Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Initial Beliefs Uncertainty

CAMA Working Paper 68/2021
July 2021

Jaqueson K. Galimberti
Auckland University of Technology, School of Economics
KOF Swiss Economic Institute
Centre for Applied Macroeconomic Analysis, ANU

Abstract

This paper evaluates how initial beliefs uncertainty can affect data weighting and the estimation of models with adaptive learning. One key finding is that misspecification of initial beliefs uncertainty, particularly with the common approach of artificially inflating initials uncertainty to accelerate convergence of estimates, generates time-varying profiles of weights given to past observations in what should otherwise follow a fixed profile of decaying weights. The effect of this misspecification, denoted as diffuse initials, is shown to distort the estimation and interpretation of learning in finite samples. Simulations of a forward-looking Phillips curve model indicate that (i) diffuse initials lead to downward biased estimates of expectations relevance in the determination of actual inflation, and (ii) these biases spill over to estimates of inflation responsiveness to output gaps. An empirical application with U.S. data shows the relevance of these effects for the determination of expectational stability over decadal subsamples of data. The use of diffuse initials is also found to lead to downward biased estimates of learning gains, both estimated from an aggregate representative model and estimated to match individual expectations from survey expectations data.

Keywords

expectations, adaptive learning, bounded rationality, macroeconomics

JEL Classification

E70, D83, D84, E37, C32, C63

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

The Centre for Applied Macroeconomic Analysis in the Crawford School of Public Policy has been established to build strong links between professional macroeconomists. It provides a forum for quality macroeconomic research and discussion of policy issues between academia, government and the private sector.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

Initial Beliefs Uncertainty

Jaqueson K. Galimberti*

Abstract

This paper evaluates how initial beliefs uncertainty can affect data weighting and the estimation of models with adaptive learning. One key finding is that misspecification of initial beliefs uncertainty, particularly with the common approach of artificially inflating initials uncertainty to accelerate convergence of estimates, generates time-varying profiles of weights given to past observations in what should otherwise follow a fixed profile of decaying weights. The effect of this misspecification, denoted as diffuse initials, is shown to distort the estimation and interpretation of learning in finite samples. Simulations of a forward-looking Phillips curve model indicate that (i) diffuse initials lead to downward biased estimates of expectations relevance in the determination of actual inflation, and (ii) these biases spill over to estimates of inflation responsiveness to output gaps. An empirical application with U.S. data shows the relevance of these effects for the determination of expectational stability over decadal subsamples of data. The use of diffuse initials is also found to lead to downward biased estimates of learning gains, both estimated from an aggregate representative model and estimated to match individual expectations from survey expectations data.

Keywords: expectations, adaptive learning, bounded rationality, macroeconomics.
JEL codes: E70, D83, D84, E37, C32, C63.

"The longer you can look back,
the farther you can look forward."
–Winston Churchill

1 Introduction

Adaptive learning can generate out-of-equilibrium expectations that help explain deviations from rational expectations and an economy's transitional dynamics towards equilibrium. Under adaptive learning, agents' beliefs are modelled through the assumption of a recursive learning

*Auckland University of Technology, School of Economics, Private mail bag 92006, Auckland 1142, New Zealand; Centre for Applied Macroeconomic Analysis, Australian National University; and KOF Swiss Economic Institute, ETH Zurich. Corresponding author: jaqueson.galimberti@aut.ac.nz, https://sites.google.com/site/jkgeconoeng/, phone: +64 9 921 9999 ext. 5087.

mechanism that updates agents' perceptions about the economy as new data observations become available. The weights given to these observations are key determinants of the degree of persistence introduced by adaptive learning in the evolution of expectations, and, hence, are important factors in the explanation of economic dynamics. Due to the recursive nature of learning, initial beliefs and an estimate of agents' uncertainty about those beliefs need to be specified and can account for some of the explanatory power of learning. In this paper I study the implications of alternative specifications of initial beliefs uncertainty regarding how new information is weighted relative to assumed initial beliefs, and the effects of these assumptions on the estimation of models with adaptive learning.

The main contribution of this paper is to show that misspecification of initials uncertainty can distort the weights given to observations in the assumed process of expectations formation of agents, and that these distortions lead to significant biases in the estimation of models with learning. There are two main components of learning involved here: initials uncertainty and data weighting. From a model estimation point of view, "initials uncertainty" is a parameter determining the confidence agents have, or are assumed to have, on their beliefs at the beginning of the estimation sample. Such initial beliefs, nevertheless, are only initial to the extent that they refer to the beginning of the econometrician's estimation sample – initial beliefs are otherwise assumed to represent the continuation of a learning process that was ongoing prior to the start of the sample, and a similar interpretation is given to the uncertainty surrounding these initial beliefs. The weighting of data under learning is regulated, or so it is often thought, by a sequence of so-called learning gains, and the specification of this sequence can be associated with different behavioural rationales for how agents process information in their expectations. Intriguingly, I find that the weighting ascribed to new data information into agents' beliefs can be substantially affected by the uncertainty assumed around agents' initial beliefs. Thus, one key new insight of this paper is to show that the weighting of data under learning is regulated by both the sequence of learning gains and the initials uncertainty.

I focus on applications with a recursive least squares, a popular learning mechanism assumed to represent agents' econometric learning in the bounded rationality literature. The pace of recursive least squares learning is regulated by the sequence of learning gains, which also determines how different pieces of information are weighted in the implied estimates of agents' perceived law of motion. One popular specification is that of a constant gain, according to which the weights given to past observations decrease geometrically and lead to the emergence of perpetual learning. Importantly, the influence of initial beliefs can become non-negligible in applications with constant-gain learning, and this is where initials uncertainty matters the most. Particularly, I find that the assumption of *diffuse initials*, representing a rather extreme situation where the agent is completely uncertain about his/her initial beliefs, implies that the profile of weights given to past observations under constant-gain learning is in fact time-varying, causing a geometric decay of weighting faster than would have been expected from asymptotic analysis of the learning algorithm. In other words, the use of diffuse initials is equivalent to the use of higher starting learning gains that decrease as the sample grows and only converges to the actual constant gain asymptotically.

I show that this result can have important implications for the estimation of models with learning, as it can lead to an overweighting of the initial sample of observations used for model estimation. Without a proper account for the initials uncertainty, the estimation of models under the assumption of a constant gain over increasing samples of data would imply agents give a decreasing weight to more recent observations. In other terms, as we accumulate more data about the economy, and use this additional data in the estimation of models, the underlying estimated expectations are likely to become less sensitive to new information than they used to be with the earlier, and hence smaller, samples of data. This can introduce a downward bias on renewed estimates of the relevance of the expectations formation mechanism in the determination of the latest economic developments. On the other hand, diffuse initials can be used as a way to speed up convergence of learning estimates. In most applications, learning is assumed to represent a process that was ongoing prior to the beginning of the econometrician's estimation sample. In the lack of proper estimates of such initial beliefs, diffuse initials offer an interesting alternative to be used in training samples. Hence, the effect of misspecified initials uncertainty I discuss in this paper is a concern mainly for its use within the model estimation sample. Considering that this approach has been considered in previous applications in the literature – discussed below – it is important to understand the effects that the associated information weighting distortions can have on model estimates.

To quantify these potential biases I simulate the estimation of a forward-looking Phillips curve model and find that, indeed, diffuse initials lead to stronger small sample distortions in the model estimates. Particularly, the misspecified initials result in a systematic underestimation of the relevance of expectations in this model, as well as an overestimation of the responsiveness of inflation rates to measures of economic slack. The simulation analysis also allows an investigation of the channels through which these effects emerge, pointing to an increase in the variance of expectations associated with the diffuse initials as the main cause for the estimation biases. This result is consistent with the analysis of information weighting of constant-gain learning using diffuse initials; particularly, the equivalent higher gains at the beginning of the estimation sample lead to higher variability in the learning estimates, which ultimately increases the variance of the implied expectations. Empirical estimates with decadal sub-samples of U.S. data also indicate that the diffuse initials can distort estimates of the relevance of expectations for the determination of inflation, particularly implying a lower degree of violations to expectational stability conditions across the different sub-samples.

Another important implication of initials uncertainty misspecification is that it can bias the estimation of learning gains. Consider again the case of diffuse initials under constant-gain learning. Because the diffuse initials distort the weighting that a constant gain gives to sample observations, particularly leading to higher equivalent gains at the beginning of the sample, the best-fitting gain values to represent agents expectations formation process will tend to be biased downwards. To validate this claim I explore the effects of initials uncertainty on the estimation of individual gains using survey data on inflation expectations from professional forecasters. Heterogeneity of learning gains constitute an important source for the emergence of heterogeneous expectations across individuals – the related literature is discussed below. Namely, allowing agents to differ on how much weight they give to real-time data when forming their expectations can naturally generate distributions of expectations. As expected, I find that gain estimates based on expectations calculated with diffuse initials are systematically lower than gain estimates obtained with initials uncertainty estimated from a pre-sample of data – i.e., with initials uncertainty consistent with the idea of an ongoing learning process. Throughout the sample period, from 1968q4 to 2019q4, the average individual gains obtained under diffuse initials fluctuate between values of 0.02 and 0.07, while those obtained with the pre-sample initials uncertainty fluctuate between 0.07 and 0.12 – it is noteworthy that typical gain calibrations sit around a value of 0.03 for applications with quarterly macroeconomic data. Considering that a higher gain attenuates the relevance of initial beliefs, it is intriguing that the misspecification of initials uncertainty, particularly in terms of making initial beliefs more diffuse, leads to such a substantial underestimation of the learning gains.

1.1 Relation to literature

This paper is related to several branches of the literature. First, the role of initial beliefs in adaptive learning has been previously studied by Carceles-Poveda and Giannitsarou (2007), who showed that initials can have non-negligible effects in applications with constant-gain learning. Slobodyan and Wouters (2012b) examined the effects that initial beliefs can have on the estimation of macroeconomic models with learning, particularly pointing to its effects on model dynamics. Berardi and Galimberti (2017b) evaluate methods used to obtain the initial estimates from data and show that misspecified initials can lead to significant biases in the estimation of other model parameters too. However, the focus of this literature has been mainly on the effects and determination of the initial beliefs, without much attention to the issue about initials uncertainty. This paper contributes to that extant literature by showing the relevance of initials uncertainty.

Initial beliefs uncertainty are especially important for the estimation of models with adaptive learning (e.g., Milani, 2007; Chevillon et al., 2010). One important feature of such applications is that the learning gain, commonly assumed to be constant, is estimated jointly with other model parameters. The rationale for a constant gain is that it allows the modelling of persistent deviations of expectations from equilibrium while agents give a higher weight to more recent observations. As I show in this paper, the misspecification of the initials uncertainty can distort the profile of weights given to sample observations by a constant-gain learning mechanism. One important implication is that the joint estimation of initials uncertainty and the learning gain, as in, e.g., Milani (2008, 2011); Slobodyan and Wouters (2012b), can distort the behavioural and statistical identification of these parameters. Namely, in this paper I show that the effect of an initially higher learning gain can be equivalently obtained with a higher degree of initials uncertainty, and vice versa.

This finding is also related with the literature assessing the impact of the specification of prior beliefs on the dynamics of expectations (Cho et al., 2002; Sargent and Williams, 2005). In this literature, prior beliefs are defined within a Bayesian estimation context and represent agents' beliefs about time-varying perceived laws of motion. Sargent and Williams (2005) show that such priors determine not only the convergence of learning to equilibria, but also the transient dynamics of expectations towards equilibria. It is important to note that these prior beliefs determine the specification of both the learning gains and the proper initialization of beliefs according to the corresponding model of agents' beliefs. For example, when agents are assumed to update their beliefs with a constant-gain recursive least squares learning mechanism, as in the applications covered here, this assumption corresponds to a specific Bayesian prior about drifting beliefs (see Berardi and Galimberti, 2013). An important contribution of this paper is to stress and quantify the importance of initial beliefs uncertainty for the implied transient dynamics of expectations estimated under this popular prior belief assumption.

In this paper I show that the use of diffuse initials, by which the uncertainty agents are assumed to assign to initial beliefs is overstated, can be particularly harmful to the estimation of models with learning. Typical applications of learning in the literature evaluate the robustness of estimation results to alternative specifications of initials, occasionally including the diffuse initials approach in the analysis. For example, Slobodyan and Wouters (2012a), in a context of Kalman filter learning, find that the fit and evolution of beliefs within a macroeconomic model can be strongly affected by the use of diffuse initials. Lubik and Matthes (2016) also find that diffuse initials can affect inferences about equilibrium indeterminacy, particularly with respect to the U.S. Great Inflation period of the 1970s. Matching survey expectations data with adaptive learning, Markiewicz and Pick (2014) favour the diffuse initials for providing a superior forecasting performance of a range of macroeconomic and financial variables. Thus, an important concern for applications of adaptive learning is the lack of guidance on what is the appropriate approach to initialize learning beliefs. This paper attempts to address this issue with respect to the specification of initials uncertainty. Particularly, the diffuse initials approach is shown to lead to a distortion of information weighting as well as biased estimates of model parameters and the learning gain.

This is also related with a recent strand of the literature attempting to understand the heterogeneity of individual inflation expectations through the lens of adaptive learning. Pfajfar and Santoro (2010) identified different mechanisms of expectations formation to model the distribution of consumers inflation expectations. Malmendier and Nagel (2016) propose an age-dependent adaptive learning model whereby individual lifetime experiences of inflation help explain the distribution of consumer inflation expectations. Within a benchmark New Keynesian macroeconomic model, Cole and Milani (2021) estimate individual learning gains that help explain the heterogeneity of individual expectations observed from survey of professional forecasters data on output growth and inflation. Intriguingly, previous applications in the literature commonly report that gain estimates obtained under diffuse initials tend to be lower than those obtained under non-diffuse initials. In this paper I show that this result is to be expected and reflects the weighting distortion associated with an artificial and unwarranted increase of initial beliefs uncertainty.

This paper also contributes to the analysis of structural estimation of forward-looking Phillips curve models (see, e.g., Mavroeidis et al., 2014; Coibion et al., 2018). One of the key findings in this paper relates to the impact of initials uncertainty on the estimation of the parameter determining the relevance of inflation expectations, which ends up downward biased when estimated under diffuse initials. In the standard New Keynesian Phillips curve model, this parameter corresponds to the subjective discount factor, a parameter that is commonly calibrated close to unity. Under learning this parameter is also a key determinant of expectational stability, and here I consider estimation of this parameter as a mean to infer the emergence of unstable expectations. This approach also relates to recent research giving emphasis to additional behavioural channels to the discounting term of forward-looking variables, such as: limited common knowledge (Angeletos and Lian, 2018), finite planning horizons (Woodford, 2019), or cognitive discounting (Angeletos and Lian, 2018). This paper contributes to this literature by showing that (mis)specification of initial beliefs uncertainty can affect estimates of the relevance of forward-looking expectations for economic dynamics.

Finally, this paper contributes to the literature drawing analytical expressions for the weighting of information implied by different specifications of the learning algorithm, particularly focusing on the recursive least squares algorithm. The formulation of the recursive least squares algorithm can be traced back to the systems identification literature (see, e.g., Ljung and Soderstrom, 1983), where this algorithm is derived as the recursive formulation of a weighted least squares estimator. Here, I describe a renewed and more general non-recursive representation of this algorithm. Namely, I show that the recursive least squares is more properly represented by a penalized weighted least squares estimator, where a penalty term accounts for the effects of the learning initial estimates. The framework proposed in this paper provides flexible analytical expressions for the calculation of data weights under alternative assumptions on the behaviour of the learning gains, including the traditional decreasing-gain (Marcet and Sargent, 1989) and constant-gain (Sargent, 1999) specifications, as well as more sophisticated mechanisms such as endogenous gain-switching (Marcet and Nicolini, 2003) and age-dependent (Malmendier and Nagel, 2016) gain specifications.

1.2 Paper organization

The remainder of this paper is split into five other sections: Section 2 outlines the learning framework and derives the main analytical expressions about information weighting and initials uncertainty; Section 3 presents simulation analysis of the distortionary effects of diffuse initials in the estimation of a forward-looking Phillips curve model with learning; Section 4 presents an empirical application of the same model to U.S. data; Section 5 shows the relevance of initials uncertainty for the estimation of individual learning gains; Section 6 concludes with some remarks. Detailed derivations and supplementary results are provided in the Appendix.

2 Learning Framework

In this section I outline the general framework of recursive learning in order to derive the information weighting implications of different specifications of the learning gains and initials. Focusing on the case of a constant-gain, I then show how the specification of uncertainty about the initial learning estimates can distort the profile of weights that this popular learning mechanism assigns to observations at the beginning of estimation samples.

2.1 Background

In models with adaptive learning a perceived law of motion (PLM) is specified relating the variables agents are assumed to observe and those variables they care and need to form expectations about. Focusing on a univariate case, a typical PLM specification is given by a linear regression model of the form

$$y_t = \mathbf{x}_t' \Phi_t + \varepsilon_t, \tag{1}$$

where y_t is assumed to be related to a vector of (pre-determined) variables, $\mathbf{x}_t = (x_{1,t}, \ldots, x_{k,t})'$, through the vector of (possibly time-varying) coefficients $\Phi_t = (\phi_{1,t}, \ldots, \phi_{k,t})'$, and ε_t denotes a white noise disturbance term. This specification can be easily extended to a multivariate context, by augmenting y_t, ε_t, and Φ_t with extra columns, and to different specifications of lag/lead in the timing of expectations, by adjusting the timing of $\mathbf{x}_t$ elements.

In a typical economic modelling context, the observations of y_t needed to estimate (1), as well as some or all of the regressors in $\mathbf{x}_t$, are endogenously determined within a hypothetical structural model. These observations are the result of market equilibrium and the interaction between the economic decisions by the many different actors that compose the economy, such as households, firms, and policymakers. Hence, due to the relevance of expectations for these agents' economic decisions, the same macroeconomic outcomes that are relevant for the formation of expectations are themselves determined by expectations, a feature often called self-referentiality.

Notwithstanding, for the purposes of deriving weighting expressions I will abstract from one side of this self-referential nature of expectations, and focus instead on the modelling of agents' PLM. I.e., I will assume agents form expectations according to equation (1) without accounting for the endogeneity of the involved variables. Notice, as is usually the case in applications of adaptive learning, this implies some degree of bounded rationality in the way agents form expectations. The effects of self-referentiality are taken into account in the simulation and empirical estimation exercises presented in later sections.

2.2 Learning and initials uncertainty

A recursive estimator is assumed to represent how agents update their PLM estimates as new observations become available. One popular algorithm is given by the Recursive Least Squares (RLS),

$$\Phi_t = \Phi_{t-1} + \gamma_t \mathbf{R}_t^{-1} \mathbf{x}_t \left(y_t - \mathbf{x}_t' \Phi_{t-1} \right), \quad (2)$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \gamma_t \left(\mathbf{x}_t \mathbf{x}_t' - \mathbf{R}_{t-1} \right), \quad (3)$$

where $\mathbf{R}_t$ stands for an estimate of regressors' matrix of second moments, $E\left[\mathbf{x}_t \mathbf{x}_t'\right]$, and γ_t is a learning gain parameter. The learning gain is an important parameter of the learning mechanism because it determines how quickly new information is incorporated into the recursive estimates, and hence, how quickly agents react to different pieces of information (this relation will be discussed in the next subsection). Moreover, as a recursive process, these estimates need to initialized: Φ_0 are initial estimates representing agents' beliefs at the beginning of the econometrician's sample of data, and the inverse of $\mathbf{R}_0$ can be interpreted as a measure of the uncertainty agents assign to these initial estimates.

The main focus of this paper is about the determination of the initial beliefs uncertainty, $\mathbf{R}_0$, in a context of econometric estimation of economic models with learning. Naturally, the initial estimates should ideally be set or estimated to be consistent with plausible agents' beliefs at the beginning of the modelled sample. Berardi and Galimberti (2017b) study methods for the estimation of Φ_0 aimed to achieve such a goal, although assuming a fixed $\mathbf{R}_0$ across model estimation exercises. As it turns out, alternative assumptions of $\mathbf{R}_0$ can play an important role in the estimation of models with learning. The main contribution of this paper is to provide an analysis of this component.

From a Bayesian point of view, $\mathbf{R}_t$ is inversely related to the uncertainty in the corresponding Kalman filter estimates of Φ_t modelled as a random walk (see Evans et al., 2010; Berardi and Galimberti, 2013). Hence, $\mathbf{R}_0 \to \mathbf{0}$, henceforth denoted as diffuse initials, can be interpreted as increasing the uncertainty about the initial estimates, in which case the observations at the beginning of the estimation sample will be given extra weight to compensate for the initials uncertainty.[1]

This effect has two main implications for the estimation of models with learning. First, diffuse initials can be used as a way to accelerate convergence of learning estimates to a process

[1] Also, notice that if $\mathbf{R}_0 = \mathbf{0}$ (exactly rather than as a limit), (2)-(3) implies that $\Phi_1 = \left(\mathbf{x}_1 \mathbf{x}_1'\right)^{-1} \mathbf{x}_1 y_1$, which will be indeterminate for $k > 1$. For this reason, in the applications that follow I approximate diffuse initials by downscaling a reference $\mathbf{R}_0$ towards zero by multiplying it by a small constant.

representing ongoing learning that was already happening prior to the beginning of the econometrician's estimation sample. This particular property makes the diffuse initials an interesting alternative to be used in training samples. Second, within an estimation sample, the overweighting of initial observations distorts the representativeness of the implied expectations, which, in turn, can affect model estimates that depend on the behaviour of such expectations. To be more precise, in what follows I show how information weighting can be traced back to the joint definition of the learning gains and initials using a renewed and more general non-recursive representation of the learning algorithm.

2.3 Non-recursive form and information weighting

The weight given to a sample observation determines the amount of information from that particular observation that is incorporated into the PLM estimates. In the RLS algorithm of equations (2)-(3), such weighting of information is controlled by the sequence of learning gains. More precisely, the sequence of learning gains can be related with the relative weights given to sample observations in the estimation process. In order to draw this relationship it is useful to consider the non-recursive formulation corresponding to this estimation problem. When initialized from arbitrary initials, $\tilde{\Phi}_0$ and $\tilde{\mathbf{R}}_0$, the RLS has a non-recursive form given by

$$\begin{aligned}\Phi_t &= \arg\min_{\tilde{\Phi}_t} \sum_{i=1}^{t} \omega_{t,i} \left(y_i - \mathbf{x}_i'\tilde{\Phi}_t\right)^2 + \omega_{t,0}\left(\tilde{\Phi}_0' - \tilde{\Phi}_t'\right)\tilde{\mathbf{R}}_0\left(\tilde{\Phi}_0 - \tilde{\Phi}_t\right), \quad (4)\\ &= \left[\sum_{i=1}^{t} \omega_{t,i}\mathbf{x}_i\mathbf{x}_i' + \omega_{t,0}\tilde{\mathbf{R}}_0\right]^{-1}\left[\sum_{i=1}^{t} \omega_{t,i}\mathbf{x}_i y_i + \omega_{t,0}\tilde{\mathbf{R}}_0\tilde{\Phi}_0\right], \quad (5)\end{aligned}$$

where the weights are related to the sequence of learning gains according to

$$\omega_{t,i} = \begin{cases} \prod_{j=1}^{t}\left(1-\gamma_j\right) & for\ i = 0\,(initial), \\ \gamma_i \prod_{j=i+1}^{t}\left(1-\gamma_j\right) & for\ 0 < i < t, \\ \gamma_t & for\ i = t. \end{cases} \quad (6)$$

Thus, when the initials are taken into account, the RLS is equivalent to a Weighted Least Squares (WLS) estimation problem augmented with a penalty on squared deviations between estimates and initials. To the best of my knowledge this non-recursive formulation of the RLS for arbitrary initials has never been outlined in the previous literature. In fact, the origins of the RLS can be traced back as the recursive formulation of the WLS solution (without the penalty on initials) to the minimization of the sum of weighted error squares in the systems identification literature (see, e.g., Ljung and Soderstrom, 1983). Hence, the innovation here stems from following the inverse approach, i.e., taking the recursive form of (2)-(3) with initials $\{\tilde{\Phi}_0, \tilde{\mathbf{R}}_0\}$ as the starting point, I obtain (5)-(6), which, in turn, can be translated as a solution to the estimation problem in equation (4).

The non-recursive formulation above allows the calculation of such weights for any arbitrary sequence of learning gains. Also notice that the weights, $\omega_{t,i}$, defined in equation (6), are already in relative terms, as obtained by dividing the absolute weights by the sum of weights given to sample observations and the initials. This follows from the fact that, under the correspondence between the RLS and the penalized WLS outlined in this paper, the sum of weights will always be equal to unity.

Moreover, it is often interesting to evaluate how the observation weights evolve relative to the last observation in the sample, i.e., redefining equation (6) in terms of lags relative to the end of the sample, $\varpi_{t,l} = \omega_{t,t-l}$, in which case we look at

$$\varpi_{t,l} = \begin{cases} \prod_{j=1}^{t} \left(1-\gamma_j\right) & for\; l = t, \\ \gamma_{t-l} \prod_{j=0}^{l-1} \left(1-\gamma_{t-j}\right) & for\; 0 < l < t, \\ \gamma_t & for\; l = 0. \end{cases} \tag{7}$$

2.4 Constant-gain learning and diffuse initials

The constant-gain (CG) learning specification was introduced in the applied learning literature by Evans and Honkapohja (1993) and became popular after Sargent (1999) for its improved capability of tracking the evolution of time-varying environments. This specification has also been under the spotlight of the most recent research on the dynamic modelling of expectations for its potential of generating escape dynamics over finite stretches of time (see Williams, 2019) and asymptotically stable distributions of beliefs (see Galimberti, 2019).

One important property of the CG-RLS relates to the persistent influence of the learning initials. Under CG-RLS learning, $\gamma_t^{cg} = \bar{\gamma}$, and the weights are given by

$$\omega_{t,i}^{cg} = \begin{cases} (1-\bar{\gamma})^t & for\; i = 0, \\ \bar{\gamma}(1-\bar{\gamma})^{t-i} & for\; 0 < i \leq t. \end{cases}$$

Hence, the relative weights given to sample information by the CG-RLS decrease with the observation lag ($l = t - i$), while the weight given to the initials decreases with the sample size. However, the duration of the effects of the initials within finite samples will depend on the learning gain. For example, the number of observations needed to equate the cumulative weight given to sample information to the weight given to the initials can be easily calculated as

$$\sum_{j=1}^{i^*} \omega_{t,j}^{cg} = \omega_{t,0}^{cg},$$

$$i^* = \frac{\log\left(1/2\right)}{\log\left(1-\bar{\gamma}\right)}.$$

For a learning gain of $\bar{\gamma} = 0.03$, a value typically found in applications with quarterly macroeconomic data (see Berardi and Galimberti, 2017a, for a review), $i^* \simeq 23$, i.e., it takes about six years of quarterly data for the CG-RLS to assign a higher weight to the sample of observations than the weight given to the learning initials in the PLM estimates. This clearly highlights the importance of properly estimating such learning initials under CG learning.

Another important property of the CG-RLS relates to its asymptotic weighting behaviour relative to lagged observations. Although the CG-RLS assigns a vanishing weight to any given sample observation i, the weight given to observations at a fixed lag l do not change with t. This property makes the CG-RLS particularly well suited for modelling the behavioural assumption that agents give a higher emphasis to the more recent observations than to those collected farther into the past.

However, without a proper initialization of $\mathbf{R}_0$, the weights given to lagged observations by the CG-RLS can decay faster than the profile of weights expected from its asymptotic operation. Particularly, under diffuse initials, $\mathbf{R}_0 \to \mathbf{0}$, the CG-RLS weights are given by

$$\varpi_{t,l}^{dcg} = \begin{cases} 0 & for\, l = t, \\ \frac{\bar{\gamma}(1-\bar{\gamma})^l}{1-(1-\bar{\gamma})^t} & for\, 0 \leq l < t, \end{cases} \tag{8}$$

where the information from the initials are completely discarded.[2] More importantly, the weight given to sample observations are declining not only with the observation lag, but also with the size of the sample. These effects are illustrated in figure 1, which depicts the lagged weights given to sample observations under diffuse initials for varying sample sizes. Notice the asymptotic weights depicted in figure 1 are in fact equivalent to $\varpi_{t,l}^{cg}$. Hence, although the relative sample weights under diffuse initials still decrease as the observation becomes outdated, the actual profile of sample weights is not time-invariant. As we will show in the next section, other than the distortion that such diffuse initials can cause to the behavioural interpretation of CG learning, such weighting distortions can generate non-negligible estimation biases in small samples.

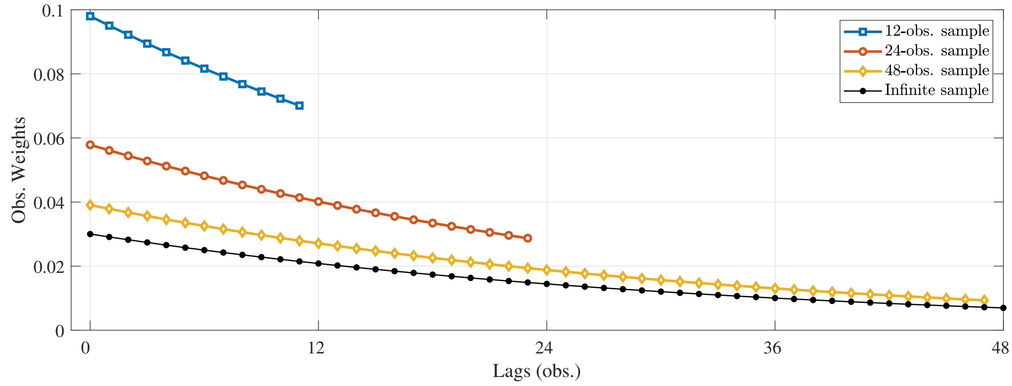
Before turning to a quantification of such estimation biases, notice that an alternative view about the distortionary effects of diffuse initials is obtained by solving for the equivalent time-varying gains. Namely, equating equation (7) to equation (8) one can find that

$$\tilde{\gamma}_t = \bar{\gamma} / \left(1 - (1 - \bar{\gamma})^t\right), \tag{9}$$

where $\tilde{\gamma}_t$ stands for the time-varying gains equivalent, in terms of information weighting, to a constant-gain $\bar{\gamma}$ under diffuse initials. The behaviour of such time-varying gains are illustrated in figure 2. Notice the equivalent gain used on the first observation under diffuse initials is

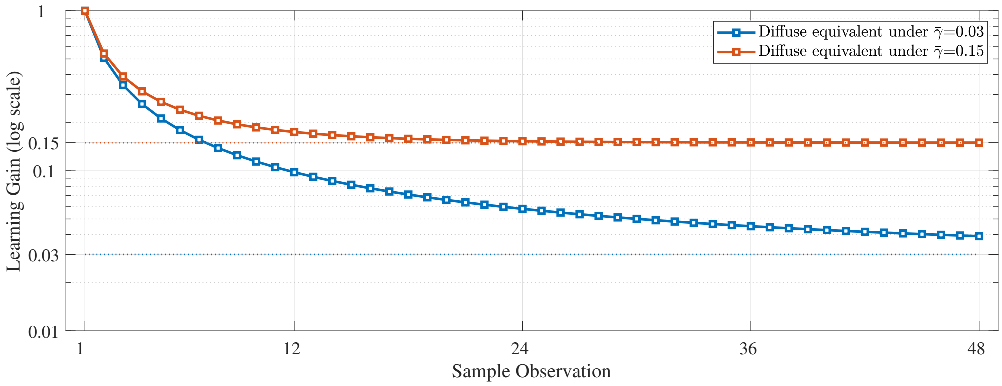
[2]To see why the initials receive zero weight under diffuse initials, notice from equation (5) that Φ_t becomes insensitive to $\tilde{\Phi}_0$ when $\tilde{\mathbf{R}}_0 \to 0$.

Figure 1: Constant-gain weights under diffuse initials.



Notes: Weights calculated using equation (8) with $\bar{\gamma} = 0.03$.

Figure 2: Equivalent time-varying gains under diffuse initials.



Notes: The equivalent time-varying gains are calculated according to equation (9).

equal to 1 – this is consistent with the observation above that a diffuse initial gives a zero weight to the initials, which means the first learning estimate will be determined entirely by the first sample observation (see also footnote (1)). Hence, the use of CG learning under diffuse initials is equivalent to the application of a decreasing sequence of gains that only converges to the underlying constant gain asymptotically.

3 Simulation Analysis

One key finding of the analysis of information weighting under least squares learning above is that the assumption of diffuse initials can distort the profile of weights given to sample observations by a constant-gain mechanism. An immediate question of interest is how much can such weighting distortions lead to biases in the estimation of models with learning. I now turn to a quantification of these effects with a simulation of the estimation of a macroeconomic model.

3.1 Model

I focus on a standard New Keynesian Phillips Curve (NKPC) model, given by

$$\begin{aligned} \pi_t &= \beta \pi^e_{t+1} + \lambda x_t + \alpha + u_t, \\ x_t &= \rho x_{t-1} + v_t, \end{aligned} \tag{10}$$

where π_t is inflation; π^e_{t+1} represents agents' expectations for next period inflation; x_t is a proxy for real marginal cost, usually assumed to be proportional to the labour share of income and the output gap; and, u_t is a disturbance that can be interpreted as a measurement error or as an unobserved cost-push supply shock. The forcing variable, x_t, is usually approximated empirically with measures of output gap, labour share of income, or unemployment rates (see Mavroeidis et al., 2014). The parameters in equation (10) can be interpreted as semi-structural when associated to deeper structural parameters of a micro-founded model of firms' staggered price setting (see Woodford, 2003, Ch.3). Particularly, under a Calvo framework, β stands for a discount factor, common across firms, while λ decreases with the fraction of firms that cannot update their prices in any given period, which leads to a "flatter" Phillips curve.

Under adaptive learning agents form expectations according to a PLM given by

$$\pi_t = \varphi_{t-1} + \phi_{t-1} x_t + z_t, \tag{11}$$

where z_t stands for an expectation error, and $\Phi_t = \{\varphi_t, \phi_t\}$ are parameters estimated with the RLS algorithm of (2)-(3). These PLM parameters are expected to converge to the rational expectations equilibrium (REE), $\phi^* = \lambda/(1-\beta\rho)$ and $\varphi^* = \alpha/(1-\beta)$, as long as $\beta < 1$ and $\beta\rho < 1$ (E-stability condition, see Evans and Honkapohja, 2001, pp. 198-200). The RE solution provides an interesting reference to simple reduced form estimates of the Phillips curve relationship between π_t and x_t. Particularly, for given α, β and λ, the implied ϕ^* and φ^* provide a description of the trade-off between inflation and, say, unemployment, after expectations have converged to equilibrium.

3.2 Simulation design

I generate 10,000 samples of artificial series of π_t and x_t assuming that $\bar{\gamma} = 0.03$, $\beta = 0.9$, $\lambda = 0.2$, $\alpha = 0$, $\rho = 0.75$, $u_t \sim N(0,3)$, and $v_t \sim N(0,1)$. These are parameter calibrations close to their empirical counterparts. The learning gain of $\bar{\gamma} = 0.03$ is in the range of calibrations reported by Berardi and Galimberti (2017a, Fig.8) to match survey forecasts of US CPI inflation from professionals, consumers, and policymakers. The discount factor, β, is often calibrated at a value close to 1; here I set this parameter to a slightly lower value in order to avoid being too close to the E-stability upper bound during the estimation. Also note that E-stability conditions are not strictly stringent in the determination of expectations stability under constant-gain learning (see Galimberti, 2019). There is less empirical agreement with respect to the slope parameter λ, although a positive value with borderline significance is often reported in the literature (see Mavroeidis et al., 2014).

For each sample I simulate the model for 2,000 observations, and then discard the first 1,000 observations to remove transient effects from PLM initials. Letting t start from -1,000, the PLM learning coefficients, $\Phi_{-1,001}$, are initialized from their implied REE values and $\mathbf{R}_{-1,001}$ from $E\left[\mathbf{x}_t\mathbf{x}_t'\right]$. I then use the remaining data for estimation of β and λ, under varying sample sizes and initial $\hat{\mathbf{R}}_0$ assumptions. Particularly, to evaluate the effect of diffuse initials I consider estimates with: (i) the correct initials uncertainty, $\hat{\mathbf{R}}_0 = \mathbf{R}_0$, which in the artificial data is given by the estimate from observation 1,000 out of the 2,000 simulated observations; (ii) approximately diffuse initials, $\hat{\mathbf{R}}_0 = \kappa\mathbf{R}_0$, where κ is set to a small value of 10^{-4}; notice $\hat{\mathbf{R}}_0$ cannot be set exactly to zero as this would lead to degenerate estimates. All other parameters, including the learning initials Φ_0 and the learning gain $\bar{\gamma}$, are set to their actual values – I discuss robustness checks about these assumptions below in section 3.5.

Since I am fixing $\bar{\gamma}$, ρ, and α, estimation of β and λ is linear on $\hat{\pi}^e_{t+1}$ and x_t, where $\hat{\pi}^e_{t+1}$ is determined by the PLM (11) and the learning estimates from the LS algorithm (2)-(3), which, in turn, depend on the simulated data, $\bar{\gamma}$, ϕ_0, and $\hat{\mathbf{R}}_0$. In order to obtain a clearer understanding of the biases introduced by the alternative $\hat{\mathbf{R}}_0$ assumptions, I conduct stepwise estimation exercises, first starting with the separate estimation of β and λ with simple regressions given by

$$(\pi_t - \lambda x_t) = \hat{\beta}\hat{\pi}^e_{t+1} + \hat{\upsilon}_{1t}, \tag{12}$$

and

$$\left(\pi_t - \beta\hat{\pi}^e_{t+1}\right) = \hat{\lambda}x_t + \hat{\upsilon}_{2t}, \tag{13}$$

respectively. On another exercise I then estimate β and λ jointly with

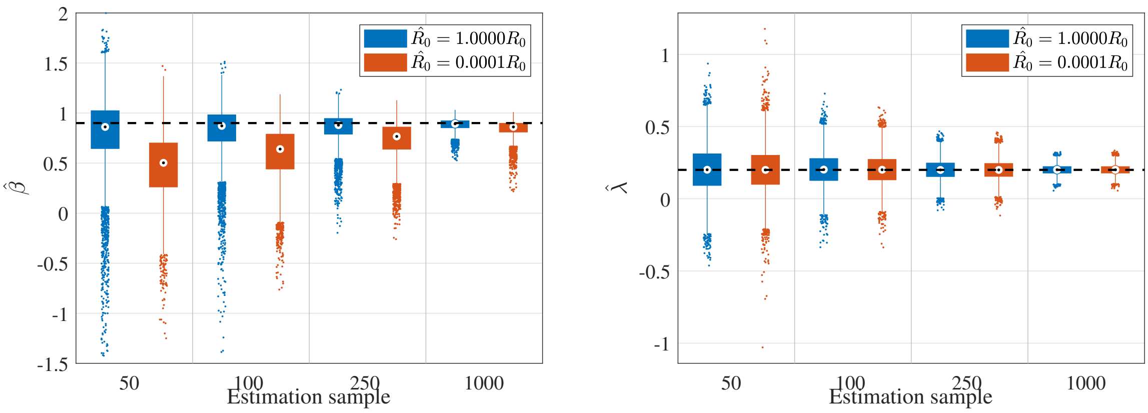
$$\pi_t = \hat{\beta}\hat{\pi}^e_{t+1} + \hat{\lambda}x_t + \hat{\upsilon}_{3t}. \tag{14}$$

All regressions are estimated using OLS. Christopeit and Massmann (2018) examine the asymptotic properties of the OLS estimator of structural parameters in models with learning, establishing its consistency in spite of non-standard distributions for traditional statistical inference. Interestingly, I found that the inclusion of an intercept in the estimation causes instabilities in the estimation of β – these effects can be attributed to strong persistence induced in $\hat{\pi}^e_{t+1}$ by the use of a low learning gain, as well as collinearity with the PLM intercept. Given that $\alpha = 0$ in the generated data, I estimate regressions without intercept.

3.3 Simulation results

I evaluate the effect of diffuse initials by considering how alternative assumptions of $\hat{\mathbf{R}}_0$ affect the estimates of β and λ. Starting with the individual estimation exercises, figure 3 depicts

Figure 3: Individual estimates of simulated Phillips curve model with constant-gain learning.



Notes: Estimates of equation (10) obtained individually, i.e., fixing λ when estimating β and vice versa, over 10,000 simulations of the model. The simulated data is generated with $\bar{\gamma} = 0.03$, $\beta = 0.9$, $\lambda = 0.2$, $\rho = 0.75$, $\sigma_u^2 = 3$, and $\sigma_v^2 = 1$. Both $\bar{\gamma}$ and ρ are pre-fixed to their true values during estimation. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively; the whiskers extend to ± 1.5 times the interquartile range, and estimates outside this range are considered outliers and are depicted as dots.

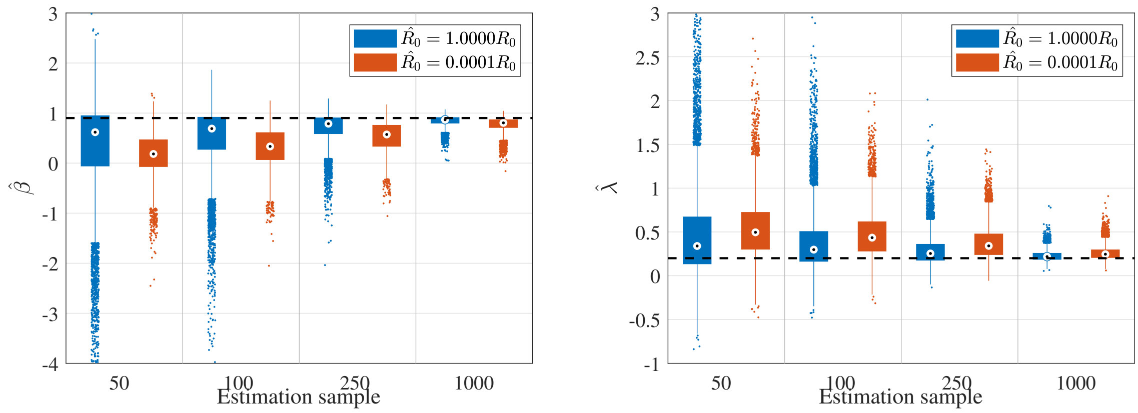
the distributions of model estimates for varying sample sizes and initials uncertainty. The β estimates are clearly downward biased under the assumption of diffuse initials, depicted in red.[3] As will be discussed below, this finding can be directly related to the diffuse initials weighting distortions discussed in the previous section.

The distributions of the β estimates are also highly skewed towards values below the true value of the parameter, especially for small samples – that is the case even for the estimates obtained under the correct initials. This result is consistent with the findings of Chevillon et al. (2010) showing that learning generates non-standard distributions of estimates of structural model parameters. Interestingly, the individual λ estimates are not affected by the varying initials, neither their distributions are affected by skewness. Hence, the inference difficulties caused by non-standard distributions under learning seem tied to the estimation of the model parameter associated with the expectations variable. As expected, these estimates tend to converge to their actual values for bigger estimation samples.

The results from the joint estimation exercise, depicted in figure 4, are similar to the previous exercise for the estimates of β, though with significant quantitative differences. For example, the median $\hat{\beta}$s estimated jointly using diffuse initials are between 0.72 ($T = 50$) and 0.10 ($T = 1{,}000$) below the true value of β, while in the individual estimation exercise these medians underestimated β by 0.40 ($T = 50$) and 0.04 ($T = 1{,}000$). The bias also increased significantly using the correct initials, e.g., rising from -0.04 in the individual estimation exercise

[3]The notion of bias is commonly defined as the mean difference between the estimate and its true value. Although the mean estimates are not directly depicted in the boxplot diagrams, the analysis based on the median estimates leads to equivalent conclusions as with the corresponding mean estimates.

Figure 4: Joint estimates of simulated Phillips curve model with constant-gain learning.



Notes: Same as figure 3 except that estimates are obtained jointly.

to -0.28 in the joint estimation one (both with $T = 50$). The estimates with the correct initials were also more strongly affected by sampling variation as reflected by the greater dispersion of estimates under the smaller estimation samples.

Another important difference in the joint estimation exercise relates to a positive bias in the estimates of λ, especially for the smaller estimation samples. The distributions of these λ estimates also show non-standard behaviour, though, in contrast to the β estimates, skewed towards values above the true value of the parameter. More importantly, here we again find that the estimates using the diffuse initials lead to greater biases. Quantitatively, the median bias in the jointly estimated $\hat{\lambda}$s using diffuse initials (+0.30 with $T = 50$, +0.05 with $T = 1,000$) are more than twice the bias obtained with the correct initials (+0.14 with $T = 50$, +0.02 with $T = 1,000$).

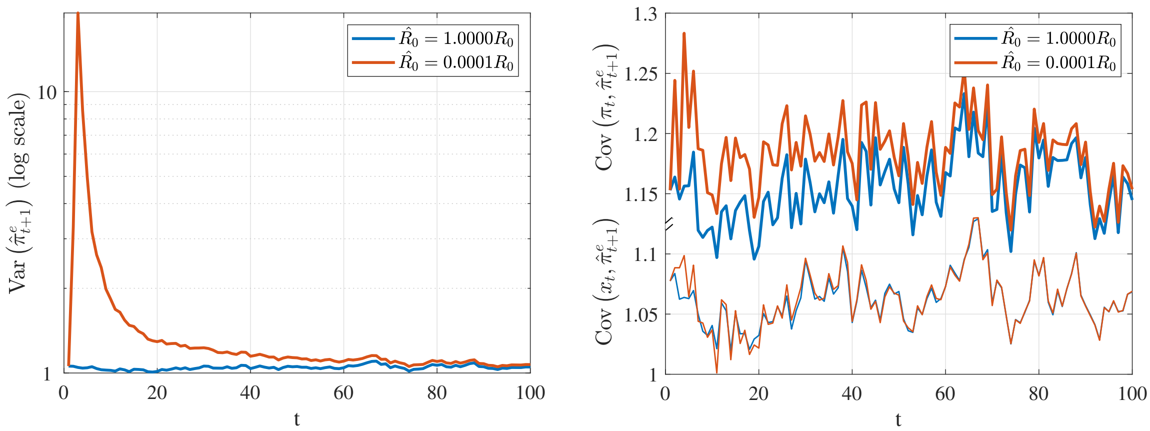
3.4 Analysis

In order to relate the biases documented above to the weighting distortions induced by diffuse initials, it is instructive to consider the limiting behaviour of the OLS estimators of the structural model parameters. Starting with the individual estimation exercises, the OLS estimate of equation (12) is given by

$$\hat{\beta}^{ols} = \frac{Cov(\pi - \lambda x, \hat{\pi}^e)}{Var(\hat{\pi}^e)}, \\ = \frac{Cov(\pi, \hat{\pi}^e) - \lambda Cov(x, \hat{\pi}^e)}{Var(\hat{\pi}^e)}, \tag{15}$$

where I drop the variables subscripts for succinctness. Clearly, $\hat{\beta}^{ols}$ would only differ, in probabilistic terms considering sampling variation, from the correct estimate of β because of deviations in $\hat{\pi}^e$ from its true value. Hence, the difference in estimation biases with respect to

Figure 5: Evolution of simulated data moments calculated across simulations.



Notes: Variances and covariances calculated for each period across the 10,000 simulations of equation (10) used in figures 3 and 4. In the second panel, the thicker lines (on top) depict $Cov\left(\pi_t, \hat{\pi}^e_{t+1}\right)$ while the thin lines (bottom) depict $Cov\left(x_t, \hat{\pi}^e_{t+1}\right)$.

$\hat{\mathbf{R}}_0$ assumptions can be understood as differences caused by such assumptions in the statistical moments of the simulated data relative to the implied expectations derived from learning.

Figure 5 presents the evolution of such statistics for the simulated data – $Cov\left(\pi, \hat{\pi}^e\right)$ and $Cov\left(x, \hat{\pi}^e\right)$ are presented together in the RHS panel. Of particular interest is the effect of increasing $Var\left(\hat{\pi}^e\right)$, which according to equation (15) would cause $\hat{\beta}^{ols}$ to decrease. As the LHS panel of figure 5 indicates, the use of diffuse initials led to a substantial inflation of the variance of the implied expectations at the beginning of the estimation sample, which explains the downward bias observed in the estimates of β under diffuse initials. This finding is also consistent with the analysis of the previous section showing that the diffuse initials lead to an overweighting of initial sample observations, or, equivalently, to an increase in the initial learning gains. As is well known, a higher learning gain leads to more volatile learning estimates (see, e.g., Evans and Honkapohja, 2001). Hence, the use of diffuse initials leads to more volatile learning estimates and their implied expectations, which ultimately translates into more biased estimates of the relevance of expectations in this model.

Similar analysis can be applied to the other model estimates. For the second exercise, the OLS estimate of equation (13) is given by

$$\hat{\lambda}^{ols} = \frac{Cov\left(\pi - \beta\hat{\pi}^e, x\right)}{Var\left(x\right)} = \frac{\overline{Cov\left(\pi, x\right)} - \beta Cov\left(x, \hat{\pi}^e\right)}{\overline{Var\left(x\right)}}. \tag{16}$$

In contrast to the first exercise, the individual OLS estimate of λ is not affected by the variance of the expectations variable – to facilitate analysis, terms not affected by the initials are depicted with a upper bar. Here, the only component that may cause differences between the initial assumptions is the covariance between the exogenous variable and the implied expectations, $Cov(x, \hat{\pi}^e)$. As the RHS panel of figure 5 indicates, this statistic was not strongly affected by the use of diffuse initials, which explains why there was no significant difference observed in the estimates reported for this exercise, in the RHS panel of figure 3.

Finally, for the joint estimation exercise the corresponding OLS estimates are given by

$$\hat{\beta}^{ols} = \frac{\overline{Var(x)}Cov(\pi, \hat{\pi}^e) - Cov(x, \hat{\pi}^e)\overline{Cov(\pi, x)}}{Var(\hat{\pi}^e)\left(\overline{Var(x)} - Cov(x, \hat{\pi}^e)\right)}, \tag{17}$$

$$\hat{\lambda}^{ols} = \frac{Var(\hat{\pi}^e)\overline{Cov(\pi, x)} - Cov(x, \hat{\pi}^e)Cov(\pi, \hat{\pi}^e)}{\overline{Var(x)}\left(Var(\hat{\pi}^e) - Cov(x, \hat{\pi}^e)\right)}. \tag{18}$$

Although in this case the effects become more convoluted, it is clear that: (i) the variance of expectations still has a negative effect on the β estimates if $Var(x) > Cov(x, \hat{\pi}^e)$, which was the case in the model simulation presented here; (ii) the estimates of λ are now also affected by initials uncertainty through its effects on the variance of expectations; particularly, it can be shown that when $Cov(x, \hat{\pi}^e) > 0$ (generally true for model 10 given that x enters the PLM) and $Cov(\pi, \hat{\pi}^e) > Cov(\pi, x)$ (also generally the case for $\beta > \lambda$), the λ estimates will be positively affected by the increasing variance of expectations associated with the diffuse initials. These two points offer an explanation for the biases caused by the diffuse initials reported in figure 4.

3.5 Robustness checks

I conducted several robustness checks on these simulation exercises, the results of which are all provided in the Appendix. First, regarding the assumption of correct initial beliefs, a more realistic situation is one where we need to obtain such estimates from the sample of data available. To address that concern I considered pre-sample estimates of Φ_0 and $\mathbf{R}_0$ and obtained similar results. Particularly, I considered initials obtained from a pre-sample of 25 observations using a variety of methods: (i) diffuse WLS initials, where Φ_0 and $\mathbf{R}_0$ are obtained using equation (5) and assuming $\tilde{\mathbf{R}}_{-26} = \kappa\mathbf{I}$; (ii) OLS initials; (iii) diffuse WLS initials departing from $\tilde{\mathbf{R}}_{-26} = \kappa\hat{\mathbf{R}}_0^{ols}$, where $\hat{\mathbf{R}}_0^{ols}$ is obtained from (ii); (iv) non-diffuse WLS initials departing from $\tilde{\mathbf{R}}_{-26} = \hat{\mathbf{R}}_0^{ols}$. All these alternatives lead to virtually equal figures as those presented in figures 3 and 4.

Another potentially controversial assumption is about the fixing of the learning gain to its actual value. For empirical purposes this parameter has to be either calibrated on the basis of survey data or actuals forecasting performance (see, e.g., Markiewicz and Pick, 2014; Berardi and Galimberti, 2017a), or jointly estimated with other structural model parameters (see, e.g., Milani, 2007; Chevillon et al., 2010; Slobodyan and Wouters, 2012b). Estimation of the learning gain introduces nonlinearities in the determination of jointly estimated structural model parameters and is complicated by weak identification and persistent dynamics (Chevillon et al., 2010). Particularly, under constant-gain learning RE weak identification issues are propagated as $\bar{\gamma} \to 0$ (no learning), and collinearity between $\hat{\pi}^e_{t+1}$ and x_t increases the lower the learning gain.

In spite of these issues, in order to address the fixed gain concern I also conducted exercises with the joint estimation of the learning gain using nonlinear least squares.[4] Figure 6 presents the resulting estimates of the learning gains – corresponding estimates of β and λ are again distributed similarly as in figures 3 and 4, hence I provide them only in the Appendix. As expected from the discussion above, the learning gain estimates tend to be widely dispersed, especially for smaller samples. Nevertheless, as the sample size increases, the median gain estimates tend to converge faster to the true gain value under the pre-sample initials than using the diffuse initials. In fact, notice that the gain estimates obtained under the diffuse initials tend to concentrate around smaller values than those obtained with a pre-sample. This is consistent with this paper's argument that the diffuse initials distort the weights associated with the learning gains upwards. Because initial sample observations are overweighted, a lower learning gain is required to explain the same degree of updating. Thus, the re-scaling of $\mathbf{R}_0$ generates a substitution effect on the role played by the learning gain.

4 Empirical Estimation of a Phillips Curve with Learning

I now turn to an empirical evaluation of the effects of diffuse initials on the estimation of the standard NKPC model with constant-gain learning. As in the simulation analysis of the previous section, the main focus of this empirical exercise is on the effect that initials uncertainty can have on the estimates of the model parameters. Particularly, I again consider two alternatives for the initial matrix of second moments: (i) an estimate obtained with the pre-sample data to represent the "correct" initials, henceforth denoted as the non-diffuse initials – details about this initialization are provided below; and, (ii) a downward re-scaled version of (i) to represent the diffuse initials approach.

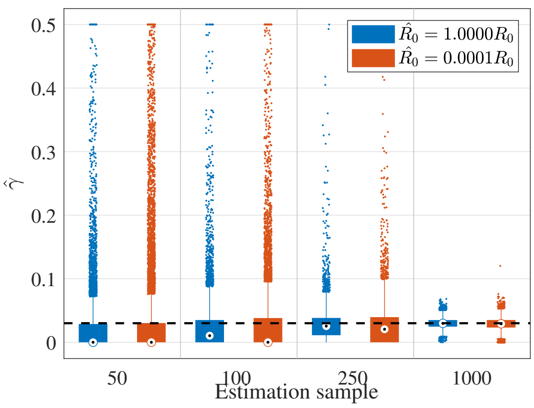
4.1 Data and estimation approach

I use U.S. quarterly data covering the period from 1947 to 2019, focusing on estimates of β and λ across decade sub-samples. The focus on sub-samples allows an analysis of the stability of the Phillips curve relationship, which has historically attracted great interest in the literature (see, e.g., Gordon, 2011). Another known issue with empirical estimates of the NKPC relates to their sensitivity with respect to the data definitions of the measures of price inflation and production slack (see Mavroeidis et al., 2014).

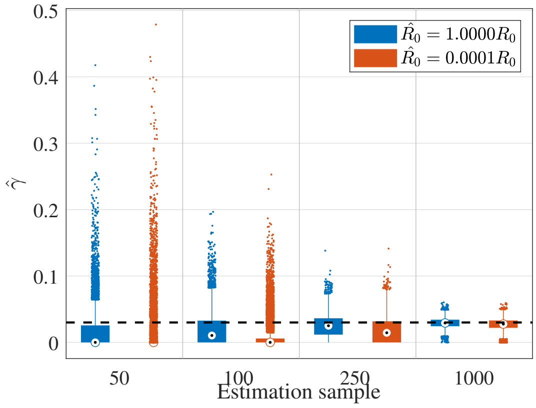
[4] Due to the sensitivity of expectations to the values of γ and β, estimation is conducted using a constrained optimization algorithm, with bounds given by: $\gamma \in [0, 0.5]$ and $\beta \in [-4, 3]$.

Figure 6: Estimates of constant-gain from simulated Phillips curve model.

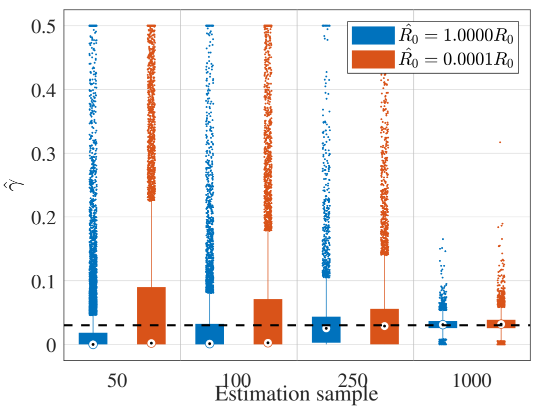
(a) Joint with β (pre-fixing λ).



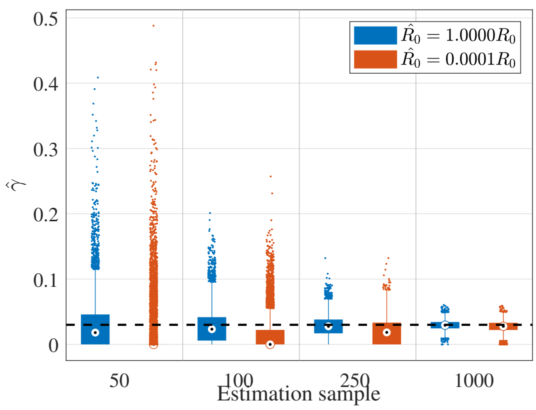
(b) Joint with λ (pre-fixing β).



(c) Joint with β and λ.



(d) Individually (pre-fixing β and λ).



Notes: Same as figure 3 except that gain estimates are obtained by nonlinear least squares jointly with estimates of β, in panel (a), jointly with estimates of λ, in panel (b), jointly with estimates of both β and λ, in panel (c), and, individually while pre-fixing β and λ to their true values, in panel (d).

To deal with such specification uncertainty, I consider combinations of three inflation measures, based on the CPI, the core CPI, and the GDP deflator, with four alternative proxies for real marginal costs, namely, an output gap measure based on real GDP data, non-farm business sector labour shares, unemployment rates, and the unemployment rate gap relative to an estimate of the natural rate. Inflation rates are annualised by multiplying the quarterly rates by four. For comparative purposes, all measures of x_t are filtered over the full sample using the Hodrick-Prescott filter (with $\lambda^{HP} = 1,600$), and all measures, including the inflation rates, are standardized prior to estimation to have zero mean and variance equal to unity over the full sample. All data series are obtained from the FRED database of the St. Louis Fed.

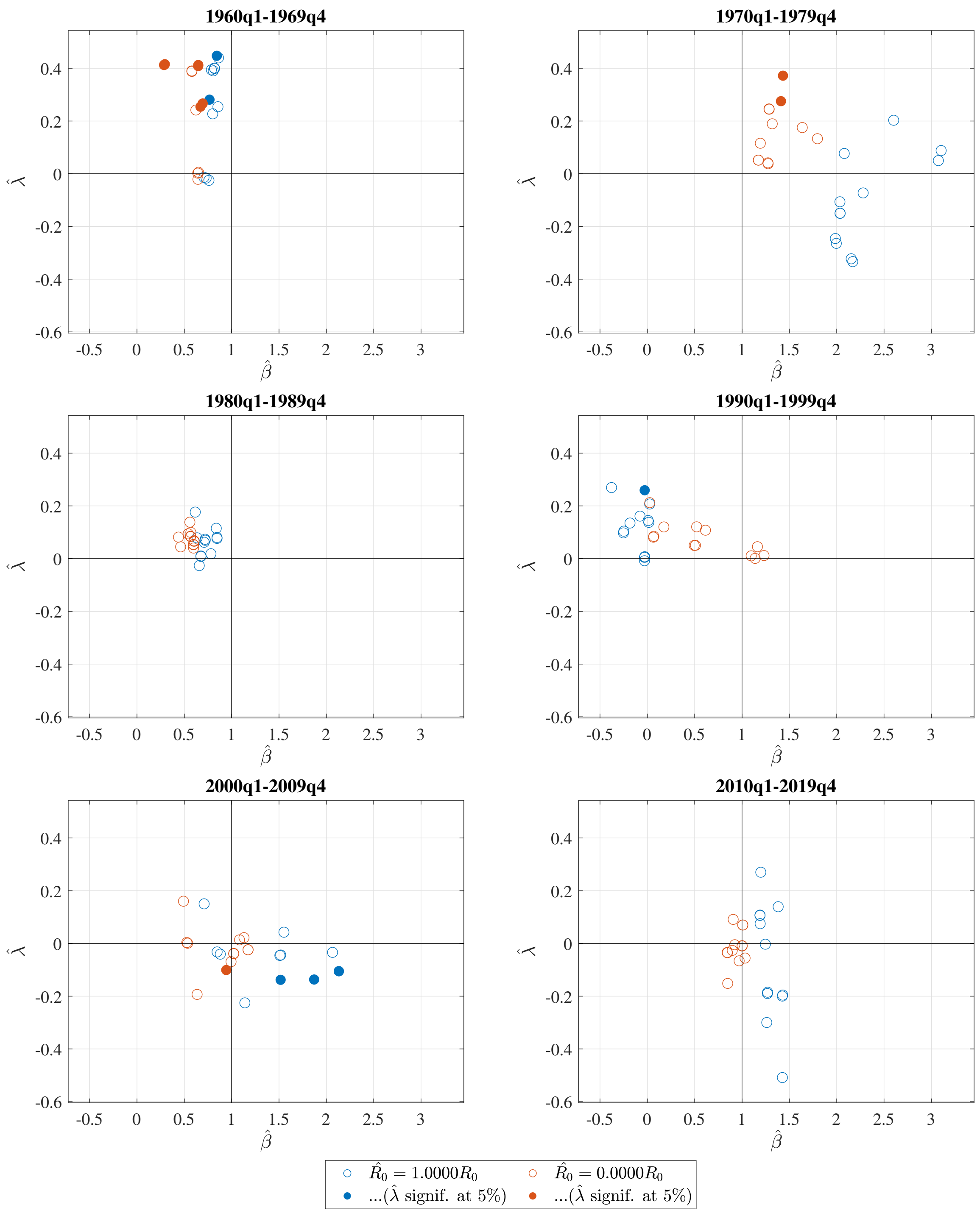
To be consistent with the previous simulation analysis, all model estimates are obtained using OLS. I focus on the joint estimation of β and λ, estimating regressions of the form of equation (14), while pre-fixing the other parameters to plausible values: $\bar{\gamma} = 0.03$ is again fixed according to the calibrations reported by Berardi and Galimberti (2017a) to match survey forecasts; ρ is pre-estimated by fitting a first-order autoregression on the full-sample of each measure used as x_t; the learning initials $\{\varphi_0, \phi_0, \mathbf{R}_0\}$ are estimated over pre-sample data using WLS in order to obtain initials consistent with the constant-gain learning adopted in the estimation sample (see Berardi and Galimberti, 2017b). The pre-samples include all data available prior to the start of each decade estimation sample. Note the PLM includes both a constant and a slope coefficient on x_t as in equation (11).

4.2 Results and analysis

Figures 7 and 8 present the estimation results. There is substantial variation in the model estimates across the variables definitions and the sub-samples. The λ estimates are mostly consistent with their expected signs up to the end of the 20th century, although rarely with statistical significance (depicted with a filled marker). In contrast, the majority of the sub-sample β estimates are statistically significant at the 5% significance level (not depicted). However, as discussed above, such inferences should be interpreted with caution considering that learning can generate non-standard distributions of statistical tests.

According to the model estimates, the 1960s may be considered as the "golden days" of the Phillips curve, as several λ estimates display statistical significance and signs according to expectations. Similarly, most β estimates in the 1960s sub-sample are statistically significant and below unity, hence satisfying E-stability conditions in this model. The estimates for the 1970s, in contrast, indicate an important change on the estimates associated with the forward-looking expectations in this model. Namely, the β estimates increase above unity, also with a more robust increase using the non-diffuse initials, which suggests a period of unstable inflation expectations relative to observed inflation rates. The β estimates then return to the E-stability range during the 1980s, while the λ estimates become less dispersed between the values of 0 and 0.2.

Figure 7: Estimates of U.S. Phillips curve with constant-gain learning by decades.



Notes: Estimates of equation (14) obtained using different combinations of data definitions for inflation, $\pi_t = \{\text{CPI, core CPI, GDP deflator}\}$, and proxy for real marginal cost, $x_t = \{\text{real GDP gap, labour share, unemployment, natural rate of unemployment gap}\}$. For comparative purposes, both π_t and x_t are standardized prior to estimation, and estimates with unemployment as x_t are depicted as $-\hat{\lambda}$. All estimates obtained under a fixed learning gain, $\bar{\gamma} = 0.03$. Statistical significance at the 5% level are depicted for $\hat{\lambda}$ with filled markers and are based on HAC standard errors. Such inferences under learning should be interpreted with caution since estimators distributions can become nonstandard.

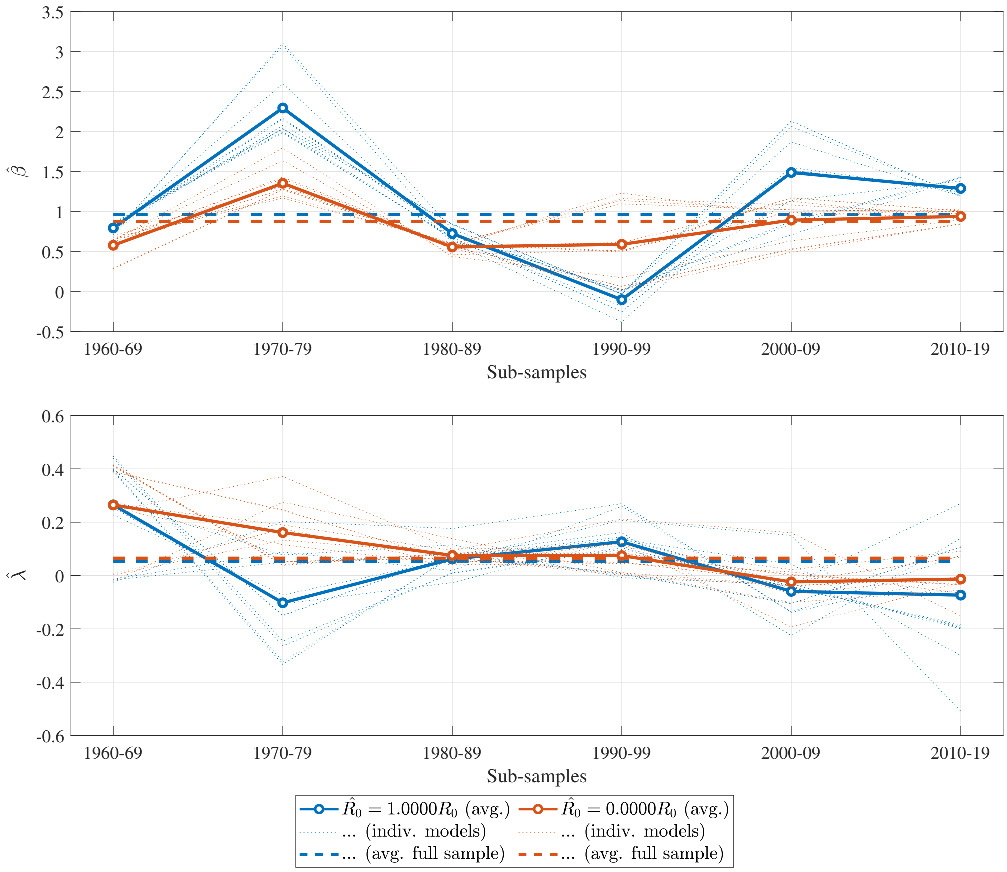
These results are consistent with the view that a strong correlation between inflation and economic activity may have misled policymakers to believe on an apparent trade-off between inflation and unemployment in the 1960s. The associated decline of active stabilization policies then led to an increase in inflation expectations in the 1970s, here reflected as a period of unstable expectations, which ultimately increased actual inflation. This is the so-called Great Inflation period, which prompted the monetary authority to revert to a more active policy of inflation and expectations stabilization in the 1980s (see, e.g., Orphanides and Williams, 2005; Primiceri, 2006; Sargent et al., 2006).

The estimates from the 1990s reflect a period of decreased relevance of inflation expectations, a result that, again, seems more robust with the use of the non-diffuse initials. This result is indicative of a build-up of credibility in the monetary authority resolve to keep inflation stable. Interestingly, in the 2000s and 2010s, the β estimates jump again outside the E-stability range, especially for the estimates based on non-diffuse initials. At the same time, the λ estimates become more dispersed and move towards negative values. This is consistent with previous evidence in the learning literature that decreasing beliefs about inflation persistence provide an explanation for lower average and volatility of inflation during the so-called Great Moderation period (1986–2006) in the U.S, as well as the flattening of the Phillips curve (Slobodyan and Wouters, 2012a).

More important to the purposes of this paper, the empirical estimates of the NKPC are found to depend on the assumption about the initials uncertainty. The evolution of the averages across the different specifications, presented in figure 8, indicate that the β estimates under the diffuse initials tend to be less sensitive to the sub-samples. This is consistent with the previous simulation evidence that diffuse initials lead to underestimation of β. Hence, the estimates obtained under diffuse initials are less informative about violations of expectational stability over time. For the λ estimates, the use of diffuse initials also point to a smoother flattening of the Phillips curve relative to the estimates with non-diffuse initials.

Table 1 presents another comparative between these estimates, focusing on one particular specification that uses the GDP deflator for inflation and the natural rate of unemployment gap as a proxy for real marginal costs – hence the expected sign of λ is negative. The β estimates obtained under the diffuse initials are mostly smaller than those obtained with the non-diffuse ones, except for the 1990-99 decade, when inflation expectations are found to lose significance. The λ estimates obtained with the two alternative initials assumptions move in different directions throughout the sub-samples, but agree on their sign and statistical significance for the 1960s. In the last two decades, $\hat{\lambda}$ turns positive under both initials, but with a greater increase and statistical significance (at the 10% level) under the non-diffuse initials for the 2010s. Nevertheless, in conjunction with the results for $\hat{\beta}$, the implied reduced form slope of the Phillips curve is always negative under the non-diffuse initials, consistent with expectations about this relationship, while the estimates under diffuse initials imply inverted Phillips curves during the 1970s, 2000s and 2010s.

Figure 8: Evolution of estimates of U.S. Phillips curve with constant-gain learning by decades.



Notes: Same as figure 7.

Table 1: Empirical estimates of a U.S. Phillips curve with constant-gain learning by decades.

	1960-19 (240 qtrs.)	1960-69 (40 qtrs.)	1970-79 (40 qtrs.)	1980-89 (40 qtrs.)	1990-99 (40 qtrs.)	2000-09 (40 qtrs.)	2010-19 (40 qtrs.)
- Under pre-initialized $\hat{R}_0$:							
$\hat{\beta}$	0.973	0.843	1.986	0.676	-0.252	1.518	1.270
	(0.207)	(0.207)	(0.295)	(0.314)	(0.384)	(0.301)	(0.165)
$\hat{\lambda}$	-0.031	-0.447	0.246	-0.008	-0.097	0.044	0.185
	(0.094)	(0.214)	(0.225)	(0.088)	(0.156)	(0.063)	(0.107)
$\hat{\phi}^*$	-0.241	-1.823	-0.316	-0.021	-0.080	-0.122	-1.354
	(0.762)	(0.998)	(0.265)	(0.229)	(0.139)	(0.180)	(1.144)
- Under diffuse $\hat{R}_0$:							
$\hat{\beta}$	0.893	0.646	1.173	0.603	1.097	1.022	1.001
	(0.174)	(0.096)	(0.154)	(0.179)	(0.350)	(0.183)	(0.127)
$\hat{\lambda}$	-0.036	-0.408	-0.052	-0.066	-0.010	0.038	0.008
	(0.090)	(0.153)	(0.134)	(0.087)	(0.094)	(0.072)	(0.083)
$\hat{\phi}^*$	-0.179	-0.968	1.047	-0.143	-0.565	0.440	0.080
	(0.457)	(0.326)	(4.857)	(0.221)	(6.851)	(1.239)	(0.858)
- PLM initials:							
$\hat{\varphi}_0$	-0.376	-0.376	-0.211	1.002	0.536	-0.161	-0.347
$\hat{\phi}_0$	-0.163	-0.163	-0.432	-0.288	-0.128	-0.111	-0.157

Notes: Estimates of equation (14) obtained using GDP deflator for inflation and the natural rate of unemployment gap as a proxy for real marginal cost. Both variables are standardized prior to estimation. All estimates obtained under a fixed learning gain, $\bar{\gamma} = 0.03$. The implied reduced form slope of the Phillips curve, $\hat{\phi}^*$, is obtained according to the REE. The PLM initials are obtained with WLS using all data available prior to the beginning of each decade estimation sample. Standard errors in parentheses are HAC robust. Inference under learning should be interpreted with caution since estimators distributions can become non-standard.

5 Estimation of Individual Learning Gains

Another field for which initials uncertainty can be important is the modelling of the distribution of individual expectations. A recent strand of the literature attempts to connect the heterogeneity of individual inflation expectations to varying aspects of adaptive learning (see, e.g., Malmendier and Nagel, 2016; Cole and Milani, 2021). Here I will explore the effects of initials uncertainty on the estimation of individual learning gains using survey micro data on inflation expectations. The idea is to model the distribution of individual expectations by allowing agents to differ only with respect to their learning gain. In other terms, agents are allowed to differ on how much weight they give to real-time data when forming their expectations. I then evaluate how the specification of initials uncertainty affect the estimates of the best-fitting learning gains.

5.1 Data, model and estimation

I focus on individual inflation forecasts data from the Survey of Professional Forecasters, obtained from the Philadelphia Fed. Particularly, I take one-quarter-ahead individual forecasts of the U.S. GDP price deflator, here denoted as $\pi^{spf}_{j,t+1}$, where j stands for the individual. After filtering out individuals with less than 10 observations, I end up with a sample of 209 individuals with quarterly forecasts collected between 1968q4 to 2019q4. This is an unbalanced dataset, in the sense that some individuals participated only at the beginning of the sample, others entered later, and so on. It is an interesting application because the number of observations for each individual is often small. To give an idea, the median number of observations across individuals here is 28.

In order to estimate individual learning gains, $\hat{\gamma}_j$, and to keep things simple, I assume agents use an AR(1) PLM. Importantly, these PLMs are fitted to real-time data and respecting the timing of the survey, for which individuals normally know only the lagged value of inflation at the time they submit their forecasts. Hence, to calculate learning-based one-quarter-ahead expectations, $\hat{\pi}^{e}_{j,t+1}$, I recursively estimate

$$\pi_{t-1} = \varphi_{t-1} + \phi_{t-1}\pi_{t-2} + \varepsilon_{t-1}, \tag{19}$$

using first-available real-time inflation rates, and then I calculate iterated forecasts according to

$$\hat{\pi}^{e}_{j,t+1} = \hat{\varphi}_{j,t-1}\left(1 + \hat{\phi}_{j,t-1}\right) + \hat{\phi}^{2}_{j,t-1}\pi_{t-1}. \tag{20}$$

The PLM coefficients, $\hat{\Phi}_{j,t} = \left\{\hat{\varphi}_{j,t}, \hat{\phi}_{j,t}\right\}$, are estimated with CG-RLS, where the individual gains are obtained by numerically minimizing the mean squared difference between the learning forecasts, $\hat{\pi}^{e}_{j,t+1}$, and the individual's forecasts observed in the survey data, $\pi^{spf}_{j,t+1}$.[5]

Because agents enter and exit the survey at different points in time, the PLM estimates can vary across individuals not only because of the different best-fitting learning gains, $\hat{\gamma}_j$, but also because of varying assumptions about the initial beliefs. Particularly, for an individual entering the survey at period $t = \tau$, $\hat{\Phi}_{j,\tau-1}$ is set according to a diffuse WLS estimate obtained with equation (5) using all inflation data available up to period $\tau - 1$, where the weights are determined by the learning gain using equation (6).[6] I then compare the individual gain estimates across different specifications of initials uncertainty, $\hat{\mathbf{R}}_{\tau-1}$, namely, (i) one using pre-sample initials uncertainty, $\hat{\mathbf{R}}_{\tau-1} = \hat{\mathbf{R}}^{wls}_{\tau-1}$, and (ii) the other with the diffuse initials approach, $\hat{\mathbf{R}}_{\tau-1} = (1/100)\,\hat{\mathbf{R}}^{wls}_{\tau-1}$, where the pre-sample initials uncertainty are re-scaled towards zero.

5.2 Results and analysis

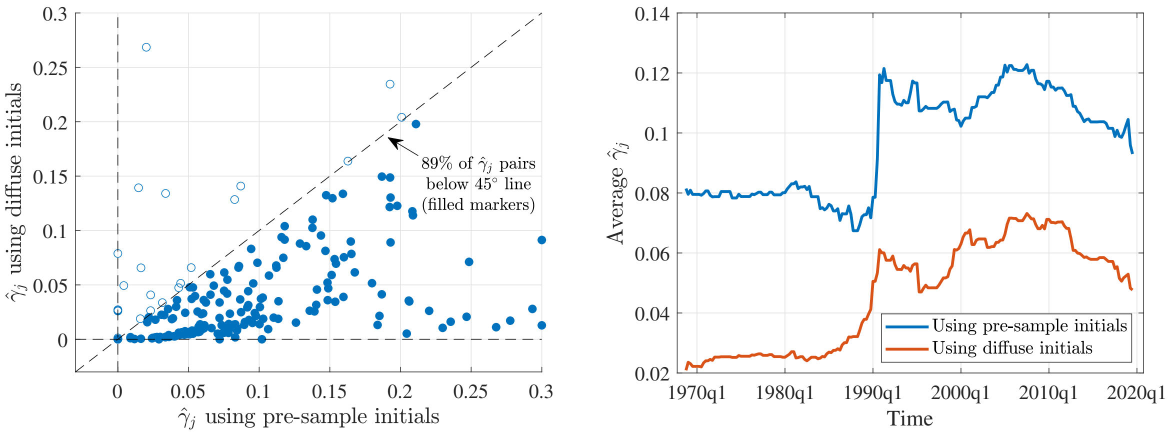
Figure 9 presents the resulting estimates. The scatterplot in the LHS panel depicts the gain estimates obtained using diffuse initials against the gains obtained under the pre-sample initials uncertainty. The diffuse initials approach leads to a downward bias in the estimates of individual learning gains. Namely, about 89% of the individual gain estimates based on diffuse initials are lower than their corresponding estimates using pre-sample initials uncertainty. This finding is in line with the distortionary effects of diffuse initials pointed out in this paper. Because the diffuse initials distort the weighting that a constant-gain gives to sample observations, particularly leading to higher equivalent gains at the beginning of the sample, the best-fitting gain values end up biased downwards.

A similar pattern emerges by looking at the time evolution of the average gains across the survey respondents, depicted in the RHS panel of figure 9. The average of individual gains under the diffuse initials are always lower than the estimates obtained with pre-sample initials, on average about 0.05 lower throughout the sample. Interestingly, both approaches indicate an increase in learning gains starting from 1990, which implies agents have been giving more weight to more recent observations when forming their inflation expectations than they used to give in the 1970-80s period.

[5]Estimation is conducted using a standard interior-point constrained optimization algorithm, where the gain is constrained between 0 and 0.3.

[6]The use of WLS pre-sample estimates makes the initial beliefs, $\hat{\Phi}_{j,\tau-1}$, depend on the gain estimate too. Results using the alternative of OLS pre-sample initial estimates are provided in the Appendix and lead to similar conclusions.

Figure 9: Estimates of individual learning gains from survey of inflation forecasts.



Notes: Individual gain estimates obtained by minimizing the mean squared difference between observed individual's inflation forecasts from the Survey of Professional Forecasters and learning-based forecasts constructed using an AR(1) PLM estimated by CG-RLS on real-time inflation data. The diffuse initials are obtained by downscaling the pre-sample initials uncertainty by a factor of 100. Filled markers in the first panel depict estimates for which the use of pre-sample initials uncertainty led to greater gain values than using diffuse initials. The average gains depicted in the second panel are calculated across the sample of survey respondents at each period.

6 Concluding remarks

In this paper I proposed a more general non-recursive representation of the recursive least squares algorithm that is used in the adaptive learning literature to represent how agents form their expectations in economic settings. According to this new formulation, the recursive learning mechanism is more properly represented by a penalized weighted least squares estimator, where a penalty term accounts for the effects of the learning initial estimates. The non-recursive formulation also allowed a renewed analysis of how information is weighted in the implied estimates of agents' perceived law of motion. Such weights are directly determined by the sequence of learning gains used in the recursive least squares algorithm, and the specification of the uncertainty around initial learning estimates. The framework proposed in this paper provides flexible analytical expressions for the calculation of information weighting under different assumptions on the evolution of the learning gains and initial beliefs.

One important finding obtained under this refreshed framework is that, without a proper account for the initial beliefs uncertainty, the estimation of models under the assumption of a constant gain over increasing samples of data would imply agents give a decreasing weight to more recent observations, distorting the real-time tracking interpretation of this mechanism. The relevance of this distortion was evidenced by simulation and empirical exercises, where the misspecified initials led to systematic biases to estimates of the relevance of expectations in a Phillips curve model, as well as to estimates of the responsiveness of inflation rates to output gaps. In another empirical exercise, estimates of individual learning gains using survey data on inflation expectations from professional forecasters were found to be significantly downward biased under inflated initial beliefs uncertainty. Thus, a proper account of how information is weighted under alternative learning mechanisms and assumptions about initial beliefs and their uncertainty are important aspects for the estimation of models with adaptive learning.

Finally, the recommended approach to address the estimation of initial beliefs uncertainty is to use a training sample for estimation of both beliefs and their uncertainty. According to the results presented in this paper, this is the most appropriate approach to elicit initial beliefs that are consistent with the view of learning as an ongoing process prior to the beginning of the econometrician's estimation sample. One approach that should be particularly discouraged is the procedure of inflating initials uncertainty, here denoted as the diffuse initials approach. This is obtained with a re-scaling of the initial matrix of second moments of the learning estimates. The results presented in this paper indicate that such approach is unwarranted and will lead to biased model estimates. However, one concession to the diffuse initials approach is for its use within training samples, for which the inflated initials uncertainty can accelerate convergence of beliefs within small pre-samples of data.

References

Angeletos, G.-M. and C. Lian (2018). Forward guidance without common knowledge. *American Economic Review 108*(9), 2477–2512.

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.

Berardi, M. and J. K. Galimberti (2017a). Empirical calibration of adaptive learning. *Journal of Economic Behavior & Organization 144*, 219 – 237.

Berardi, M. and J. K. Galimberti (2017b). On the initialization of adaptive learning in macroeconomic models. *Journal of Economic Dynamics and Control 78*, 26 – 53.

Carceles-Poveda, E. and C. Giannitsarou (2007). Adaptive learning in practice. *Journal of Economic Dynamics and Control 31*(8), 2659–2697.

Chevillon, G., M. Massmann, and S. Mavroeidis (2010). Inference in models with adaptive learning. *Journal of Monetary Economics 57*(3), 341–351.

Cho, I.-K., N. Williams, and T. J. Sargent (2002). Escaping nash inflation. *The Review of Economic Studies 69*(1), 1–40.

Christopeit, N. and M. Massmann (2018). Estimating structural parameters in regression models with adaptive learning. *Econometric Theory 34*, 68–111.

Coibion, O., Y. Gorodnichenko, and R. Kamdar (2018). The formation of expectations, inflation, and the phillips curve. *Journal of Economic Literature 56*(4), 1447–91.

Cole, S. J. and F. Milani (2021). Heterogeneity in individual expectations, sentiment, and constant-gain learning. *Journal of Economic Behavior & Organization 188*, 627–650.

Evans, G. W. and S. Honkapohja (1993). Adaptive forecasts, hysteresis, and endogenous fluctuations. *Economic Review*, 3–13.

Evans, G. W. and S. Honkapohja (2001). *Learning and expectations in macroeconomics*. Frontiers of Economic Research. Princeton, NJ: Princeton University Press.

Evans, G. W., S. Honkapohja, and N. Williams (2010). Generalized stochastic gradient learning. *International Economic Review 51*(1), 237–262.

Galimberti, J. K. (2019). An approximation of the distribution of learning estimates in macroeconomic models. *Journal of Economic Dynamics & Control 102*, 29–43.

Gordon, R. J. (2011). The history of the phillips curve: Consensus and bifurcation. *Economica 78*(309), 10–50.

Ljung, L. and T. Soderstrom (1983). *Theory and Practice of Recursive Identification*. The MIT Press.

Lubik, T. A. and C. Matthes (2016). Indeterminacy and learning: An analysis of monetary policy in the great inflation. *Journal of Monetary Economics 82*, 85 – 106.

Malmendier, U. and S. Nagel (2016). Learning from inflation experiences. *Quarterly Journal of Economics 131*(1), 53–87.

Marcet, A. and J. P. Nicolini (2003). Recurrent hyperinflations and learning. *American Economic Review 93*(5), 1476–1498.

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

Markiewicz, A. and A. Pick (2014). Adaptive learning and survey data. *Journal of Economic Behavior & Organization 107*, 685 – 707.

Mavroeidis, S., M. Plagborg-Moeller, 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). 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.

Orphanides, A. and J. C. Williams (2005). The decline of activist stabilization policy: Natural rate misperceptions, learning, and expectations. *Journal of Economic Dynamics and Control 29*(11), 1927–1950.

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. *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.

Sargent, T. J. and N. Williams (2005). Impacts of priors on convergence and escapes from nash inflation. *Review of Economic Dynamics 8*(2), 360 – 391.

Slobodyan, S. and R. Wouters (2012a). Learning in a medium-scale dsge model with expectations based on small forecasting models. *American Economic Journal: Macroeconomics 4*(2), 65–101.

Slobodyan, S. and R. Wouters (2012b). Learning in an estimated medium-scale dsge model. *Journal of Economic Dynamics and Control 36*(1), 26 – 46.

Williams, N. (2019). Escape dynamics in learning models. *Review of Economic Studies 86*(2), 882–912.

Woodford, M. (2003). *Interest and Prices*. Princetion, NJ: Princeton University Press.

Woodford, M. (2019). Monetary policy analysis when planning horizons are finite. *NBER Macroeconomics Annual 33*, 1–50.

A Proofs and Derivations

A.1 Correspondence between penalized WLS and RLS

To see how the RLS of (2)-(3) can be derived from the penalized WLS formulation of (5) and (6), first notice that iterating (3) recursively from $\mathbf{R}_0$ we have that

$$\mathbf{R}_t = \sum_{i=1}^{t} \omega_{t,i}\mathbf{x}_i\mathbf{x}_i' + \omega_{t,0}\mathbf{R}_0,$$

which is the inverse of the first term in (5), leading to

$$\hat{\phi}_t = \mathbf{R}_t^{-1}\left[\sum_{i=1}^{t} \omega_{t,i}\mathbf{x}_i y_i + \omega_{t,0}\mathbf{R}_0\phi_0\right]. \tag{21}$$

For the second term notice that

$$\begin{aligned}
\sum_{i=1}^{t} \omega_{t,i}\mathbf{x}_i y_i &= \sum_{i=1}^{t-1} \omega_{t,i}\mathbf{x}_i y_i + \gamma_t\mathbf{x}_t y_t, \\
&= (1-\gamma_t)\sum_{i=1}^{t-1} \omega_{t-1,i}\mathbf{x}_i y_i + \gamma_t\mathbf{x}_t y_t,
\end{aligned}$$

and

$$\omega_{t,0}\mathbf{R}_0\phi_0 = (1-\gamma_t)\,\omega_{t-1,0}\mathbf{R}_0\phi_0,$$

where we use

$$\omega_{t,i} = (1-\gamma_t)\,\omega_{t-1,i},$$

which follows from (6). Hence, (21) is equivalent to

$$\hat{\phi}_t = \mathbf{R}_t^{-1}\left[\gamma_t\mathbf{x}_t y_t + (1-\gamma_t)\left(\sum_{i=1}^{t-1} \omega_{t-1,i}\mathbf{x}_i y_i + \omega_{t-1,0}\mathbf{R}_0\phi_0\right)\right]. \tag{22}$$

Lagging (21) one period we find that

$$\mathbf{R}_{t-1}\hat{\phi}_{t-1} = \sum_{i=1}^{t-1} \omega_{t-1,i}\mathbf{x}_i y_i + \omega_{t-1,0}\mathbf{R}_0\phi_0,$$

which can be substituted into (22) to yield

$$\hat{\phi}_t = \mathbf{R}_t^{-1}\left[\gamma_t\mathbf{x}_t y_t + (1-\gamma_t)\,\mathbf{R}_{t-1}\hat{\phi}_{t-1}\right]. \tag{23}$$

From (3) notice that

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

which substituted into (23) and after rearranging leads to

$$\begin{aligned}\hat{\phi}_t &= \mathbf{R}_t^{-1}\left[\gamma_t\mathbf{x}_t y_t + \left(\mathbf{R}_t - \gamma_t\mathbf{x}_t\mathbf{x}_t'\right)\hat{\phi}_{t-1}\right],\\ &= \gamma_t\mathbf{R}_t^{-1}\mathbf{x}_t y_t + \hat{\phi}_{t-1} - \gamma_t\mathbf{R}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\hat{\phi}_{t-1},\\ &= \hat{\phi}_{t-1} + \gamma_t\mathbf{R}_t^{-1}\mathbf{x}_t\left(y_t - \mathbf{x}_t'\hat{\phi}_{t-1}\right),\end{aligned}$$

establishing the correspondence between the penalized WLS solution of (5) and the RLS of (2).

A.2 Absolute and relative weights

Letting W_t^n stand for the sum of weights starting from weight n up to weight t, from the definition of the absolute weights, (6), this sum of weights can be expanded according to

$$\begin{aligned}W_t^0 &= \sum_{i=0}^{t}\omega_{t,i},\\ &= \prod_{j=1}^{t}\left(1-\gamma_j\right) + \sum_{i=1}^{t-1}\gamma_i\prod_{j=i+1}^{t}\left(1-\gamma_j\right) + \gamma_t. \qquad (24)\end{aligned}$$

Expanding the first term of (24) we have that

$$\begin{aligned}\omega_{t,0} &= (1-\gamma_1)(1-\gamma_2)\dots(1-\gamma_{t-1})(1-\gamma_t),\\ &= (1-\gamma_2)\dots(1-\gamma_{t-1})(1-\gamma_t) - \gamma_1\prod_{j=2}^{t}\left(1-\gamma_j\right),\\ &= 1-\gamma_t - \sum_{i=1}^{t-1}\gamma_i\prod_{j=i+1}^{t}\left(1-\gamma_j\right). \qquad (25)\end{aligned}$$

Returning to (24) we then have

$$\begin{aligned}W_t^0 &= 1-\gamma_t - \sum_{i=1}^{t-1}\gamma_i\prod_{j=i+1}^{t}\left(1-\gamma_j\right) + \sum_{i=1}^{t-1}\gamma_i\prod_{j=i+1}^{t}\left(1-\gamma_j\right) + \gamma_t,\\ &= 1.\end{aligned}$$

A.3 Equivalent time-varying gains under diffuse initials

The sequence of gains, $\tilde{\gamma}_t$, that generates equivalent weightings as a constant-gain under diffuse initials needs to solve

$$\begin{aligned}\varpi_{t,l} &= \varpi_{t,l}^{dcg},\\ &= \frac{\bar{\gamma}(1-\bar{\gamma})^l}{1-(1-\bar{\gamma})^t} \qquad (26)\end{aligned}$$

for all t and l. From equation (7), starting with $l = 0$ we simply have that

$$\tilde{\gamma}_t = \bar{\gamma} / \left(1 - (1 - \bar{\gamma})^t\right). \tag{27}$$

It only remains to validate if equation (27) also solves equation (26) for $l > 0$. Substituting equation (27) into equation (7) for $0 < l < t$,

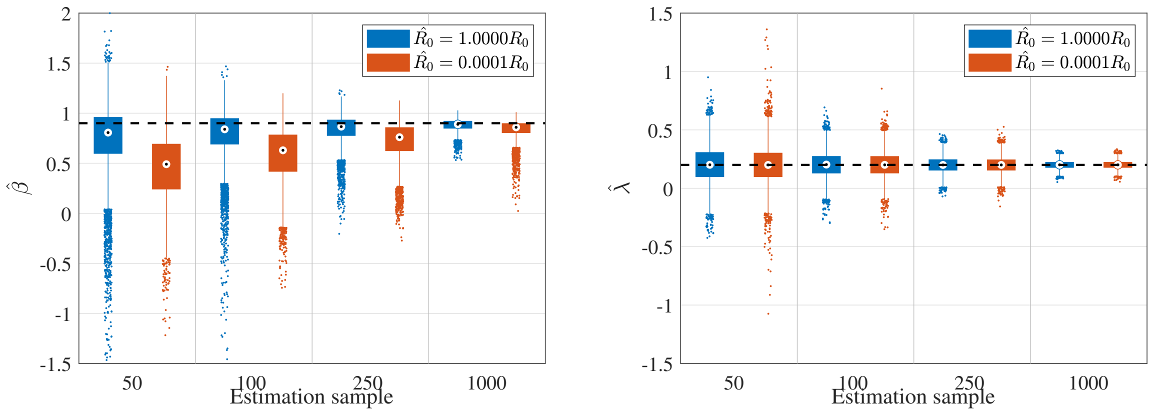
$$\begin{aligned}
\varpi_{t,l} &= \frac{\bar{\gamma}}{1 - (1 - \bar{\gamma})^{t-l}} \prod_{j=0}^{l-1} \left(1 - \frac{\bar{\gamma}}{1 - (1 - \bar{\gamma})^{t-j}}\right), \\
&= \frac{\bar{\gamma}(1 - \bar{\gamma})^l}{1 - (1 - \bar{\gamma})^{t-l}} \prod_{j=0}^{l-1} \left(\frac{1 - (1 - \bar{\gamma})^{t-j-1}}{1 - (1 - \bar{\gamma})^{t-j}}\right), \\
&= \frac{\bar{\gamma}(1 - \bar{\gamma})^l}{1 - (1 - \bar{\gamma})^{t-l}} \left(\frac{1 - (1 - \bar{\gamma})^{t-l}}{1 - (1 - \bar{\gamma})^{t}}\right), \\
&= \frac{\bar{\gamma}(1 - \bar{\gamma})^l}{1 - (1 - \bar{\gamma})^{t}},
\end{aligned}$$

which solves equation (26) for $l > 0$. Finally, notice that under diffuse initials the weight given to the learning initials is null, i.e., $\varpi_{t,t}^{dcg} = 0$. This is equivalent to using a $\gamma_1 = 1$, which is again satisfied by equation (27).

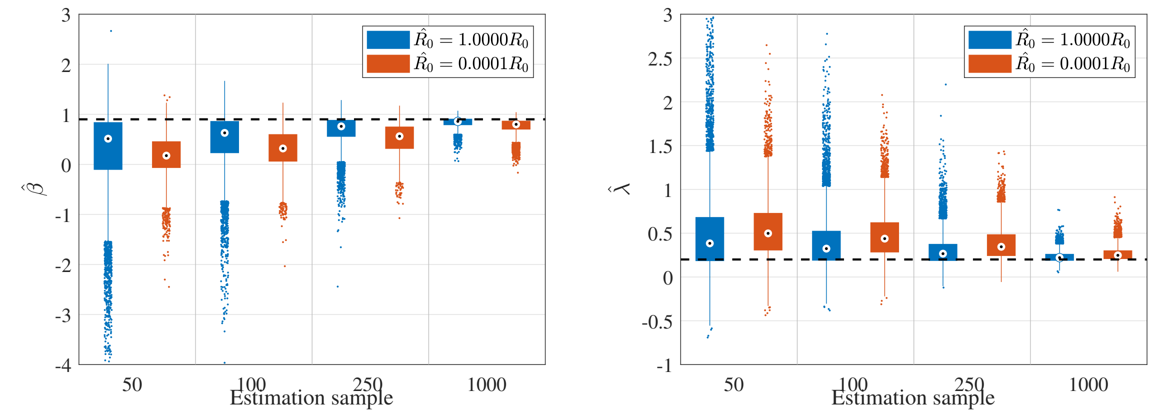
B Supplementary Results

Figure 10: Estimates of simulated Phillips curve model with constant-gain learning – using diffuse WLS pre-sample initials.

(a) Individual estimates.



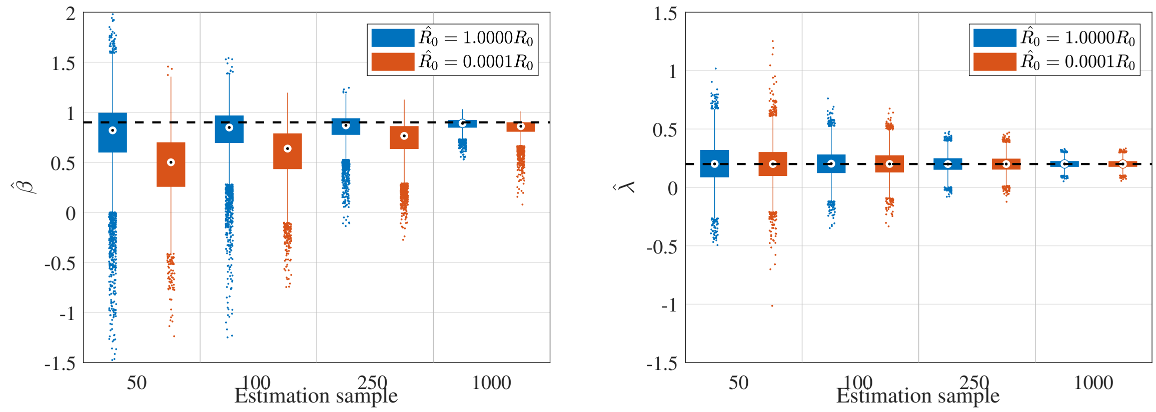
(b) Joint estimates.



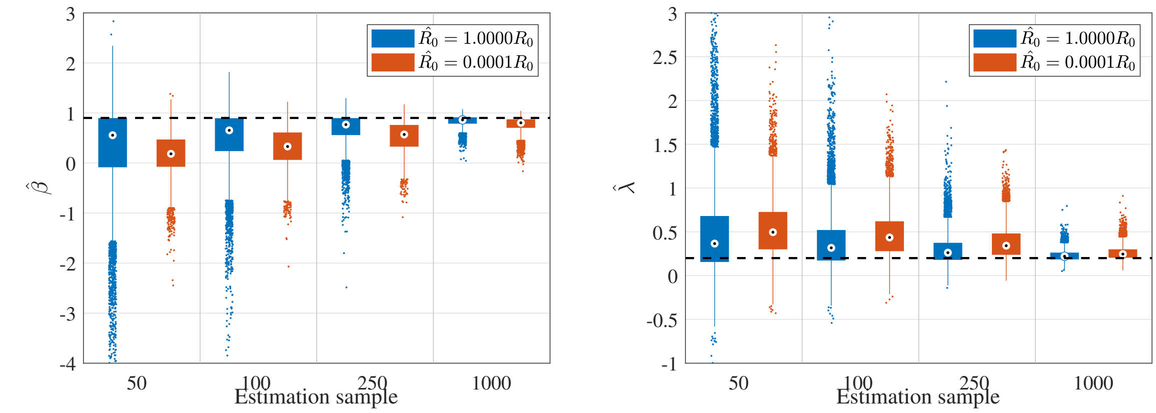
Notes: Same as figures 3 and 4 except that using diffuse WLS pre-sample estimates for $\mathbf{R}_0$.

Figure 11: Estimates of simulated Phillips curve model with constant-gain learning – using OLS pre-sample initials.

(a) Individual estimates.



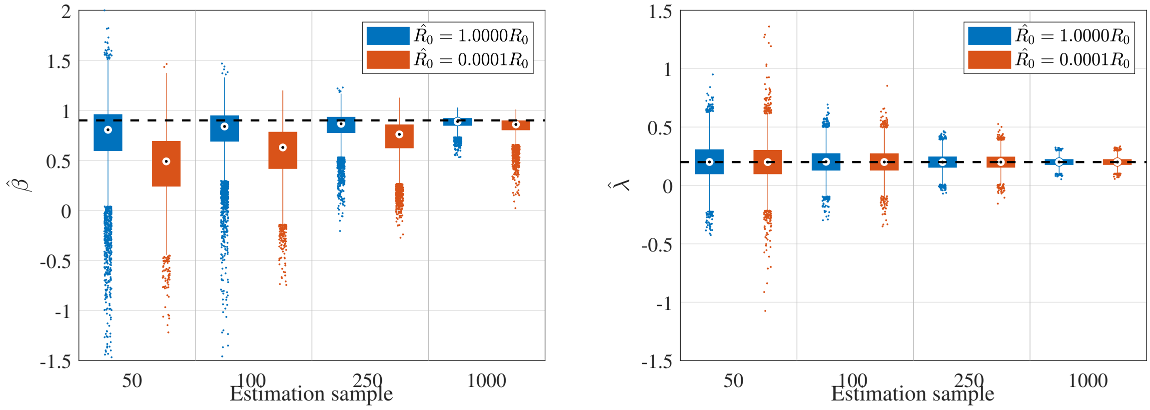
(b) Joint estimates.



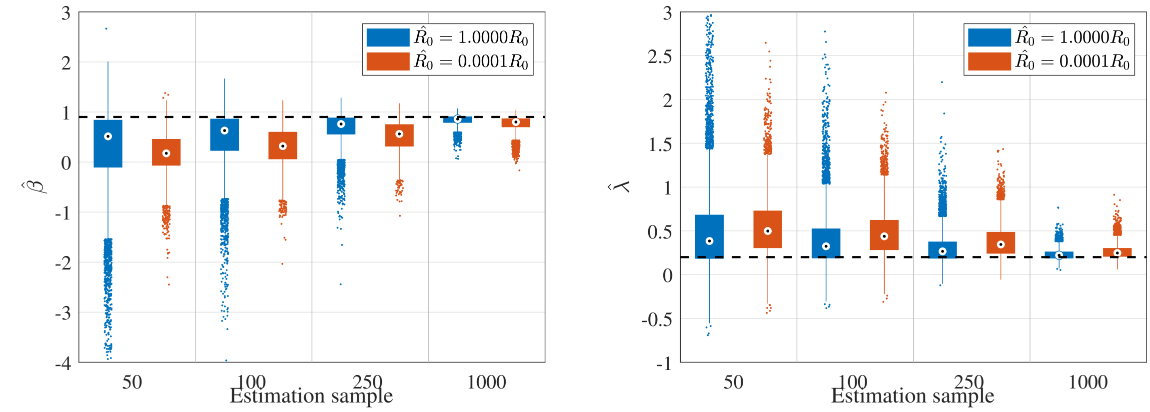
Notes: Same as figures 3 and 4 except that using OLS pre-sample estimates for $\mathbf{R}_0$.

Figure 12: Estimates of simulated Phillips curve model with constant-gain learning – using OLS plus diffuse WLS pre-sample initials.

(a) Individual estimates.



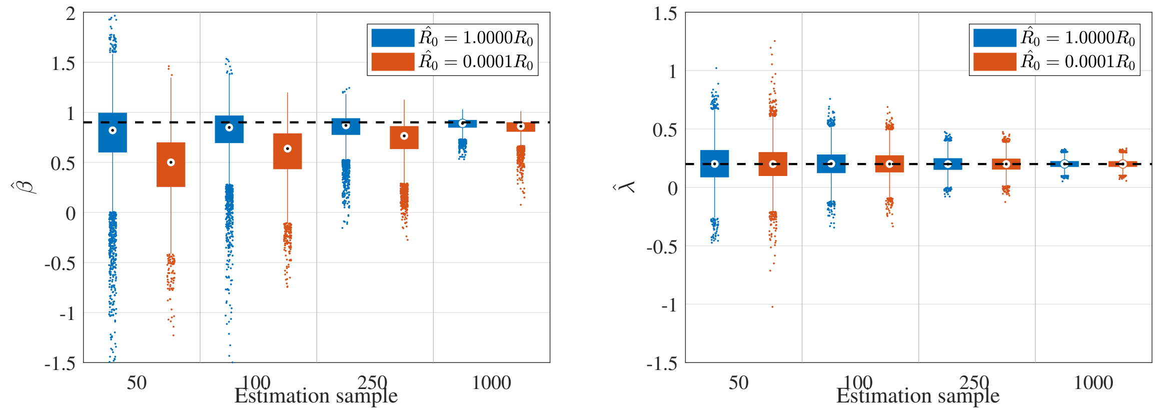
(b) Joint estimates.



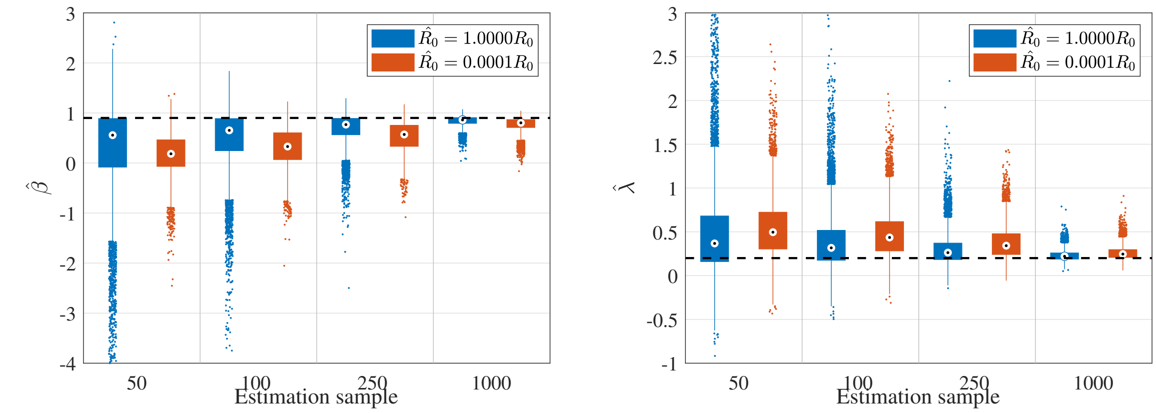
Notes: Same as figures 3 and 4 except that using OLS plus diffuse WLS pre-sample estimates for $\mathbf{R}_0$.

Figure 13: Estimates of simulated Phillips curve model with constant-gain learning – using OLS plus WLS pre-sample initials.

(a) Individual estimates.



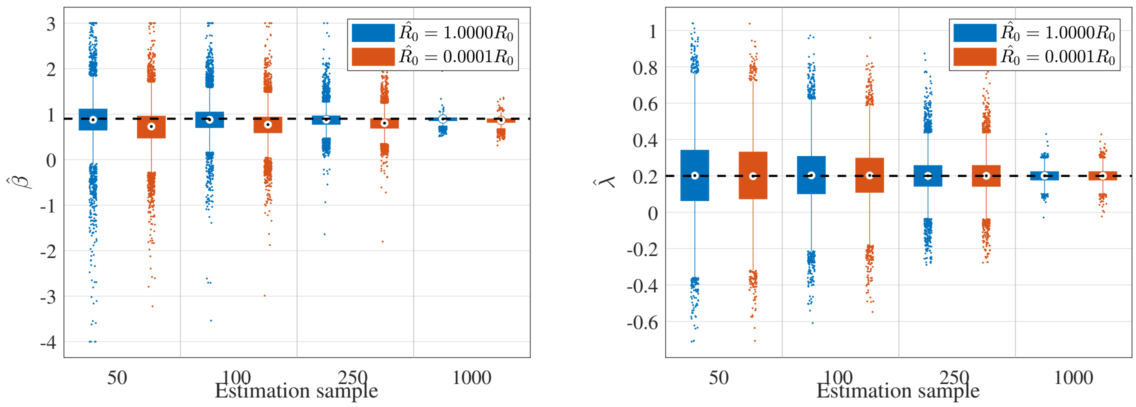
(b) Joint estimates.



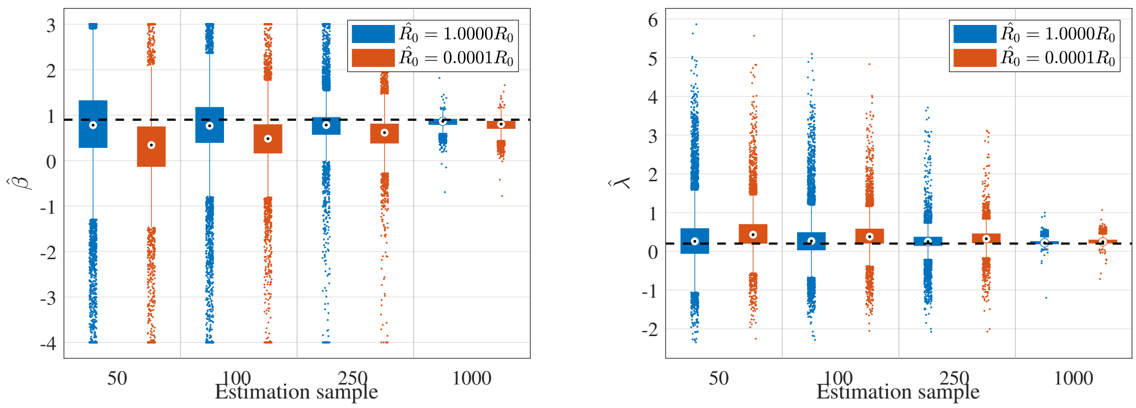
Notes: Same as figures 3 and 4 except that using OLS plus WLS pre-sample estimates for $\mathbf{R}_0$.

Figure 14: Estimates of simulated Phillips curve model with constant-gain learning – with jointly estimated learning gains.

(a) Individual estimates (each jointly with γ).

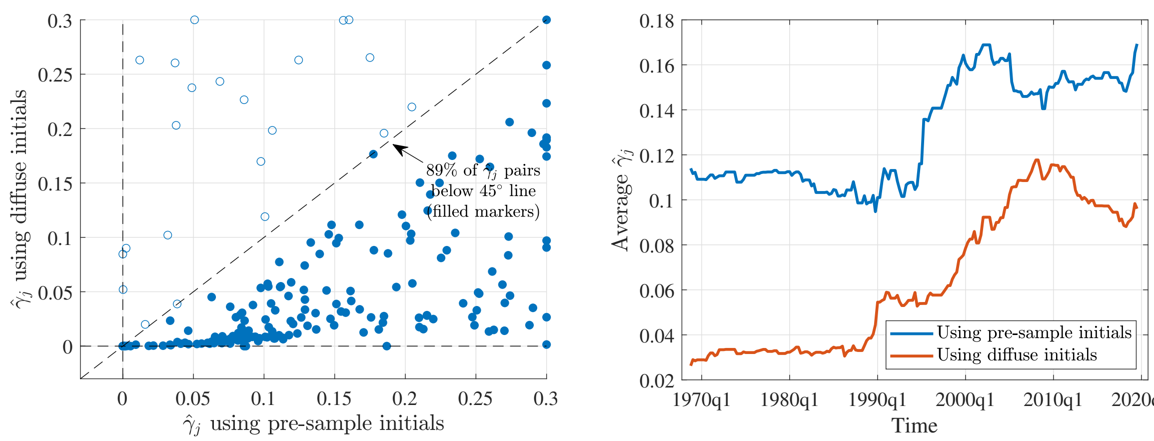


(b) Joint estimates.



Notes: Same as figures 3 and 4 except that jointly estimating the learning gains as in figure 6.

Figure 15: Estimates of individual learning gains from survey forecasts – using OLS pre-sample initials.



Notes: See notes to Figure 9.