econometricians, one natural choice for that purpose has been the Least Squares (LS) algorithm, which can be generally defined as follows.

Algorithm 1 (LS). *Let agents' PLM of y_t be given by a linear regression of the form*

$$y_t = \mathbf{x}_t' \boldsymbol{\phi}_{t-1} + \varepsilon_t, \quad (3)$$

where $\mathbf{x}_t = (x_{1,t}, \dots, x_{K,t})'$ is a set of pre-determined variables, possibly including a constant (e.g., $x_{1,t} = 1$) and lags of y_t , $\boldsymbol{\phi}_t = (\phi_{1,t}, \dots, \phi_{K,t})'$ stands for a vector of coefficients, possibly time-varying, and ε_t denotes an unpredictable disturbance term. Under this context, the LS estimates of $\boldsymbol{\phi}_t$, conditional on observations up to time t , are given by

$$\hat{\boldsymbol{\phi}}_t = \hat{\boldsymbol{\phi}}_{t-1} + \gamma_t \mathbf{R}_t^{-1} \mathbf{x}_t (y_t - \mathbf{x}_t' \hat{\boldsymbol{\phi}}_{t-1}), \quad (4)$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \gamma_t (\mathbf{x}_t \mathbf{x}_t' - \mathbf{R}_{t-1}), \quad (5)$$

where γ_t is a learning gain parameter, and \mathbf{R}_t stands for an estimate of the regressors matrix of second moments.

The LS algorithm is originally motivated as the result from the minimization of a weighted sum of squared errors, where the weights are determined by the learning gain parameter (see Berardi and Galimberti, 2013). Hence, the learning gain stands for a parameter determining how quickly

a given information is incorporated into the algorithm's coefficients estimates. There are two particular cases of interest: (i) when $\gamma_t = 1/t$, every observation receives the same weight and (4)-(5) reduces to the (recursive) Ordinary Least Squares (OLS); and, (ii) under a constant gain, past observations receive geometrically decaying weights and (4)-(5) can be viewed as a (recursive) Weighted Least Squares (WLS) with weights given by $(1 - \gamma)^j$, where j indexes for the number of periods between the weighted observation and the last observation in the sample. Our focus is on the constant gain specification due to its relevance for applied purposes: because it allows for a continuous operation of the algorithm's tracking capabilities, the constant gain can capture time-varying effects of different sources, such as structural breaks or the out-of-equilibrium dynamics generated by stochastic shocks.

In the context of our simple example model, because agents' PLM includes only a constant term, the LS recursions in (4)-(5) simplify to

$$\phi_t = \phi_{t-1} + \gamma (y_t - \phi_{t-1}). \quad (6)$$

Once a learning mechanism is specified, one of the main issues in the theoretical literature on adaptive learning has been to single out the conditions under which this learning process converges towards an equilibrium. Following Evans and Honkapohja (2001), these conditions are equivalent to the conditions governing the E-stability of RE solutions, which in the case of our simple model, together with the assumption of a sufficiently small gain, is given by $\beta < 1$. Thus, by ruling out the indeterminate RE solution with $\beta = 1$, convergence of adaptive learning provides a successful equilibrium selection criterion in model (1). Importantly, under constant gain learning most of the theorems for convergence are local, meaning that the initial conditions are also important for the determination of whether the algorithm converges or not.

In contrast to this theoretical issue, in the applied side of the literature the key issue is about the actual behavior of the learning algorithm in finite samples. For instance, one may be interested in understanding how much of macroeconomic persistence can actually be attributed to learning (as in Orphanides and Williams, 2005b; Milani, 2007), and what part of business cycles fluctuations can be explained by a model with learning and expectational shocks (as in Eusepi and Preston, 2011; Milani, 2011). The empirical nature of this kind of questions naturally raises further quantitative issues about the approach required for the estimation/calibration of the model and learning parameters with actual data. One such an issue is how to properly initialize the learning recursions.

2.2 The initialization problem

Recursive learning algorithms naturally demand an initial starting point, and it is the numerical specification of these conditions that we denote as the initialization problem. By the recursive

nature of learning, any error in the initial estimates will propagate recursively into the predictions obtained with the model. Consider the case of our example model, (2)-(6); letting $\hat{\phi}_0$ stand for a guess of the true value of ϕ_0 , the model prediction of y_{t+1} associated with this initial is given by $\hat{y}_{t+1} = \beta \hat{\phi}_t$, where $\hat{\phi}_t$ is obtained from (6) as

$$\begin{aligned}\phi_0 &= \hat{\phi}_0, \\ \hat{\phi}_1 &= \hat{\phi}_0 + \gamma (y_1 - \hat{\phi}_0), \\ &\vdots \\ \hat{\phi}_t &= \hat{\phi}_{t-1} + \gamma (y_t - \hat{\phi}_{t-1}).\end{aligned}\tag{7}$$

Let the corresponding prediction error be denoted by $\hat{\Delta}_{t+1} = y_{t+1} - \hat{y}_{t+1}$; then, the mean squared prediction error (MSPE) from this model amounts to¹

$$\begin{aligned}E [\hat{\Delta}_{t+1}^2] &= E \left[\left(\beta (\phi_t - \hat{\phi}_t) + u_{t+1} \right)^2 \right], \\ &= \beta^2 (1 - \gamma)^{2t} \hat{\delta}_0^2 + \sigma_u^2,\end{aligned}\tag{8}$$

where $\hat{\delta}_0 = \phi_0 - \phi_0$ is the initialization error and σ_u^2 is the variance of u_t . Clearly, assuming that $0 < \gamma < 1$, (8) shows that the effects of initialization errors tend to disappear as the distance from the initial point increases. Also notice that as γ increases, the prediction error associated to an initialization error decreases. Hence, the smaller the learning gain, the more important are the learning initials for the accuracy of the predictions obtained with the model. Importantly, (8) also establish a connection between the relevance of the learning initials and the main measure we use to evaluate the different initialization methods, namely, the Mean Squared Deviation (MSD) between the true and the estimated parameters of interest.

Definition 1 (MSD). The Mean Squared Deviation between the true values of a vector of parameters, θ_t , which may include learning initials (e.g., ϕ_0 , $vec(\mathbf{R}_0)$) and time-invariant model parameters, and a corresponding vector of estimates, $\hat{\theta}_t$, is given by

$$\mathcal{D}_t = E [\delta_t^2],\tag{9}$$

where $\delta_t = \|\theta_t - \hat{\theta}_t\|$ stands for the Euclidean norm of the vector of parameters deviations.

For empirical purposes both the model and learning parameters, say β and γ in our example model, are not known *a priori* and may therefore require estimation. In that case parameters'

¹See Appendix A.1 for details on derivations.

estimation error will also affect the accuracy of the model predictions. To see that consider the case where only β is unknown, and an estimate of its value is given by $\hat{\beta}$. Assuming $\hat{\beta}$ is uncorrelated with u_t , i.e., $E[\hat{\beta}u_t] = 0$, the MSPE is then given by

$$\begin{aligned} E[\hat{\Delta}_{t+1}^2] &= E\left[\left(\beta\phi_t - \hat{\beta}\hat{\phi}_t + u_{t+1}\right)^2\right], \\ &= E[\hat{\delta}_t^2] + \sigma_u^2, \end{aligned} \tag{10}$$

where $\hat{\delta}_t = \beta\phi_t - \hat{\beta}\hat{\phi}_t$, which can no longer be solved as a deterministic function of the initials and the model parameter's errors². The effects of parameters' uncertainty can be further aggravated considering that the accuracy of the estimates can depend on the initialization errors itself. In fact, as we will show in our simulation analysis further below, the estimation of the model and learning parameters can be severely affected by the misspecification of the learning initials.

3 Review of Initialization Methods

3.1 Equilibrium-related methods

One way to initialize learning algorithms is obtained by using the existing knowledge about the law of motions generating the data. Particularly, conditional on the knowledge about the model specification and the parameter values, one can easily obtain the REE implied values of agents' PLM coefficients and use these equilibrium values as reference for the initial estimates. Although this method was naturally appealing in earlier works with theoretical simulations, such as in Bray and Savin (1986), its debut into the applied literature came in the seminal contribution of Sargent (1999). Its usage has since been prominent in studies on the effects of replacing the assumption of frictionless REE by the sticky process of expectations formation through adaptive learning (e.g., Marcet and Nicolini, 2003; Bullard and Eusepi, 2005; Orphanides and Williams, 2005b).

For simulations, one way to obtain robust inferences is to draw the initials from a distribution centered around the REE values, such as the normal distribution with a variance given by the asymptotic variance of the learning coefficients' estimates (see Carceles-Poveda and Giannitsarou, 2007). In the case of our example model, such initials would be given by³ $\hat{\phi}_0 \sim N\left(0, \frac{\gamma^2 \sigma_u^2}{1 - (\gamma(1 - \beta))^2}\right)$. Empirically, it is the uncertainties about the true model parameters that may complicate the adoption of this method. One alternative is to approach the issue in two stages: first, model estimates are obtained under the REE assumption; second, these estimates are used to

²In the Appendix we show that $\hat{\delta}_t$ depends on the initial error, $\hat{\delta}_0$, plus a weighted sum of $\{y_i\}_{i=1}^t$.

³See Appendix A.2 for the derivation of ϕ_t long-run variance.

calculate the PLM coefficient values corresponding to the REE, which are then plugged back in as initial estimates for the algorithm's recursion for the analysis under learning (see Slobodyan and Wouters, 2012; Ormeño and Molnár, 2015). One criticism to this practical solution is that it seems very likely that the REE estimates obtained in the first step will be biased for not taking the learning effects into account. Besides, the REE-based initials do not provide ideal initial estimates for cases where there is prior information that the economy was in a transient phase at the beginning of the sample.

One alternative is provided by the *ad-hoc* initialization method, where the initials are hand-picked by the researcher. When taking the REE-based initials as a reference, this method provides a way to validate the sensitivity of results obtained under the former approach (e.g., Milani, 2007; Carceles-Poveda and Giannitsarou, 2008). In fact, one of the main uses of *ad-hoc* initials is to deal with the possibility of structural changes around the periods of the initials: when the changes affect the REE, agents may not be able to instantaneously adjust to the new equilibrium, and could therefore be forming expectations consistent with the previous equilibrium at the time of the initialization (see also Carceles-Poveda and Giannitsarou, 2007, p. 2679).

3.2 Training sample-based methods

Another common approach in learning applications is to use a pre-sample of observations in order to obtain the initial estimates. This is especially recommended for the cases where there is not enough previous knowledge about the system under estimation so as to allow an educated guess. The origins of this approach can be traced back into the engineering literature (see, e.g., Ljung and Soderstrom, 1983, pp. 299-303), where it is often suggested that the coefficients should be initialized with the value of zero and an initial training sample should be left aside to let the algorithm adjust its estimates according to the underlying calibration.

For applied purposes, it is often easier to adopt the non-recursive version of the learning algorithm to estimate the initials over the training sample. Letting P denote the number of observations set aside for the initialization, application of (4)-(5) in the training sample results in⁴

$$\mathbf{R}_P = \gamma_P \sum_{i=1}^P w_i \mathbf{x}_i \mathbf{x}_i' + w_0 \mathbf{R}_\emptyset, \quad (11)$$

$$\hat{\phi}_P = \gamma_P \mathbf{R}_P^{-1} \sum_{i=1}^P w_i \mathbf{x}_i y_i, \quad (12)$$

⁴For detailed derivations, please refer to Berardi and Galimberti (2013).

where $\{w_i\}_{i=0}^P$ stands for the sequence of weights given to each observation in the training sample, and \mathbf{R}_0 may incorporate prior information regarding the uncertainty surrounding the determination of the coefficients estimates. Under the assumption of a Gaussian random walk parameter drift model for ϕ_t , Berardi and Galimberti (2013) have shown that \mathbf{R}_t is inversely related to the matrix of mean squared errors associated to the Kalman filter coefficients estimates, $E \left[\left(\phi_t - \hat{\phi}_t \right) \left(\phi_t - \hat{\phi}_t \right)' \right]$. Hence, in a Bayesian interpretation, as $\mathbf{R}_0 \rightarrow \mathbf{0}$ the prior becomes more diffuse, since it is associated with a higher uncertainty about the coefficients estimates⁵.

Depending on the weighting scheme and the prior estimates, there are two main variations of this method in the learning literature: the OLS-based (e.g., Williams, 2003; Orphanides and Williams, 2005a; Sargent et al., 2006) and the WLS-based (e.g., Primiceri, 2006; Milani, 2007, 2008, 2011, 2014; Huang et al., 2009; Chevillon et al., 2010; Eusepi and Preston, 2011; Lubik and Matthes, 2014) initials. The OLS-based method, as the abbreviation suggests, is based on the Ordinary Least Squares estimator that is widely known among econometricians for possessing some well desired properties, such as consistency and efficiency in the estimation of linear models. For training sample initialization, it is obtained by setting $\gamma_P = 1/P$, $w_i = 1$, and $\mathbf{R}_0 = \mathbf{0}$ in (11)-(12). One important advantage of this method relates to its convergence speed: the fact that a relatively higher gain value is used in the first iterations of the algorithm within the training sample tends to accelerate its convergence to the true initials.

The WLS-based method derives from the Weighted Least Squares interpretation of the learning algorithm under a constant gain specification. In the training sample the initials associated to this method are obtained by setting $\gamma_P = \gamma$, and $w_i = (1 - \gamma)^{P-i}$ in (11)-(12). For the prior, \mathbf{R}_0 , we consider two possibilities. One, based on REE reasoning, is to set it to the sample estimate of the long-run covariance matrix of the regressors. Ideally, the sample used for such estimation should be restricted to the training sample itself, in order to prevent contamination of the initials due to the effects of changes in the statistical properties of the data that were not present before the initialization period. The second alternative we consider here is to follow a diffuse approach and set $\mathbf{R}_0 = \mathbf{0}$.

3.3 Estimation-based methods

Another approach to the initialization of learning coefficients is to add the initials to the set of the model's parameters, and estimate them jointly. The idea can be traced back to the landmark work by Sargent et al. (2006), where the estimation of the monetary authority's initial and consecutive stream of beliefs provided evidence in favor of Sargent's (1999) hypothesis on the "Conquest of American Inflation": namely, that the rise and fall of post-WWII inflation in the US can be at-

⁵Notice that when $\mathbf{R}_0 = \mathbf{0}$, a necessary condition for \mathbf{R}_P to be invertible, as required in (12), is that $P \geq K$.

tributed to the evolution of the monetary authority's beliefs about the trade-off between inflation and unemployment. In spite of some early criticisms (see discussion in Carboni and Ellison, 2009), the approach of joint estimation of initials has been slowly incorporated into broader applications of adaptive learning. After being hinted as a possibility in Milani (2007, p. 2071) and Huang et al. (2009, p. 397), initial attempts have focused on the effects of the joint estimation of the initial matrix of second moments, \mathbf{R}_0 (e.g., Milani, 2008, 2011), and more recently on the estimation of the complete set of learning initials (as in, e.g., Slobodyan and Wouters, 2012; Gaus and Ramamurthy, 2014).

Naturally, an estimation-based initialization method is only relevant for empirical applications and its implementation details depend on the adopted model estimation methodology. Nevertheless, most estimation approaches share a common idea of looking for the combination of parameter values that maximize the fit of the model, or its implications, to available macroeconomic data. Hence, the joint estimation of learning initials can have an appealing motivation for providing those estimates of the initial beliefs that are the most consistent with the data according to the chosen empirical criterion. In fact, that would be the case if incorporating the initials into estimation routines did not cause side effects on the estimation of the other model's structural parameters. Unfortunately, as we will show in our simulation analysis in the next section, that is not the case. As we have previously alluded, the effect of the initials over the estimation criterion can perpetuate long enough into the sample so as to affect the identification of the structural parameters of interest.

3.4 Mixed approaches

Initialization methods can have several nuances that may not be, strictly speaking, reflected into the classes we proposed above. Particularly, there are many possibilities involving a mixture of the different approaches. For example, the REE-based initials could be computed on the basis of estimates of the model parameters obtained using data solely from the training sample. A similar approach has been used in Slobodyan and Wouters (2012), though adding the OLS-based method to the mixture: after estimates of the model under RE are obtained, using either the training or the whole sample, the initials are set to the REE-implied OLS estimates of agents' PLM. Another example is given in Milani (2011), where the mix is between the WLS-based method and the estimation-based approach: for every draw in the Bayesian estimation routine, a training sample of observations is used to compute the initial matrix of second moments according to (11), plugging in the corresponding estimated learning gain.

Another approach, developed in Berardi and Galimberti (2012), relies on the use of Kalman smoothing within a sample of training data in order to accelerate the convergence of the WLS-based initialization method. Although this approach requires additional computations, it has been

found to provide important speed improvements under alternative specifications of the learning mechanism, such as the Stochastic Gradient (SG) algorithm (Barucci and Landi, 1997; Evans and Honkapohja, 1998), which is obtained by replacing \mathbf{R}_t^{-1} by an identity matrix in the LS specification of (4). Hence, the SG does not benefit from the LS “normalization” step given by the inverse of the matrix of second moments, which prevents the use of a diffuse prior on its initialization. As our results will show, the diffuse prior on \mathbf{R}_0 provides an easy way to accelerate the LS algorithm convergence during the training sample.

Finally, although we have focused our discussion on the use of actual data on the variables included in agents’ PLM, another alternative is to use data from survey-based forecasts in order to get information about the initial conditions. Data on survey forecasts have been broadly taken as proxy for agents’ actual expectations. In most of the cases, the initialization methods we discussed above can be adjusted to take advantage of this information. For example, the REE-based initials can be calculated using model’s estimates obtained by replacing expectation terms by direct measures from surveys (see, e.g., Orphanides and Williams, 2005a). Learning initials consistent with surveys’ information can also be obtained by adjusting the estimation-based method to maximize the fit of the forecasts implied by the learning estimates to those obtained from survey forecasts (Pfajfar and Santoro, 2010). Although we recognize the potential value of these alternatives, we restrict the scope of our analysis to the definitions of initialization methods covered by our classification.

4 Simulation Analysis

4.1 Baseline Phillips curve model

In order to shed some light on the comparative evaluation of the initialization methods reviewed above, we now analyze their quantitative properties with simulations. To provide a meaningful economic example, we focus on a standard New Keynesian Phillips curve (NKPC) model, given by

$$\pi_t = \beta\pi_{t+1}^e + \lambda x_t + u_t, \quad (13)$$

$$x_t = \rho x_{t-1} + v_t, \quad (14)$$

where π_t is inflation, π_{t+1}^e represents agents’ expectations for next period inflation, x_t is a proxy for real marginal cost, and u_t is a disturbance which can be interpreted as a measurement error or as an unobserved cost-push shock. The parameters in (13) are taken as semi-structural in the sense that they can be associated to deeper structural parameters of a microfounded model (see,

e.g., Mavroeidis et al., 2014). Particularly, β is the subjective discount factor and

$$\lambda = \frac{(1 - \theta)(1 - \theta\beta)}{\theta} \kappa, \quad (15)$$

where $\theta \in (0, 1)$ represents the fraction of firms that cannot change their prices in any given period, i.e., an index of price rigidity under the Calvo framework, and $\kappa \leq 1$ is a function of the labor elasticity of production and the price elasticity of demand.

The RE solution of this model is given by

$$\pi_t = \frac{\lambda}{1 - \beta\rho} x_t + u_t. \quad (16)$$

It can be shown that this equilibrium is E-stable if $\beta\rho < 1$, a condition that is automatically met under the usual assumptions that $0 < \beta < 1$ and $|\rho| < 1$ (see Evans and Honkapohja, 2001, pp. 198-200). Consistent with this solution, under adaptive learning agents form expectations according to a PLM given by

$$\pi_t = \phi_{t-1} x_t + z_t, \quad (17)$$

where ϕ_t is a parameter estimated with the univariate version of the LS algorithm given by (4)-(5), also substituting $y_t \equiv \pi_t$. Iterating (17) forward and substituting the expectations in (13) we obtain the ALM under learning

$$\pi_t = (\beta\rho\phi_{t-1} + \lambda) x_t + u_t. \quad (18)$$

4.2 Simulation and estimation approach

We generate 10,000 samples of artificial series of π_t and x_t assuming that $u_t \sim N(0, \sigma_u^2)$, $v_t \sim N(0, \sigma_v^2)$, and that $Correl(u_t, v_t) = 0$. The number of observations used for the learning initialization, in the training sample-based methods, and the estimation of the model's parameters will be a dimension of our analysis, but in general we simulate the model for 1,000 observations and assume the sample of data available for estimation starts at the 1,001 observation, i.e., $t = 0$ is observation 1,000 in our artificial series. The model parameters are set to $\beta = 0.99$, $\theta = 0.65$, $\kappa = 0.25$, $\rho = 0.9$, $\sigma_u^2 = 3$, $\sigma_v^2 = 1$, whereas for the learning gain we evaluate two options, $\gamma_1 = 0.02$ and $\gamma_2 = 0.10^6$. The initial parameters of the learning recursions are set to their associated RE equilibrium value, i.e., $\phi_{-1000} = \lambda / (1 - \beta\rho)$, and $R_{-1000} = \sigma_v^2 / (1 - \rho^2)$.

⁶Our findings are qualitatively insensitive to these choices of parameters values, but not quantitatively. As evidenced in (8), though under a simpler model specification, the impact of initialization errors over the accuracy of the model's predictions is positively related to the magnitude of the parameter associated with the forward-looking term, β .

In order to estimate the model and learning parameters, we follow a generalized method of moments (GMM) approach. Following Chevillon et al. (2010), we obtain the moment conditions from the common assumption that the unobserved disturbance term, u_t in our model, is a martingale difference sequence, which means $E_{t-1}[u_t] = 0$. Using the model's ALM under learning, (18), we can define the residual function according to

$$h_t(\Theta) = \pi_t - \beta\rho\phi_{t-1}x_t - \lambda x_t, \quad (19)$$

where Θ denotes the set of parameters requiring estimation. For a given set of pre-determined instruments, Z_t , the corresponding moment conditions are given by

$$E[Z_t h_t(\Theta)] = 0. \quad (20)$$

Here we take the first two lags of π_t and x_t , and a constant as instruments⁷. The model parameters are then estimated by minimization of the associated GMM objective function

$$g_T(\hat{\Theta}) = \left[T^{-1} \sum_{t=1}^T Z_t h_t(\hat{\Theta}) \right]' W_T \left[T^{-1} \sum_{t=1}^T Z_t h_t(\hat{\Theta}) \right], \quad (21)$$

which is constructed from the sample counterpart of the moment conditions in (20) and a weighting matrix, W_T . This weighting matrix is optimally defined as a consistent estimator of the inverse of the long-run variance of the moment conditions. Because the variance of (20) depends on the values of Θ , we adopt an iterative GMM estimator (see, e.g., Hall, 2005): we first set $W_T^{(0)} = I$ to obtain the preliminary estimates $\hat{\Theta}^{(0)}$ that minimize (21); we then use the Newey and West (1987) heteroskedasticity and autocorrelation consistent estimator of the variance of the moment conditions evaluated at $\hat{\Theta}^{(0)}$ to obtain a new estimate of $W_T^{(1)}$; we repeat this process until a convergence criterion is achieved⁸.

⁷Chevillon et al. (2010) demonstrated that inference in models with adaptive learning is plagued by weak identification and persistence problems, suggesting the use of the Anderson-Rubin statistic and the replacement of lags of the endogenous variable by lags of $h_t(\hat{\Theta})$ in the list of instruments to deal with these problems. Since we are not primarily interested in statistical inference, here we follow the standard GMM estimation approach, though experimenting with the alternative setting of instruments was found to have no major effects over our main conclusions about initial estimates.

⁸Details about this procedure and the numerical optimization routine used to minimize (21) are provided in Appendix B.

4.3 Results

We conduct several simulation exercises to evaluate the accuracy of four methods of initialization, namely: the REE-based initials, the OLS-based initials, the WLS-based initials, and GMM estimation-based initials. For the former three methods, which we denote as pre-determined initials, we also feed the initials into the estimation of the model parameters to check how treating the initial learning estimates as fixed parameters may affect the identification of the model parameters. Given the stochastic environment, in order to assess the different initialization methods we focus on averaged MSD statistics, calculated according to (9) across the repeated simulations of the model. Particularly, we analyse both the accuracy of the learning initials, i.e., looking at MSDs based on $\delta_0^\phi = \phi_0 - \hat{\phi}_0$ and $\delta_0^R = R_0 - \hat{R}_0$, and how the different initials affect the accuracy of estimates of other model's parameters, such as the learning gain and the price rigidity parameter; in the latter cases, MSD measures are obtained on the basis of $\delta^\gamma = \gamma - \hat{\gamma}$ and $\delta^\theta = \theta - \hat{\theta}$, respectively.

Pre-determined initials

We begin looking at the MSD of the pre-determined initials, presented in Table 1 for the two different gain values. Under the REE-based method, we calculate the initials based on the true model's parameter values, which are given by $\phi_0^{REE} \simeq 0.44$, and $R_0^{REE} \simeq 5.26$. For the methods based on a training sample, i.e., the OLS and the WLS initials, we first set aside an initial portion of the simulated series, $\{\pi_t, x_t\}_{t=1-P}^0$, and then compute the initials based on (11)-(12). Here the main dimension of interest is the size of the training sample of data, which we vary over $P = \{25, 50, 75, 100, 200\}$. Two further assumptions are required for the WLS case. First, we assume the learning gain is known⁹, and set it according to what was used to generate the artificial series (either $\gamma = 0.02$ or $\gamma = 0.10$). Second, we explore two alternative specifications of the prior R_0 , namely, a diffuse prior with $R_0 = 0$, and a REE prior with $R_0 = R_0^{REE}$. There are some interesting points sprouting from Table 1 results:

1. The REE-based initials are overall the less accurate among the pre-determined initials, especially for the PLM coefficient, ϕ_0 . Also, the performance of the REE-based initials deteriorate substantially for the higher gain calibration. This last result is in agreement with the idea that higher learning gains lead to noisier estimates of agents' PLM parameters, which drive out-of-equilibrium dynamics farther from the REE implied parameter values.
2. Whereas there is little difference between the OLS and the WLS initials under the lower gain calibration, the latter is clearly the best performing method under the higher gain. Also,

⁹Later in our empirical application we relax this assumption and let the WLS initials to be determined jointly with the estimation of the learning gain and other model parameters.

Table 1: Mean-square deviations of pre-determined initials.

Methods	Initials	Training samples (P)				
		25	50	75	100	200
Data generated with $\gamma = 0.02$:						
REE	ϕ_0^{REE}		0.056	
	R_0^{REE}		4.890	
OLS	$\hat{\phi}_0^{OLS}$	0.026	0.006	0.0032	0.002	0.006
	\hat{R}_0^{OLS}	7.155	2.017	1.145	1.195	2.446
WLS.1 (diffuse prior on \hat{R}_{-P})	$\hat{\phi}_0^{WLS.1}$	0.026	0.006	0.002	0.001	0.000
	$\hat{R}_0^{WLS.1}$	11.975	4.331	1.567	0.567	0.010
WLS.2 (REE prior on \hat{R}_{-P})	$\hat{\phi}_0^{WLS.2}$	0.110	0.045	0.018	0.007	0.000
	$\hat{R}_0^{WLS.2}$	1.860	0.668	0.234	0.084	0.002
Data generated with $\gamma = 0.10$:						
REE	ϕ_0^{REE}		0.376	
	R_0^{REE}		17.984	
OLS	$\hat{\phi}_0^{OLS}$	0.017	0.058	0.104	0.142	0.245
	\hat{R}_0^{OLS}	4.529	10.685	13.300	14.507	16.287
WLS.1 (diffuse prior on \hat{R}_{-P})	$\hat{\phi}_0^{WLS.1}$	0.004	0.000	0.000	0.000	0.000
	$\hat{R}_0^{WLS.1}$	0.243	0.001	0.000	0.000	0.000
WLS.2 (REE prior on \hat{R}_{-P})	$\hat{\phi}_0^{WLS.2}$	0.011	0.000	0.000	0.000	0.000
	$\hat{R}_0^{WLS.2}$	0.100	0.001	0.000	0.000	0.000

Statistics averaged over 10,000 simulations of the model.

notice that increasing the size of the training sample always improves the performance of the WLS initials, while the relationship is not monotonic for the OLS initials.

3. Between the two alternative specifications of the WLS prior on the learning coefficients uncertainty, the diffuse prior provides the most accurate initial estimates, though not in terms of the initial for the regressors' variance, where the REE prior obtains a better fit. Hence, the use of a diffuse prior provides an interesting way to speed up the convergence of the learning estimates within the training sample.

Estimated initials

The results for the estimation-based method are presented in Table 2, where we consider a total of four experiments varying which parameters are included in the joint estimation by GMM. First, we assume there is only ϕ_0 to be estimated, so that all the model parameters, the learning gain, and the initial estimate of the variance of x_t , R_0 , are pre-fixed to their true values. In the second experiment, we add the constant learning gain to the estimation problem. Next, we consider the case where a model parameter is estimated jointly with the initial and the learning gain, namely, the price rigidity parameter θ . Finally, we relax the assumption of knowledge of R_0 , and estimate it jointly with all the previous parameters. Another dimension of interest in these exercises is the size of the sample used for estimation. Whereas the estimates of the model parameters are expected to converge to a normal distribution centered around their true values as the sample size increases, due to the consistency of the GMM estimator, for the initial parameters this might not be the case. In order to evaluate whether the sample size has any effect over the estimates we vary it over $T = \{100, 200, 500, 1000, 5000\}$.

Overall, we find that the MSDs obtained under the estimation-based approach are of several orders of magnitude higher than those obtained under the pre-determined initials. For instance, the worst performing pre-determined initial under the higher gain calibration, namely, the REE-based initial, achieved a MSD ten times smaller than the initials estimated with a sample of 100 observations. Focusing on these estimation-based results we also have some interesting observations regarding the dimensions considered:

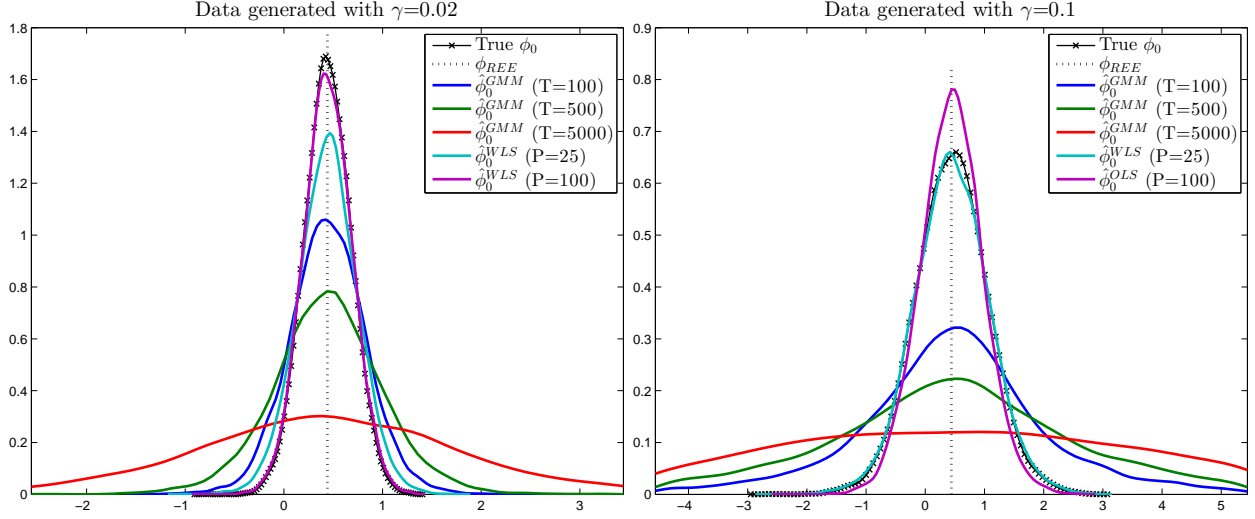
1. Adding more parameters to the joint estimation approach clearly leads to a deterioration of the initials accuracy. Hence, we find that there is a negative spillover effect from the inaccuracy of model parameter estimates into the estimates of the learning initials.
2. Larger estimation samples have a perverse effect on the accuracy of the estimated initials, even though the spillover effect tends to reduce as the precision of the model parameters estimates increases with the sample size. Thus, the estimation of initials induces a problem

Table 2: Mean-square deviations of estimated initials and parameters.

Experiments		Estimation samples (T)				
No.	Parameters	100	200	500	1000	5000
Data generated with $\gamma = 0.02$:						
1	ϕ_0	0.098	0.130	0.274	0.504	2.193
2	ϕ_0	1.499	1.103	1.104	1.384	3.107
	γ	0.009	0.003	0.001	0.001	0.000
3	ϕ_0	5.973	4.609	3.837	4.372	9.288
	γ	0.009	0.003	0.001	0.001	0.000
	θ	0.115	0.099	0.081	0.067	0.023
4	ϕ_0	7.456	7.209	7.854	9.138	13.834
	γ	0.009	0.004	0.001	0.001	0.000
	θ	0.107	0.095	0.081	0.071	0.029
	R_0	523.2	509.4	561.6	597.1	543.8
Data generated with $\gamma = 0.10$:						
1	ϕ_0	3.175	4.045	6.248	8.728	15.213
2	ϕ_0	3.177	4.014	6.246	8.887	15.764
	γ	0.013	0.007	0.003	0.002	0.000
3	ϕ_0	6.822	6.543	8.667	11.980	18.944
	γ	0.013	0.007	0.003	0.002	0.000
	θ	0.113	0.083	0.046	0.027	0.002
4	ϕ_0	8.933	10.175	12.904	15.151	19.497
	γ	0.013	0.008	0.004	0.002	0.001
	θ	0.107	0.083	0.050	0.032	0.005
	R_0	695.3	762.3	755.2	759.5	902.3

Statistics averaged over 10,000 simulations of the model.

Figure 1: Densities of initial estimates over simulations.



Densities estimated using the normal kernel smoothing function over simulations of the model. The coefficient estimates are from experiment 1 (only ϕ_0 estimated). Simulations with exact boundary estimates are removed (see Table 4 in the Appendix).

of too many degrees of freedom in estimation, where larger samples drive the estimation of the initial parameters towards indeterminacy.

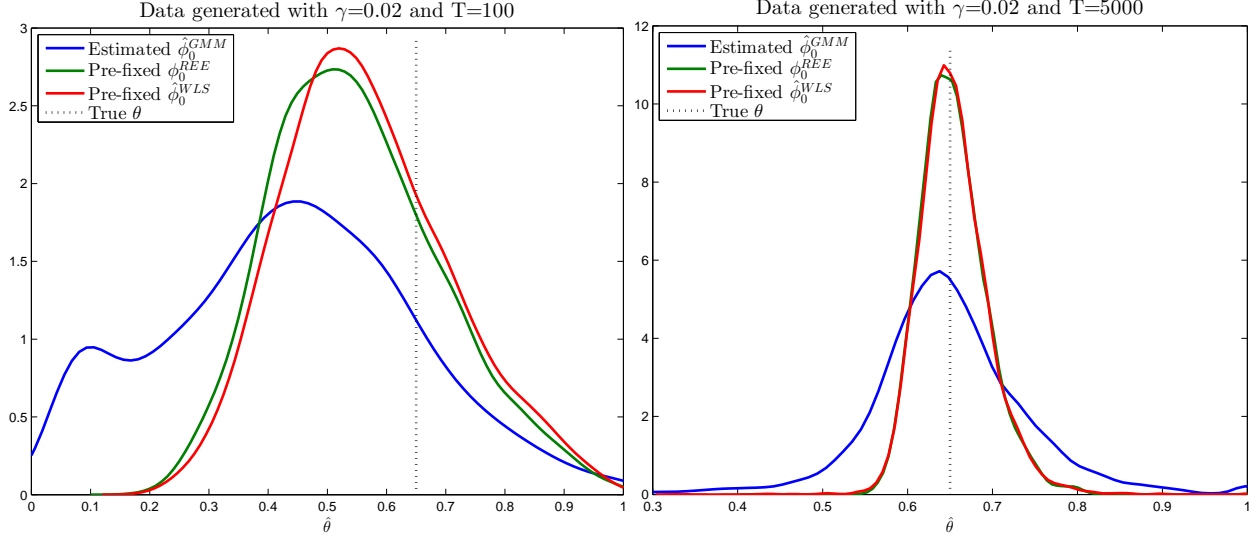
3. Finally, notice that the MSDs of estimated initials are bigger under the higher learning gain calibration, while for the other estimated parameters increasing the learning gain has smaller effects over their accuracy.

Distributions of estimates

To have a clearer understanding of how these initialization approaches compare to each other we can go beyond the averaged statistics and look at the actual distributions of the corresponding estimates. The densities of the initial estimates over the simulations are presented in Figure 1. There we can clearly see that the main problem with the estimated initials relates to their dispersion, which increases with the size of the estimation sample. We also see that, although the true initials were distributed around the REE implied coefficient value, their dispersion increased with the learning gain, which explains the failures of both the REE-based and the OLS-based initialization methods. Also notice that increasing the training sample size brings the distribution of the WLS-based initials very close to the actual initials distribution.

The densities of the estimates of the model structural parameter representing the index of price rigidity, θ , are presented in Figure 2. We look at the estimates obtained under the joint estimation of this model parameter and the learning gain, plus the initial for the estimation-based case, $\hat{\phi}_0^{GMM}$.

Figure 2: Densities of θ estimates over simulations.



Densities estimated using the normal kernel smoothing function over simulations of the model with $\gamma = 0.02$. The θ coefficient estimates are from experiment 3, i.e., ϕ_0 is jointly estimated with γ and θ for the GMM case, and only γ and θ are estimated for the other cases where ϕ_0 is fixed from a training sample estimate. Simulations with exact boundary estimates of ϕ_0 , γ , and θ are removed (see Tables 4 and 5 in Appendix B) leaving 2546 (6943), 3896 (9623), and 3851 (9819) observations for the computation of the densities of $\hat{\theta}$ associated to $\hat{\phi}_0^{GMM}$, $\hat{\phi}_0^{REE}$, and $\hat{\phi}_0^{WLS}$, respectively under $T = 100$ (5000).

Chevillon et al. (2010) have shown that learning induces non-standard distributions in finite sample estimates of model parameters, a finding that is clearly corroborated in the left hand side (l.h.s.) panel of Figure 2. Here we add an interesting observation to this finding; namely, we find that the joint estimation of the learning initials can lead to much stronger deviations from asymptotic distributions. Moreover, notice that even with a larger estimation sample, as in the right hand side (r.h.s.) panel of Figure 2, the θ estimates are still more dispersed under the case where the initial was jointly estimated than under the cases with pre-determined initials.

Learning curves

The statistics on the learning initials refer only to the accuracy of the learning algorithm estimates at the initial point of the estimation sample. Another interesting question is how long does it take for the different initializations to converge to the stationary estimates implied by a long run operation of the learning algorithm within the model under analysis. To answer that question we draw the MSD learning curves, which represent how the accuracy of the estimates evolves through time. These curves are presented in Figure 3, where we focus on some cases of interest: under the lower gain calibration, in the top panels, we focus on the transitional dynamics of the WLS

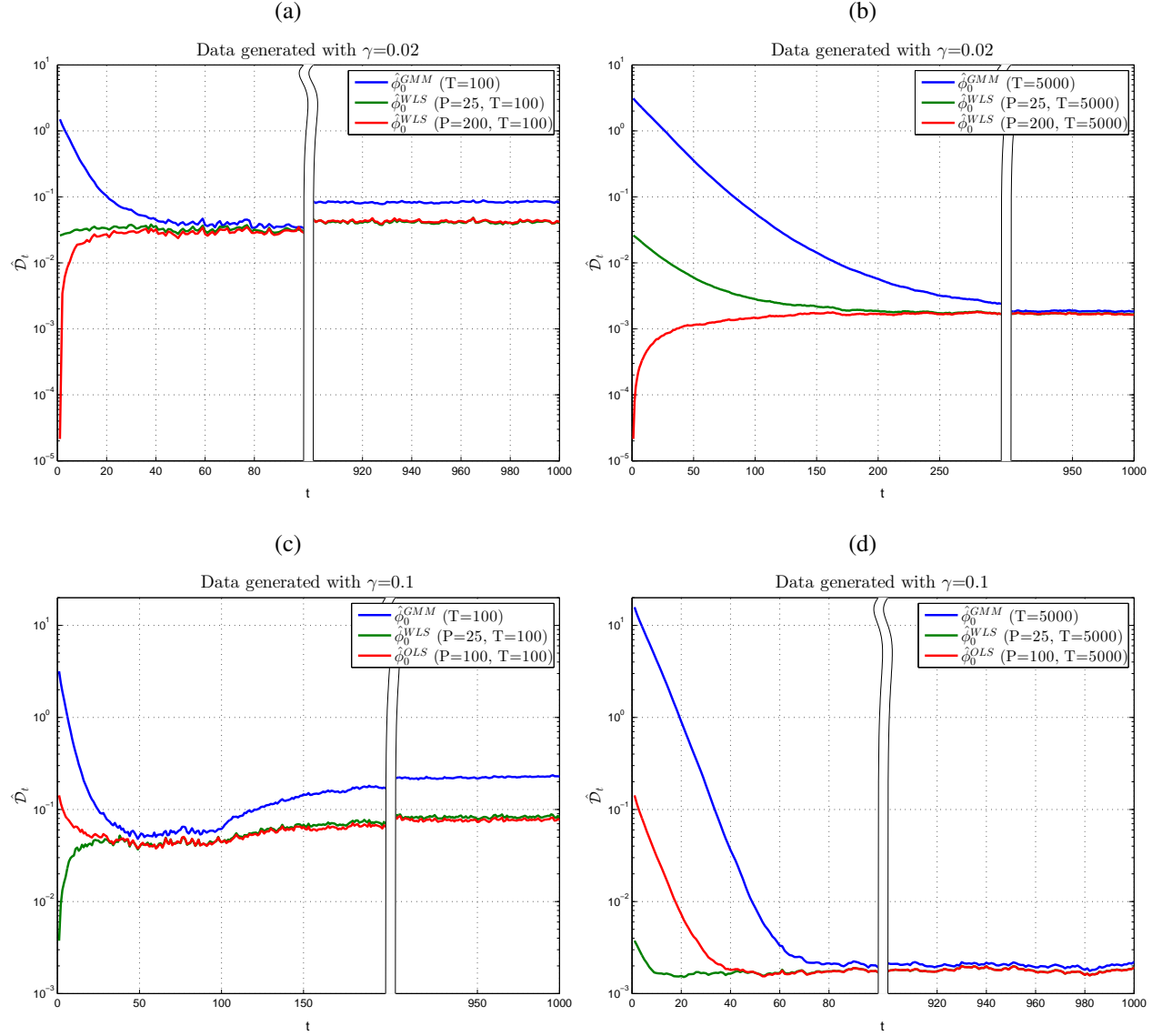
and the GMM estimates of the initial, since there was little difference between the WLS and the OLS methods; the OLS case is illustrated in the bottom panels for the higher gain calibration. The difference between the l.h.s. panels and those in the r.h.s. is only about the size of the sample used for estimation of the learning gain, and the initials in the $\hat{\phi}_0^{GMM}$ case: 100 observations in the former, and 5000 in the latter. There are three important observations from these results:

1. The inaccuracy associated with every initial learning estimate does tend to vanish as the time goes on, though the time to convergence may vary depending on the magnitude of the initial error and the estimate of the learning gain. This is more clearly evident in panel (d), where convergence is achieved in about 20, 40, and 70 periods after the WLS, the OLS, and the GMM-based initials, respectively. As expected, a longer convergence time is required under the lower gain calibration, as presented in panel (b), although the bias in the learning gain estimates seem to be more pronounced under this case, which leads to the next points.
2. Fixing the learning initials according to the pre-determined methods leads to smaller biases in the estimation of the learning gain with the smaller sample. This is evident from the long run MSD levels achieved in the l.h.s. panels, which are higher under the estimated initials. As the estimation sample increases, consistent with our previous results, the learning gain estimation errors diminish and there is not much difference between the long run MSD levels observed in the r.h.s. panels.
3. Even fixing the initials with pre-determined methods, the estimation of the learning gain is still biased and, in contrast with the biases in the learning initials, the learning gain inaccuracy will persist to affect the long run distribution of learning estimates. This is evident from the learning curves by: first, comparing the different long run MSD levels implied by the use of only different estimation samples, panel (a) versus panel (b), and panel (c) versus panel (d); second, noticing that, in some cases, the WLS initials (here using the true gain within the training sample) achieve MSDs smaller than what eventually becomes the long run MSDs levels under the estimated learning gains.

4.4 Discussion

Our simulation results can be summarized as follows. First, we found that initialization methods based on a training sample of data provide more accurate initial estimates than joint estimation approaches. Second, the joint estimation of the initials can affect the distribution of other model parameter estimates, and these distortions can persist in large samples too. Finally, among the pre-determined initialization methods we have found important evidence in favor of the use of a WLS-based approach over a training sample, particularly assuming a diffuse prior. In order

Figure 3: MSD learning curves of $\hat{\phi}_t$.



Statistics averaged over 10,000 simulations of the model. The estimates come from experiment 2, where ϕ_0 is jointly estimated with γ for the GMM case, and only γ is estimated for the other cases where ϕ_0 is fixed from a pre-sample estimate.

to obtain a relative assessment of the initialization methods under analysis, we now relate these results to two principles that we consider of relevance for applied adaptive learning research: (i) the initials COHERENCE to the learning process; and, (ii) the initials SUSCEPTIBILITY to bias the model’s explanatory power and the estimation of its parameters.

In empirical settings, a proper initialization of the learning algorithm requires to find out what were agents’ beliefs at the beginning of the sample of data. To achieve this goal it is important to understand the statistical properties of the learning process we are trying to mimic. Recursive estimation algorithms are statistically characterized by two main distinct phases: a transient phase, where the estimates are so far apart from the true parameter values that the upcoming sequence of updates can easily achieve substantial improvements to the accuracy of the estimates; and a stationary phase, where most of the updates to the estimates are essentially just tracking tiny disturbances that may affect the system under estimation. Hence, if the initial beliefs should reflect the continuation of an estimation process that was already in motion prior to the sample beginning, an initialization method will satisfy the COHERENCE criterion when it can deliver estimates as close as possible to the algorithm’s long run operation¹⁰. Under our evaluation measure this corresponds to a minimization of the initials’ MSD.

Another empirical issue is how much can the learning initials affect the accuracy of the estimates of other model parameters. Under standard likelihood-based estimation approaches, every data point is given the same weight on the estimation of a structural parameter that is assumed to be constant throughout the sample period. Under learning this weighting profile can be easily manipulated by tweaking the initial learning estimates so as to induce a transient phase in the portion of the sample that follows the initialization, which potentially increases the model’s explanatory power. In the context of our evaluation exercises, such SUSCEPTIBILITY to biases was measured by looking at the MSDs of some key parameter estimates across the different initialization methods.

These principles may inherently generate a trade-off for the assessment of the initialization methods: one can always give up some degree of the COHERENCE delivered by a learning initial in exchange for some SUSCEPTIBILITY to tweak that initial in order to improve the model fit to the data. In spite of this trade-off, our results were overall favorable to the training sample-based methods, particularly the WLS with diffuse priors; whereas the coherence of the WLS method is due to its closer resemblance to the actual learning mechanism, especially for assuming a coherent sequence of gains, the diffuse priors proved useful to accelerate convergence within small training samples. Because the implied initials are then pre-determined for the model estimation, the method also minimizes the susceptibility to the introduction of biases through the learning initials.

¹⁰Admittedly, one may also be interested in obtaining initials that reflect the transient phase that follows the occurrence of a large shock that shifted agents’ beliefs away from equilibrium just before the initialization data point.

5 Empirical Application

5.1 Hybrid Phillips curve model

In order to evaluate the relevance of using different initialization methods, we now pursue an empirical application augmenting our baseline model and estimating it with US macroeconomic data. We follow Gali and Gertler (1999) and estimate a hybrid NKPC model given by

$$\pi_t = \psi_f \pi_{t+1}^e + \psi_b \pi_{t-1} + \delta x_t + \eta_t, \quad (22)$$

$$x_t = \rho x_{t-1} + \nu_t, \quad (23)$$

with

$$\psi_f = \beta \theta \zeta^{-1}, \quad \psi_b = \omega \zeta^{-1}, \quad \delta = (1 - \omega)(1 - \theta)(1 - \beta \theta) \zeta^{-1}, \quad (24)$$

$$\zeta = \theta + \omega(1 - \theta(1 - \beta)), \quad (25)$$

where ω is the fraction of firms that set their prices according to a backward looking rule of thumb, and the remaining parameters have the same interpretation as in the baseline specification.

The REE is given by the values of a , b , and c that solve the following equalities

$$a = \psi_f a(1 + b), \quad b = \psi_f b^2 + \psi_b, \quad c = \psi_f bc + \psi_f \rho c + \delta. \quad (26)$$

There are multiple solutions to this system, which are characterized in the Appendix A.3 together with E-stability conditions. Under adaptive learning, agents form expectations using estimates of these parameters obtained according to the LS algorithm of (4)-(5), where $\hat{\phi}_t \equiv (\hat{a}_t, \hat{b}_t, \hat{c}_t)'$, $\mathbf{x}_t \equiv (1, \pi_{t-1}, x_t)'$ and $y_t \equiv \pi_t$.

5.2 Data and estimation approach

We use quarterly US data covering the period from 1947q1 to 2007q4. To measure inflation we use CPI inflation, whereas for the forcing variable, x_t , we use non-farm business sector labor shares. To remove trends in the latter we obtain gap measures using the Hodrick-Prescott (HP) filter. For simplicity, we are neglecting real-time data issues by focusing on a unique snapshot of the realization of these series. All our data series are obtained from the FRED database of the St. Louis Fed.

We adopt a GMM estimation approach similar to that used in our simulation analysis. Here, the set of parameters determining inflation is given by $\Theta = (\beta, \theta, \omega, \rho, \gamma, \Phi_0, \mathbf{R}_0)$. To facilitate estimation we fix some of these parameters to values we consider reasonable: $\beta = 0.99$, $\rho =$

0.75. As instruments we use a constant plus four lags of inflation, the labor income share gap, the HP-filtered output gap on real GDP, and a long-short interest rate spread given by the difference between the 10-year Treasury Bill rate and the Federal Funds rate. We also include an unrestricted constant in the model specification.

In order to evaluate the training sample-based initializations, we set 100 initial observations aside and obtain initials with a varying number of training sample observations. To keep the estimates comparable, we fix the estimation sample accordingly; hence, our estimation sample starts from 1972q2. For the case of WLS-based initials, we use a diffuse prior and the initials are re-calculated within the estimation routine in order to take into account the estimates of the learning gain. This is in contrast with the approach we followed in our simulations above, where the gain was fixed to the true value when calculating the WLS initials; but, since empirically we do not know the true learning gain value, it seems more reasonable to include the WLS initialization in the estimation routine.

We also benchmark our estimates of the model under learning with corresponding estimates under RE. For that purpose we follow the standard approach of replacing the expectation term in (22) by actual observations of next period inflation. Imposing the same identifying assumption we used under learning, which in the present context is given by $E_{t-1}[\eta_t] = 0$, leads to moment conditions of the form of (20), except that the residual term now also includes an one-step-ahead inflation forecast error (see Mavroeidis et al., 2014, pp. 133-4).

5.3 Results

The results obtained with the GMM estimation of the hybrid NKPC model are presented in Table 3. The RE benchmark estimates seem to be in accordance to previous estimates in the literature (e.g., Gali and Gertler, 1999) where the Calvo's index of price stickiness, θ , is in general found to be smaller than the fraction of backward looking price setters, ω . Also notice that the RE estimates are statistically different from zero at standard levels of significance.

Most important for our purposes are the estimates obtained under learning, particularly comparing the differences due to changes in the initialization method. Overall, we observe that the introduction of learning reduces the estimates of θ . This result is consistent with previous findings obtained under fully-fledged DSGE models, which point that the introduction of learning reduces the relevance of intrinsic sources of macroeconomic persistence (e.g., Milani, 2007). One interesting difference relates to our estimates of the learning gain, where previous estimates in the literature are usually confined to lie below the value of 0.10. Hence, our results put into question too restrictive prior assumptions on the learning gain.

Table 3: Empirical estimates of US NKPC - 1972q2-2007q4.

Estimation Exercises	Parameter Estimates					
	θ	ω	γ	a_0	b_0	c_0
Rational expectations	0.911 (0.086)	0.106 (0.052)	-	-	-	-
Learning (estimated initials)	0.643 (0.066)	0.000 (0.187)	0.296 (0.060)	7.340 (2.640)	-1.585 (0.358)	2.797 (1.085)
Learning (pre-determined initials)						
- WLS-25 (1966q1-1972q1)	0.775 (0.055)	0.298 (0.059)	0.268 (0.037)	0.658	0.289	0.132
- WLS-50 (1959q4-1972q1)	0.775 (0.055)	0.298 (0.059)	0.268 (0.037)	0.657	0.290	0.132
- WLS-75 (1953q3-1972q1)	0.775 (0.055)	0.298 (0.059)	0.268 (0.037)	0.657	0.290	0.132
- WLS-99 (1947q3-1972q1)	0.775 (0.055)	0.298 (0.059)	0.268 (0.037)	0.657	0.290	0.132
- OLS-25 (1966q1-1972q1)	0.787 (0.057)	0.290 (0.062)	0.254 (0.039)	0.661	0.372	0.121
- OLS-50 (1959q4-1972q1)	0.817 (0.071)	0.306 (0.062)	0.240 (0.040)	0.166	0.764	0.032
- OLS-75 (1953q3-1972q1)	0.822 (0.075)	0.307 (0.062)	0.238 (0.040)	0.142	0.761	0.019
- OLS-99 (1947q3-1972q1)	0.764 (0.052)	0.304 (0.058)	0.297 (0.037)	0.224	0.642	0.005

Parameters estimated by GMM. For the cases with pre-determined initials, the initial learning coefficients are obtained from training samples of size indicated in the methods names. Values in parentheses are standard errors of the estimates and are computed based on numerical approximations of the objective function first derivatives. The standard errors under learning should be interpreted with caution since the estimators distribution, and corresponding test statistics, can become non-standard (see Chevillon et al., 2010).

Nevertheless, our results also point to an important sensitivity on the estimates of the ω parameter to the chosen method of initialization. Namely, we find that when the initials are jointly estimated with the other structural parameters, $\omega = 0$. A different picture emerges when the initials are pre-determined using a training sample: overall we find that the estimated fraction of firms using a backward looking price setting rule increases in relation to the benchmark estimates under RE. That is a very important difference since the former estimate implies strong evidence against the hybrid specification of the NKPC.

In light of what was found in our simulation analysis, we interpret these results as providing further evidence against the estimation-based approach to the initialization of learning. It is also clear from the initial estimates presented in Table 3 that the problem with such an estimation approach is due to a tendency to overfitting in the after-initials portion of the estimation sample. Namely, the learning initials under the joint estimation approach are clearly unreasonable estimates to represent a behavioral model of forecasts for inflation.

Among the pre-determined initialization methods we find only small variations when comparing the structural parameters estimates associated to the WLS and the OLS-based methods. Consistent with our simulation results, we also see that the OLS-based initials are less robust to the use of different training sample sizes. The WLS estimates appear to have achieved convergence already within the smaller training sample considered, 25 observations, though this result is likely to be driven by the higher gain estimates.

6 Concluding remarks

In this paper we provided a critical review on the several methods previously proposed in the literature of learning and expectations in macroeconomics in order to initialize its learning algorithms. Most importantly, we have also provided one of the first attempts in the literature to evaluate how these methods compare to each other, and how their performance may be evaluated with respect to their learning and expectations rationale. Throughout the paper, most of our analysis is carried in the context of forward looking Phillips curve models.

We proposed a taxonomy of initialization methods that can be broadly defined in three major classes: equilibrium-related methods, training sample-based methods, and estimation-based methods. We conducted extensive simulation exercises comparing different initialization methods that can be conceived within this classification. We also evaluated the relevance of the differences between these methods in an empirical application on the determination of US inflation rates under the Phillips curve framework.

Our analysis led us to draw the following recommendations. First, though equilibrium-related initialization methods seem to provide rather conservative initials, they are often incoherent with

the dynamics implied by learning, particularly under high learning gains. Second, among the training sample-based methods, the use of standard OLS estimates can also turn out to provide incoherent estimates since it does not take into account the particular specification of the learning gain. Direct application of the learning algorithm into the training sample, the WLS-based method in our terminology, was overall favored by most of our evaluation criteria. Particularly, we found that a diffuse specification of this method leads to an accelerated convergence, facilitating the feasibility of the method in macroeconomic contexts.

Finally, and foremost, we strongly discourage the approach of joint estimation of learning initials, since these will tend to be biased due to overfitting in the initial portion of the estimation sample. Moreover, we found evidence of spillover effects from the biases introduced by estimation of the initials into the estimates of the model's structural parameters. In our empirical application, these distortions were found to be strong enough to overturn evidence that is generally in favor of the hybrid specification of the New Keynesian Phillips curve.

A Derivations

A.1 MSPE implied by initialization error

From (2) we have that $y_{t+1} = \beta\phi_t + u_{t+1}$ and $\hat{y}_{t+1} = \beta\hat{\phi}_t$, so that the prediction error is given by

$$\begin{aligned}\hat{\Delta}_{t+1} &= y_{t+1} - \hat{y}_{t+1}, \\ &= \beta(\phi_t - \hat{\phi}_t) + u_{t+1}.\end{aligned}\tag{27}$$

Defining $\hat{\delta}_t = \phi_t - \hat{\phi}_t$, from (6) and (7) we find that

$$\begin{aligned}\hat{\delta}_t &= \phi_{t-1} + \gamma(y_t - \phi_{t-1}) - \hat{\phi}_{t-1} - \gamma(y_t - \hat{\phi}_{t-1}), \\ &= (1 - \gamma)(\phi_{t-1} - \hat{\phi}_{t-1}), \\ &= (1 - \gamma)\hat{\delta}_{t-1},\end{aligned}$$

which can be solved recursively to result in $\hat{\delta}_t = (1 - \gamma)^t \hat{\delta}_0$. Substituting this back into (27) and taking the expectation of the squared value results in

$$\begin{aligned}\hat{\Delta}_{t+1} &= \beta(1 - \gamma)^t \hat{\delta}_0 + u_{t+1}, \\ E[\hat{\Delta}_{t+1}^2] &= \beta^2(1 - \gamma)^{2t} \hat{\delta}_0^2 + \sigma_u^2.\end{aligned}$$

For the case where β is unknown, we defined $\hat{\delta}_t = \beta\phi_t - \hat{\beta}\hat{\phi}_t$, which using (6) and (7) results in

$$\hat{\delta}_t = (1 - \gamma) \hat{\delta}_{t-1} + \gamma (\beta - \hat{\beta}) y_t.$$

Solving this recursively yields

$$\hat{\delta}_t = (1 - \gamma)^t \hat{\delta}_0 + \gamma (\beta - \hat{\beta}) \sum_{i=0}^{t-1} (1 - \gamma)^i y_{t-i}.$$

A.2 Long run variance of learning estimates

Substituting (2) into (6) we obtain

$$\phi_t = \lambda\phi_{t-1} + \gamma u_t,$$

where $\lambda = 1 - \gamma(1 - \beta)$. This recursion is equivalent to

$$\phi_t = \lambda^t \phi_0 + \gamma \sum_{i=0}^{t-1} \lambda^i u_{t-i}, \quad (28)$$

In the limit, as $t \rightarrow \infty$, $E[\phi_\infty] = 0$ as long as $|\lambda| < 1$, or $\gamma - 2/\gamma < \beta < 1$. Hence, the long run variance of ϕ_t , denoted by $\bar{\sigma}_\phi^2$, is given by

$$\begin{aligned} \bar{\sigma}_\phi^2 &= \lim_{t \rightarrow \infty} E[\phi_t^2], \\ &= \lim_{t \rightarrow \infty} E \left[\lambda^{2t} \phi_0^2 + 2\lambda^t \phi_0 \gamma \sum_{i=0}^{t-1} \lambda^i u_{t-i} + \left(\gamma \sum_{i=0}^{t-1} \lambda^i u_{t-i} \right)^2 \right], \end{aligned}$$

which, because u_t is assumed to be serially independent, simplifies to

$$\begin{aligned} \bar{\sigma}_\phi^2 &= \lim_{t \rightarrow \infty} \lambda^{2t} \phi_0^2 + \gamma^2 \sum_{i=0}^{t-1} \lambda^{2i} \sigma_u^2, \\ &= \lim_{t \rightarrow \infty} \lambda^{2t} \phi_0^2 + \gamma^2 \sigma_u^2 \frac{1 - \lambda^{2t}}{1 - \lambda^2}, \\ &= \frac{\gamma^2 \sigma_u^2}{1 - \lambda^2}, \end{aligned}$$

where the limit is solved under the assumption that $|\lambda| < 1$. Notice that $\partial \bar{\sigma}_\phi^2 / \partial \gamma > 0$, i.e., as the gain increases the dispersion of the learning estimates tends to increase as well.

A.3 Hybrid Phillips curve REE

There are six solutions to the RE conditions in (26). Starting with b , there are two possible solutions given by $b^\pm = \frac{1 \pm \sqrt{1 - 4\psi_f\psi_b}}{2\psi_f}$. For a there are three possibilities: $a = 0$ or a is indeterminate with $\beta = 1$ and $\omega \neq 1$, or $\omega = 1$ and $\beta \neq 1$. Finally, c is uniquely determined by b . Putting these combinations together we have the following RE solutions:

$$RE.1 = \{a = 0, b^+, c^+\}; RE.2 = \{a = 0, b^-, c^-\};$$

$$RE.3 = \{a = any, b^+, c^+, \beta = 1, \omega \neq 1\}; RE.4 = \{a = any, b^+, c^+, \omega = 1, \beta \neq 1\};$$

$$RE.5 = \{a = any, b^-, c^-, \beta = 1, \omega \neq 1\}; RE.6 = \{a = any, b^-, c^-, \omega = 1, \beta \neq 1\}.$$

To check for E-stability of these solutions we first define the T -mapping associated to this model:

$$T \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{bmatrix} \psi_f a (1 + b) \\ \psi_f b^2 + \psi_b \\ \psi_f bc + \psi_f c\rho + \delta \end{bmatrix}.$$

E-stability requires that the eigenvalues of the Jacobian matrix of T , evaluated at the given RE solution, are smaller than unity. These eigenvalues depend only on the value of b and are given by

$$\{(1 + b) \psi_f, 2b\psi_f, (b + \rho) \psi_f\}.$$

Focusing on the range of reasonable parameter values, $0 < \beta \leq 1$, $0 \leq \theta \leq 1$, $0 \leq \omega \leq 1$, and $-1 < \rho < 1$, we find that only the RE solutions with b^- can be E-stable but with some additional restrictions on the structural parameters: if $\beta = 1$ then $\theta \neq 1$ and $\theta < \omega$; if $\beta \neq 1$ then $\omega \neq 0$ if $\theta = 0$.

B GMM numerical estimation

A numerical optimization routine is used to find the values of $\hat{\Theta}$ that minimize the GMM objective function, (21). For that purpose we adopt a sequential quadratic programming algorithm, namely the 'sqp' option in the fmincon function in Matlab optimization toolbox. The convergence criterion for the iterative estimation of the weighting matrix is based on the Euclidean distance between the successive parameters estimates, i.e., $\|\hat{\Theta}^{(i)} - \hat{\Theta}^{(i-1)}\| < \epsilon$. In our simulations we set $\epsilon = 10^{-6}$, for which convergence is achieved in about 10 (15) iterations, on average, under $\gamma = 0.02$ ($\gamma = 0.10$).

Whereas the model parameters are reasonably constrained by theory implied boundaries, the parameters associated to the learning algorithm require artificial constraints to avoid numerical

Table 4: Number of simulations where estimates hit lower/upper bounds.

Estimation	Estimation Experiments				
sample	1	2	3	4	
(T)	$\hat{\phi}_0$	$\hat{\phi}_0$	$\hat{\phi}_0$	$\hat{\phi}_0$	\hat{R}_0
Data generated with $\gamma = 0.02$:					
100	0	295	1676	1937	2986
200	0	187	1156	1890	2767
500	1	128	729	2179	2794
1000	1	154	824	2685	2682
5000	65	344	2053	4238	2066
Data generated with $\gamma = 0.10$:					
100	423	629	1818	2298	4107
200	563	733	1497	2780	4456
500	1101	1221	1878	3899	4019
1000	1873	1964	3021	4776	3576
5000	4454	4714	6223	6346	3595

Statistics based on 10,000 simulations of the Phillips curve model under learning.

instabilities during estimation. Our experimental analysis led us to adopt the following constraints: $0 \leq \gamma \leq 0.5$, $\phi^{REE} - 5 \leq \hat{\phi}_0 \leq \phi^{REE} + 5$, and $0 < \hat{R}_0 \leq 50$. Although these constraints were never violated in the artificial data ($\sup |\phi_0| = 1.2893$ (2.7735), and $\sup R_0 = 24.11$ (49.08), across the simulations for $\gamma = 0.02$ (0.10)), the numerical estimation of $\hat{\phi}_0$ and \hat{R}_0 often resulted in boundary solutions. These cases are summarized in Table 4, where we observe that the joint estimation of R_0 tends to increase the number of boundary solutions, and increasing the number of estimated parameters also hinders the search for interior solutions. These effects are also amplified when the data true learning gain increases. Similar statistics for the cases of jointly estimated model parameters are reported in Table 5.

Table 5: Number of simulations where model estimates hit lower/upper bounds.

Estimation sample	Estimation Experiments									
	2		3		4					
(T)	$\hat{\gamma}_{\phi_0}^{GMM}$	$\hat{\gamma}_{\phi_0}^{REE}$	$\hat{\gamma}_{\phi_0}^{GMM}$	$\hat{\gamma}_{\phi_0}^{REE}$	$\hat{\theta}_{\phi_0}^{GMM}$	$\hat{\theta}_{\phi_0}^{REE}$	$\hat{\gamma}_{\phi_0}^{GMM}$	$\hat{\gamma}_{\phi_0}^{REE}$	$\hat{\theta}_{\phi_0}^{GMM}$	$\hat{\theta}_{\phi_0}^{REE}$
Data generated with $\gamma = 0.02$										
100	4910	1208	3682	2855	2348	3627	3492	2860	2249	3641
200	4108	661	3219	2183	2617	3087	3042	2188	2457	3090
500	2497	394	2002	1348	2879	2187	1964	1352	2802	2215
1000	1212	217	1044	843	2690	1265	1049	844	2714	1269
5000	8	47	107	354	1016	41	114	353	1289	41
Data generated with $\gamma = 0.10$										
100	2606	522	1880	1242	3086	3673	1684	1248	3049	3679
200	801	139	550	362	3160	2838	487	370	3155	2850
500	33	12	17	26	2306	1317	16	28	2423	1316
1000	0	0	2	3	1332	493	1	3	1638	490
5000	0	0	0	0	15	0	0	0	122	0

Statistics based on 10,000 simulations of the Phillips curve model under learning. The model parameters are estimated either jointly with the learning initial ($\hat{\phi}_0^{GMM}$), or by fixing the initial to its REE value (ϕ_0^{REE}).

References

- Barucci, E. and L. Landi (1997). Least mean squares learning in self-referential linear stochastic models. *Economics Letters* 57(3), 313–317.
- Berardi, M. and J. K. Galimberti (2012). On the initialization of adaptive learning algorithms: A review of methods and a new smoothing-based routine. Centre for Growth and Business Cycle Research Discussion Paper Series 175, Economics, The University of Manchester.
- Berardi, M. and J. K. Galimberti (2013). A note on exact correspondences between adaptive learning algorithms and the kalman filter. *Economics Letters* 118(1), 139–142.
- Bray, M. M. and N. E. Savin (1986). Rational expectations equilibria, learning, and model specification. *Econometrica* 54(5), 1129–1160.
- Bullard, J. and S. Eusepi (2005). Did the great inflation occur despite policymaker commitment to a taylor rule? *Review of Economic Dynamics* 8(2), 324 – 359.
- Carboni, G. and M. Ellison (2009). The great inflation and the greenbook. *Journal of Monetary Economics* 56(6), 831 – 841.
- Carceles-Poveda, E. and C. Giannitsarou (2007). Adaptive learning in practice. *Journal of Economic Dynamics and Control* 31(8), 2659–2697.
- Carceles-Poveda, E. and C. Giannitsarou (2008). Asset pricing with adaptive learning. *Review of Economic Dynamics* 11(3), 629 – 651.
- Chevillon, G., M. Massmann, and S. Mavroeidis (2010). Inference in models with adaptive learning. *Journal of Monetary Economics* 57(3), 341–351.
- Eusepi, S. and B. Preston (2011, October). Expectations, learning, and business cycle fluctuations. *American Economic Review* 101(6), 2844–72.
- Evans, G. W. and S. Honkapohja (1998). Stochastic gradient learning in the cobweb model. *Economics Letters* 61(3), 333–337.
- Evans, G. W. and S. Honkapohja (2001). *Learning and expectations in macroeconomics*. Frontiers of Economic Research. Princeton, NJ: Princeton University Press.
- Gali, J. and M. Gertler (1999, October). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics, Elsevier* 44(2), 195–222.

- Gaus, E. and S. Ramamurthy (2014, August). Estimation of constant gain learning models. Working Papers 12-01, Ursinus College, Department of Economics.
- Hall, A. R. (2005). *Generalized Method of Moments*. Advanced Texts in Econometrics. Oxford University Press.
- Huang, K. X., Z. Liu, and T. Zha (2009). Learning, adaptive expectations and technology shocks. *The Economic Journal* 119(536), 377–405.
- Ljung, L. and T. Soderstrom (1983). *Theory and Practice of Recursive Identification*. The MIT Press.
- Lubik, T. A. and C. Matthes (2014, January). Indeterminacy and learning: An analysis of monetary policy in the great inflation. Working Paper 14-2, Federal Reserve Bank of Richmond.
- Marcet, A. and J. P. Nicolini (2003). Recurrent hyperinflations and learning. *American Economic Review* 93(5), 1476–1498.
- Mavroeidis, S., M. Plagborg-Møller, and J. H. Stock (2014). Empirical evidence on inflation expectations in the new keynesian phillips curve. *Journal of Economic Literature* 52(1), 124–88.
- Milani, F. (2007, October). Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics* 54(7), 2065–2082.
- Milani, F. (2008). Learning, monetary policy rules, and macroeconomic stability. *Journal of Economic Dynamics and Control* 32(10), 3148 – 3165.
- Milani, F. (2011). Expectation shocks and learning as drivers of the business cycle. *The Economic Journal* 121(552), 379–401.
- Milani, F. (2014). Learning and time-varying macroeconomic volatility. *Journal of Economic Dynamics and Control* 47(0), 94 – 114.
- Newey, W. K. and K. D. West (1987, May). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–08.
- Ormeño, A. and K. Molnár (2015, June). Using survey data of inflation expectations in the estimation of learning and rational expectations models. *Journal of Money, Credit and Banking* 47(4), 673–699.

- Orphanides, A. and J. C. Williams (2005a, November). The decline of activist stabilization policy: Natural rate misperceptions, learning, and expectations. *Journal of Economic Dynamics and Control* 29(11), 1927–1950.
- Orphanides, A. and J. C. Williams (2005b). Inflation scares and forecast-based monetary policy. *Review of Economic Dynamics* 8(2), 498 – 527.
- Pfajfar, D. and E. Santoro (2010). Heterogeneity, learning and information stickiness in inflation expectations. *Journal of Economic Behavior & Organization* 75(3), 426–444.
- Primiceri, G. E. (2006). Why inflation rose and fell: Policy-makers’ beliefs and u. s. postwar stabilization policy. *The Quarterly Journal of Economics* 121(3), 867–901.
- Sargent, T., N. Williams, and T. Zha (2006). Shocks and government beliefs: The rise and fall of american inflation. *American Economic Review* 96(4), 1193–1224.
- Sargent, T. J. (1999). *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.
- Slobodyan, S. and R. Wouters (2012). Learning in an estimated medium-scale dsge model. *Journal of Economic Dynamics and Control* 36(1), 26 – 46.
- Williams, N. (2003, January). Adaptive learning and business cycles. Mimeo.