Inference in models with adaptive learning, with an application to the new Keynesian Phillips curve

Sophocles Mavroeidis*
Brown University

Guillaume Chevillon†
ESSEC Business School
and CREST-INSEE

Michael Massmann‡
Vrije Universiteit Amsterdam

February 13, 2009§

Abstract

Replacing rational expectations by adaptive learning algorithms complicates the dynamics of economic models. Identification of the structural parameters may improve relative to rational expectations, but it deteriorates when learning converges. Learning also induces persistent dynamics, and this makes the distribution of estimators and test statistics non-standard. We show that valid inference can be conducted using the Anderson-Rubin statistic with appropriate choice of instruments. Application of this method to the new Keynesian Phillips curve with US data provides evidence against standard versions of the model and against least squares learning with a constant gain parameter.

Keywords: Adaptive learning, Weak identification, Anderson-Rubin statistic, Phillips curve
JEL classification: C22, E31

---

*corresponding author: sophocles_mavroeidis@brown.edu.

†chevillon@essec.fr. This author gratefully acknowledges research support from the Economics Department, University of Oxford and a grant by CERESSEC.

‡mmassmann@feweb.vu.nl.

§We would like to thank Jörg Breitung, Norbert Christopeit, Davide Delle Monache, Stefano Eusepi, Stéphane Gregoir, Peter Howitt, Frank Kleibergen, Albert Marcet, Adrian Pagan, Bruce Preston, Tom Sargent, Jim Stock, Harald Uhlig, Mark Watson and Tao Zha, as well as the participants in the NBER summer institute and the Learning & Macroeconomic Policy Conference in Cambridge for helpful comments and discussions. We also benefited from comments received in the NBER/NSF time series conference, several Econometric Society meetings, the NESG and T2M conferences, as well as from seminar participants at Duke, Essec, Maastricht, NYU, OSU, UCL, VU Amsterdam and Harvard-MIT.

1 Introduction

This paper studies inference on structural equations involving expectations that are modeled using adaptive learning. A growing number of studies consider adaptive learning as an alternative to rational expectations (RE), see for instance Sargent (1993), Evans and Honkapohja (2001; 2008), Orphanides and Williams (2004; 2005a), Primiceri (2006), Milani (2005; 2007). Structural models with learning are self-referential and their dynamics are considerably more complicated than the dynamics under RE.[1] As a result, little is known about the properties of structural estimation and inference in these models.

The motivation for studying this problem is twofold. On the one hand, it is well-understood that learning typically induces more persistence in the data than what is implied by models with RE. In fact, one of the motivations for replacing RE with adaptive learning in forward-looking models is to match the dynamics in the data without the need to introduce any intrinsic sources of persistence, which are thought of as ad hoc, see Milani (2005; 2007). On the other hand, it is well-known that forward-looking models suffer from identification problems, see Canova and Sala (2009), Mavroeidis (2005) and Cochrane (2007a; 2007b). Hence, the objective of this paper is to study the implications of those two issues, persistent dynamics and weak identification, for inference on the structural parameters of models with adaptive learning. Our main results can be summarized as follows.

First, we show that identification of structural models is improved under learning relative to rational expectations. The intuition for this result is simple: expectations are more variable under learning than under perfect knowledge, and this improves the accuracy of

[1] Models with Bayesian learning, where non-fully informed agents update their beliefs about the state of the economy using Bayes rule, also induce more complex dynamics than full-information RE, see Schorfheide (2005).

estimators in models where expectations appear as regressors. However, we also find that when the gain of the learning algorithm is decreasing or constant but small, identification becomes weak. The problem can be expressed as near-multicollinearity in regression models, or as 'weak instruments' in models identified by exclusion restrictions. Moreover, it is shown that identification is stronger when the gain parameter is larger. Weak identification invalidates inference using conventional methods, such as Wald statistics, see Stock et al. (2002). However, there is one additional complication which prevents us from using standard identification-robust methods. Learning induces persistence in the data and can cause nearly non-stationary behavior. Thus, methods that rely on normal asymptotic theory become inapplicable.

Second, we show that there is a straightforward and easy-to-implement solution to the problem of inference. In particular, we propose to use a statistic developed by Anderson and Rubin (1949) and popularized recently by the weak instruments literature, with an appropriate choice of instruments, such as lags of the identified structural shocks, so that the required regularity conditions hold. The limiting distribution of the test statistic is χ^2 and does not depend on any nuisance parameters. Simulations show that our proposed method is reliable and useful in the sense that it does not reject the null hypothesis, when it is true, more often than the desired level of significance, and it has nontrivial power when the null hypothesis is false.

Third, since the real test of a method lies in its application to a problem of general interest, we apply our method to study the new Keynesian Phillips curve, a very popular model of inflation dynamics, under learning. The papers most closely related to our empirical study are those by Milani (2005; 2007). We test the baseline model with indexation under constant gain least squares learning using quarterly data on US inflation and the labor share since 1960. Our main conclusion, which is remarkably robust to a wide range of variations in the data, identification assumptions, learning dynamics and estimation sample, is that the baseline specification does not fit the data, in the sense that there is no value of the parameters (including the gain) for which the identifying restrictions hold. We manage to find a version of the model that fits the data when we allow for delayed price changes, as in Woodford (2003, chapter 3). Consistently with Milani, we find that indexation is unnecessary when inflation expectations are formed by some form of adaptive learning. However, unlike Milani, we find that learning with a constant gain parameter does not fit the data, since there is evidence of shifts in the gain parameter over the past fifty years. Specifically, we find that the gain parameter was significantly higher during a period of macroeconomic instability (1974 to 1987) than it was before and after that period.

The paper is structured as follows. Section 2 discusses the problems of inference due to weak identification and persistence in the data, with a textbook example of a model with learning from Evans and Honkapohja (2001). Section 3 introduces our proposed method and provides simulation evidence on its size and power properties in finite samples. Section 4 contains an application of the method to the new Keynesian Phillips curve with adaptive learning. Technical and computational details are presented in an appendix at the end.

2 The problem

To fix ideas, we consider a simple model taken from Evans and Honkapohja (2001, section 14.2):

$$y_t = \beta y_t^e + \delta x_{t-1} + \eta_t \tag{1}$$

where η_t is an innovation process with variance σ_η^2, y_t^e denotes expectations based on information available at time $t-1$, and x_{t-1} is a vector of exogenous and predetermined variables. This is the model studied by Bray and Savin (1986). Evans and Honkapohja (2001) motivate this as a reduced form price equation arising from either a simple cobweb model or the well-known Lucas (1973) aggregate supply model.

Provided $\beta \neq 1$, the unique rational expectations equilibrium (REE) of the model is found to be

$$y_t = \alpha x_{t-1} + \eta_t, \quad \alpha = \frac{\delta}{1-\beta}. \tag{2}$$

Equation (2) describes the law of motion under the REE. Suppose that agents perceive this as the law of motion (PLM) of y_t, but they do not know α. In order to form their forecast y_t^e, they estimate α by a_t using a stochastic recursive algorithm (SRA) with gain sequence $\{\gamma_t\}$. For instance, least squares (LS) is a SRA that can be written recursively as

$$a_t = a_{t-1} + \gamma_t \left(y_t - a_{t-1}x_{t-1}\right) x_{t-1}' R_t^{-1} \tag{3}$$

$$R_t = R_{t-1} + \gamma_t \left(x_{t-1}x_{t-1}' - R_{t-1}\right) \tag{4}$$

for $t = 1, 2, ...$, given some initial conditions a_0 and R_0, and the recursive estimate of α

based on information available at time $t-1$ is a_{t-1}. Two well-studied versions of LS learning are recursive least squares (RLS), obtained from (3) and (4) with $\gamma_t = 1/t$, and constant gain least squares (CGLS) with $\gamma_t = \gamma \in (0, 1)$. The latter is also sometimes referred to as perpetual learning, see e.g. Orphanides and Williams (2005a), and is particularly popular in empirical work. Agents' forecasts are then given by

$$y_t^e = a_{t-1} x_{t-1}. \tag{5}$$

The dynamics of y_t under learning are characterized by the so-called Actual Law of Motion (ALM):

$$y_t = \beta a_{t-1} x_{t-1} + \delta x_{t-1} + \eta_t \tag{6}$$

which is derived by substituting $a_{t-1} x_{t-1}$ for y_t^e in the structural model (1). It is clear that the dynamics of y_t under the ALM in equation (6) are more involved than under the REE, see equation (2). Here, we are interested in the implications of the learning dynamics for inference on the structural parameters β and δ.

In the rest of this section, we simplify the problem of inference by assuming that the SRA is known to the econometrician, which implies that y_t^e is observed. We do so because even under this oversimplifying and generally unrealistic assumption, inference on (β, δ) is difficult and distribution theory for estimators of the structural parameters is non-standard. Our proposed solution, introduced in the following section, does not require that y_t^e be observed, and allows for the parameters in the SRA to be estimated jointly with the structural parameters.

We consider the ordinary least squares (OLS) estimator of (β, δ), which is also the maximum likelihood estimator when the innovations η_t are Gaussian and homoskedastic. Letting $X_t = (y_t^e : x_{t-1})$ denote the regressors in equation (1), the OLS estimator can be written as

$$\begin{pmatrix} \widehat{\beta} - \beta \\ \\ \widehat{\delta} - \delta \end{pmatrix} = \left(\sum_{t=1}^{T} X_t' X_t \right)^{-1} \sum_{t=1}^{T} X_t' \eta_t. \tag{7}$$

Consistency and asymptotic normality of $\left(\widehat{\beta}, \widehat{\delta}\right)$ require that the matrix $\sum_{t=1}^{T} X_t' X_t$, scaled appropriately, should be invertible with probability (approaching) one. This is the rank condition for the identification of β, δ. To establish asymptotic normality at rate $\sqrt{T}$, we need conditions that guarantee $T^{-1} \sum_{t=1}^{T} X_t' X_t$ converges in probability to a non-stochastic and invertible matrix and that the process $T^{-1/2} \sum_{t=1}^{T} X_t' \eta_t$ satisfies a central limit theorem. Under these conditions, the OLS estimators $\left(\widehat{\beta}, \widehat{\delta}\right)$ and the associated t statistics are asymptotically normal, and the Wald statistics are asymptotically χ^2. So, our question of interest is whether these asymptotic results hold and whether they provide a good approximation to the distributions of the statistics in finite samples.

We start by reporting some Monte Carlo simulations on the distribution of OLS estimators and test statistics for the model (1). For simplicity, we make the regressor x_{t-1} in the model a scalar constant, i.e., $x_{t-1} = 1$, and we normalize the true value of the coefficient δ to zero. We simulate the distribution of the OLS estimators and t statistics for β and δ, for samples of size 100, 1000 and 10000 observations. Figure 1 shows the densities of the t statistics under the null for each sample size, and compares those densities to the standard normal asymptotic approximation. It is clear from those pictures that the normal distribution provides a very poor approximation to the sampling distribution of the statistics even for samples of 10000 observations. Non-normality is also evident in the distributions of the estimators of β and δ, but the pictures are omitted for brevity. Note that the non-normality of the distributions does not depend on the initialization of the learning algorithm: the reported results correspond to the case when the SRA is initialized at the REE, α, but they are very similar under alternative initializations.

The above results suggest that there appears to be some convergence to normality under CGLS. So, a relevant question is how large the sample needs to be for the asymptotic approximations to become accurate. One way to address this question is to compare the nominal significance level of a test to its actual rejection frequency and use a particular tolerance level to measure the discrepancy between the asymptotic and finite sample distributions.[2] We consider a 5% level t-test on β and find the smallest sample size that is needed for this test to reject the null hypothesis no more than 10% of the time when it is true. We vary the parameters β and γ within the ranges $[.9, .99]$ and $[.01, .1]$, respectively. The resulting minimum sample sizes are reported in table 1. The main message is that the required sample size increases as β gets closer to unity and γ closer to zero. It is noteworthy that when $\beta = 0.99$ and $\gamma = 0.01$, the required sample size is around 100,000 observations![3]

We now show that these non-standard distributions are the result of identification problems and persistence in the data, taking each explanation in turn.

[2]We thank Ulrich Müller for this suggestion. A similar approach was used by Stock and Yogo (2005) to define weak instruments.

[3]Failure of conventional asymptotic theory at such large sample sizes is not unprecedented in economics. A classic example is found in Bound et al. (1995), who reported problems with the two-stage least squares estimator of the returns to education in a sample of 329,000 observations.

2.1 Identification

It is immediately obvious from equation (2) that the parameters β and δ are not separately identified under rational expectations. One way to see this identification problem is to observe that, under the REE in equation (2), the regressor y_t^e is equal to αx_{t-1}, and hence, it is perfectly collinear with the regressors x_{t-1} that also enters the model. In contrast, under learning, y_t^e becomes $a_{t-1}x_{t-1}$, and this breaks the perfect collinearity with x_{t-1} as long as a_{t-1} varies with t. So, learning *improves* the identifiability of the structural parameters relative to the REE.

The above observation reveals that the identification of the structural parameters β and δ hinges upon the behavior of a_t, which is a well-studied problem in the learning literature. For the simple model (1) with $x_t = 1$, it can be shown that provided $\beta < 1$, agents' estimator a_t converges to α under RLS learning with probability one, see Evans and Honkapohja (2001, Theorem 2.1). Hence, the regressors $y_t^e = a_{t-1}x_{t-1}$ and x_{t-1} in (1) become perfectly collinear in large samples, and this is an example of a phenomenon known in econometrics as near multicollinearity, see, e.g., Judge et al. (1985). In other words, under RLS the identification of the coefficients β and δ breaks down, and this explains the lack of convergence of the OLS estimators shown in Figure 1.

Note that the fact that the PLM coincides with the REE is not necessary for the identification problem just discussed. Near multicollinearity will also arise when a_t converges to some value other than α, as in the case of self-confirming or restricted perceptions equilibria (RPE) under misspecification of the PLM, in the sense that the PLM is strictly nested within the REE, see Evans and Honkapohja (2001, section 3.6).

The conditions under which RLS learning converges to the REE, or to some RPE under misspecification of the PLM, are referred to as E-stability conditions. These are restrictions on the structural parameters, e.g. $\beta < 1$ in equation (1), and regularity assumptions on the process x_t, see e.g. Fourgeaud et al. (1986). Thus, we see that when the structural parameters of the model are not identified under the REE or the relevant RPE, RLS will lead to weak identification when the E-stability conditions hold.

With constant gain learning, it is well-known that a_t does not converge to a non-stochastic limit. Evans and Honkapohja (2001) discuss the behavior of a_t under constant gain learning for a large class of SRAs that includes CGLS as a special case, see Evans and Honkapohja (2001, section 14.2). They show that when the constant gain parameter γ is small and $\beta < 1$, the variance of a_t is proportional to γ. Hence, it is clear that if γ is close to zero, a_t will be approximately constant, and there will be near multicollinearity in equation (1). This situation is in fact empirically relevant, because researchers are often interested in estimating the dynamics of the economy when there are only small departures from rational expectations, i.e., when γ is small; see, e.g., Milani (2007).

When γ is bounded away from zero, the multicollinearity problem disappears. This explains why in Figure 1 there appeared to be convergence under CGLS, since γ was kept fixed as we increased T. In fact, since the variability of a_t is increasing in γ, and since the accuracy with which the coefficients (β, δ) in (1) can be estimated is positively related to the variability of the regressors, other things equal, the parameters will be better identified (i.e., more accurately estimable) the higher γ is. In other words, under constant gain learning, identification improves as the speed of learning decreases.

The regressors in model (1) are exogenous, so the model can be estimated by OLS. In models with endogenous regressors, which are typically identified by exclusion restrictions on a set of instrumental variables, decreasing or small constant gain learning leads to the problem of 'weak instruments', see Staiger and Stock (1997). To see this, consider a variant of model (1) with non-predetermined regressors w_t, such as the New Classical Phillips curve, see Woodford (2003, section 3.1.3):

$$y_t = \beta y_t^e + \delta w_t + \zeta_t. \tag{8}$$

Under the assumption that $E_{t-1}\zeta_t = 0$, the parameters (β, δ) in equation (8) can be estimated by instrumental variables regression, using any predetermined variables as instruments. Identification can be checked using the infeasible second-stage regression that is based on the optimal instruments $E_{t-1}w_t$. The parameters (β, δ) are identified if and only if the regressors in the second-stage regression are not perfectly collinear. Letting $x_{t-1} \equiv E_{t-1}w_t$ and $\eta_t \equiv \zeta_t + \delta\left(w_t - E_{t-1}w_t\right)$, the second-stage regression is given by equation (1). The REE is again given by equation (2), and therefore, there is perfect multicollinearity in the second-stage regression under the REE. Learning breaks the perfect multicollinearity, so instruments become relevant but weak.

It is important to point out that learning does not always improve identification relative to a rational expectations equilibrium. It does so unambiguously only when the model is unidentified under the REE. This was true of the class of models discussed above that involve only contemporaneous expectations and where the REE is unique. This class does not include forward-looking models, such as the NKPC, which can have multiple REEs, and may be identified under some or all of those REEs. Specifically, two leading situations in which learning dynamics may induce weaker identification than a REE are: (i) when the model is identified under the unique REE, but the PLM is misspecified and learning converges to a RPE under which the model is unidentified; and (ii) when there are multiple REEs, and learning converges to a REE under which the model is unidentified. As example of the latter suppose there are multiple REEs indexed by sunspots, and the model is identified under all REEs except that in which there are no sunspot dynamics, and learning converges to that particular equilibrium.

We summarize the above discussion in the following proposition.

Proposition 1 *When the structural parameters of a model are not identified under a particular rational expectations equilibrium, learning dynamics induce identification. However, identification becomes weak when learning converges to that REE or to a restricted perceptions equilibrium, or when it induces only small deviations from that REE or RPE.*

2.2 Persistence

Next, we turn to the issue of persistence of the data under learning dynamics. We first observe that in the simple model (1) the persistence of y_t and y_t^e under the REE (2) is determined solely by the dynamics of the driving process x_t, but learning adds further dynamics to y_t. Thus, we need to examine how much persistence learning generates, and what implications this has for inference on the structural parameters.

We shall focus our discussion on CGLS learning, since it is more relevant empirically than RLS learning. To keep the exposition simple, we discuss only the case in which the regressor x_t is a scalar constant, because in that case, the ALM reduces to a linear time series model, which most readers are familiar with.

When $x_t = 1$ in model (6), the SRA reduces to $a_t = a_{t-1} + \gamma (y_t - a_{t-1})$ with $R_t = 1$. Substituting for y_t using (1) and the fact that $y_t^e = a_{t-1}$, the law of motion for a_t can be written as a first-order autoregression with autoregressive coefficient $1 - (1-\beta)\gamma$ and scale parameter γ :

$$a_t - \alpha = (1 - (1-\beta)\gamma)(a_{t-1} - \alpha) + \gamma \eta_t, \quad t = 1, 2, \ldots. \tag{9}$$

Hence, when the autoregressive root is stable, i.e., when $|1 - (1-\beta)\gamma| < 1$, the process a_t admits a stationary solution and is ergodic. This implies that the asymptotic distribution theory for OLS estimators and Wald tests is standard. Since $\gamma \in [0,1]$ and $\beta \leq 1$, this condition holds provided that $\gamma > 0$ and $1 - 2/\gamma < \beta < 1$. Even though violations of both constraints on β cause standard asymptotic theory to break down, we shall discuss only the upper bound on β, because it is more relevant in practice.

For values of the parameters γ and β that are close to the boundaries of zero and one, respectively, it is evident from equation (9) that a_t follows an autoregressive process with a near unit root. Since it is well-known that distribution theory for nearly integrated autoregressive processes is non-standard (see Phillips, 1987), we expect that this may have an impact on the distribution of OLS estimators and test statistics for the coefficients in equation (1), as well. To approximate the behavior of the regressor $y_t^e = a_{t-1}$ and the OLS estimators defined in equation (7) when $X_t = (a_{t-1} : 1)$, we let γ lie in a neighborhood of zero and β in a neighborhood of one as the sample size grows, i.e., we set $\gamma \in O(T^{-\nu})$ and $1 - \beta \in O(T^{-\omega})$ with $\nu, \omega > 0$. This nesting is such that model parameters are constant in any given sample, but they are allowed to get closer to the boundaries as the sample grows (for notational convenience, we suppress the dependence of β and γ on T in the results given below). This approach is known as local asymptotic approximation, and it can lead to better asymptotic approximations to the finite-sample distributions of statistics that involve persistent data. We follow this approach because it has been used successfully in the econometrics literature to approximate the finite-sample properties of nearly integrated autoregressive processes, see Chan and Wei (1987) and Phillips (1987).

The rates ν and ω define, respectively, the proximity to zero of γ and $1-\beta$ in terms of T and different choices for them give rise to alternative local asymptotic approximations to the behavior of a_t and of the OLS estimators $\left(\widehat{\beta}, \widehat{\delta}\right)$. In the following proposition, we only give the results for the case $\nu = \omega = 1/2$, since this localization was found to give by far the most accurate asymptotic approximation to the finite sample distributions.[4] The symbol "$\Rightarrow$" denotes weak convergence.

Proposition 2 *Consider the stochastic process a_t that satisfies equation (9) with initial condition a_0. Suppose $(1-\beta)\gamma = 1 - e^{\phi/T}$ and $\gamma = \psi/\sqrt{T}$ with $\phi < 0$ and $\psi > 0$, and let $[Tr]$ denote the integer part of Tr, for $0 \leq r \leq 1$. Then, as $T \to \infty$,*

$$a_{[Tr]} \Rightarrow \alpha + e^{\phi r}\left(a_0 - \alpha\right) + \psi \sigma_\eta J_\phi\left(r\right) \equiv K_{\psi,\phi}\left(r\right) \tag{10}$$

where $J_\phi\left(r\right)$ is an Ornstein-Uhlenbeck diffusion with parameter ϕ and $J_\phi\left(0\right) = 0$, driven by the standard Brownian motion $W\left(r\right)$, which is such that $T^{-1/2}\sum_{t=1}^{[Tr]} \eta_t \Rightarrow \sigma_\eta W\left(r\right)$. Moreover, the asymptotic distribution of the OLS estimators $\left(\widehat{\beta}, \widehat{\delta}\right)$ defined in equation (7) with

[4] Results for all other cases of ν and ω are available from the authors on request.

$X_t = (a_{t-1} : 1)$ *is*

$$\begin{bmatrix} \sqrt{T}\left(\widehat{\beta} - \beta\right) \\ \sqrt{T}\left(\widehat{\delta} - \delta\right) \end{bmatrix} \Rightarrow \begin{bmatrix} \int_0^1 K_{\psi,\phi}^2(r)\,dr & \int_0^1 K_{\psi,\phi}(r)\,dr \\ \int_0^1 K_{\psi,\phi}(r)\,dr & 1 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_\eta \int_0^1 K_{\psi,\phi}(r)\,dW(r) \\ \sigma_\eta W(1) \end{bmatrix}. \quad (11)$$

In the above result, the parameters ϕ and ψ measure, respectively, the distance of the autoregressive root from unity and of the gain parameter from zero, relative to the sample size. The Ornstein-Uhlenbeck diffusion is a continuous time autoregressive process whose persistence is inversely related to ϕ, where the limiting case $\phi = 0$ corresponds to a random walk. Proposition 2 therefore shows that the persistence in a_t is increasing the closer is $(1-\beta)\gamma$ to zero.

Regarding the asymptotic distribution of the OLS estimators, we see that it is non-normal. This is because the second moment matrix of the regressors, $T^{-1}\sum_{t=1}^{T} X_t' X_t$ in equation (7), does not converge to a non-stochastic limit, and the sample moment conditions involving the persistent regressor, $T^{-1/2}\sum_{t=1}^{T} a_{t-1}\eta_t$, do not satisfy a standard central limit theorem. In the special case $\alpha = a_0 = 0$, the distribution of the OLS estimators given by the right-hand side of equation (11) corresponds almost exactly to the local-to-unit root approximation in the model considered by Phillips (1987) and the resulting distributions are of the Dickey-Fuller type. Yet, contrary to the pure unit-root case, $\widehat{\beta}$ does not converge faster than at rate $\sqrt{T}$. This is because of the dampening effect of a vanishing γ on the variance of the regressor a_{t-1}, which prevents it from exhibiting a stochastic trend.

Figure 2 shows that the local asymptotic distribution given by the right-hand side of expression (11) provides a very accurate approximation to the finite sample distribution of the OLS estimators (7) for a sample of size $T = 100$ and for $\beta = 0.99$ and $\gamma = 0.02$, while the standard fixed-parameter asymptotic approximation is poor. The approximation is also very good for other values of β and γ, the results being omitted for brevity.

3 Robust inference

3.1 The Anderson-Rubin statistic

The previous section showed that weak identification and persistence of the regressors in models with learning render inference using conventional test statistics, such as the Wald statistic, unreliable. In this section, we propose a test statistic whose asymptotic distribution under the null is χ^2 without any assumptions on identification or weak dependence in the data. Hence, inference based on this statistic is fully robust to violations or near violations of these conditions. The proposed method is an application of the Anderson and Rubin (1949) statistic, which has been recently revived by the weak instruments literature, see Dufour (1997) and Staiger and Stock (1997). The exact Anderson-Rubin ($\mathcal{AR}$) test applies to a linear instrumental variable (IV) model with strongly exogenous instruments and Gaussian independently and identically distributed (i.i.d.) data. However, Stock and Wright (2000) extended it to nonlinear models with dependent and heterogeneous data that are estimable by the generalized method of moments (GMM), under mild regularity conditions. Here we show how to obtain versions of the $\mathcal{AR}$ statistic for which the regularity conditions in Stock and Wright (2000) can be verified for models with learning. For a detailed description of the $\mathcal{AR}$ statistic, the reader is referred to the excellent surveys of Stock et al. (2002), Dufour

(2003) and Andrews and Stock (2005).

To explain our proposed method of inference, it is useful to illustrate the basic principle of the $\mathcal{AR}$ test in its original setting, the prototypical linear IV regression model. Suppose we are interested in testing the null hypothesis $H_0 : \theta = \theta_0$ on the parameters of the model $y_t = Y_t\theta + u_t$, where the regressor Y_t is endogenous and u_t is a disturbance term, and suppose there exists an exogenous vector of instruments Z_t such that $EZ_t u_t = 0$. The principle of the $\mathcal{AR}$ test is not to test H_0 directly, but rather take a somewhat lateral approach by testing the exclusion restrictions that are implied by H_0. Since, under H_0, the disturbance term is observed, $u_t^0 \equiv y_t - Y_t\theta_0$, the $\mathcal{AR}$ test can be obtained as the usual F test of the exclusion of Z_t in the auxiliary regression of u_t^0 on Z_t. Moreover, the distribution of the $\mathcal{AR}$ statistic does not depend on the correlation between the regressors and the instruments, i.e. on whether the parameters θ are identified or not, and is therefore fully robust to weak instruments.[5]

The models with learning that we consider can be expressed in the form $h(\mathcal{Y}_t; \theta) = \eta_t$, where $\mathcal{Y}_t$ denotes the observed data and η_t is an unobserved disturbance, which could be a vector when the model consists of several equations. It is important to stress that the parameter vector θ includes both the structural parameters of the model and the parameters of the SRA that characterizes learning dynamics, such as the gain parameter under CGLS. Identifying assumptions are usually placed on the dynamics of the disturbance term, which is typically interpreted as a structural shock. A common assumption is that η_t is a martingale difference sequence (MDS), such that $E_{t-1}\eta_t = 0$. Based on this over-identifying assumption, we can identify the parameters by the moment conditions $EZ_t h(\mathcal{Y}_t; \theta) = 0$, for

[5] When the model is just-identified, the $\mathcal{AR}$ test even has certain optimality properties, see Andrews and Stock (2005).

any predetermined instruments Z_t.

Consider now the problem of testing the hypothesis $H_0 : \theta = \theta_0$. Note that the disturbances of the model under H_0, denoted η_t^0, can be computed from the function $h(\mathcal{Y}_t, \theta_0)$. Therefore, the $\mathcal{AR}$ statistic for H_0 can be obtained as the Wald statistic for testing the exclusion of Z_t in the auxiliary regression of η_t^0 on Z_t:[6]

$$\mathcal{AR}(\theta_0) = \frac{1}{T}\left(\sum_{t=1}^{T} \eta_t^{0\prime} Z_t'\right) \widehat{V}_{Z\eta}^{-1} \left(\sum_{t=1}^{T} Z_t \eta_t^0\right) \tag{12}$$

where $\widehat{V}_{Z\eta}$ is an estimator of the asymptotic variance of $T^{-1/2}\sum_{t=1}^{T} Z_t\eta_t^0$, such as White's (1980) heteroskedasticity consistent estimator, which is consistent under the assumption $E_{t-1}\eta_t = 0$ and some additional mild regularity conditions, see Nicholls and Pagan (1983).

When the sample moments $T^{-1/2}\sum_{t=1}^{T} Z_t\eta_t^0$ satisfy a central limit theorem, the distribution of $\mathcal{AR}(\theta_0)$ is, in large samples, $\chi^2(k)$ under H_0, where k is the number of moment conditions. Stock and Wright (2000) provide sufficient primitive conditions to establish this result, but when Z_t is highly persistent these conditions may not hold. For example, in the model we studied in section 2, we saw that limit theory involving the persistent regressor y_t^e is nonstandard, and this has an impact on the $\mathcal{AR}$ statistic as well.[7] Therefore, to avoid having to work out special asymptotic theory for the $\mathcal{AR}$ statistic in every application, we need to use instruments that ensure that the $\mathcal{AR}$ statistic is asymptotically χ^2 under the null hypothesis. This includes predetermined variables that are weakly dependent, but it

[6] When η_t^0 is a vector, i.e., when the model consists of a system of structural equations, Z_t must be defined as a matrix whose dimension is comformable to η_t^0, and the auxiliary regression is a system of seemingly unrelated regressions (SUR).

[7] If we were to use $y_t^e = a_{t-1}$, which is predetermined, as an instrument for equation (1), then $T^{-1/2}\sum_{t=1}^{T} Z_t\eta_t \Rightarrow \sigma_\eta \int_0^1 K_{\psi,\phi}(r)\, dW(r)$, which is non-normal, see Proposition 2.

excludes lags of the endogenous variable y_t or its forecast y_t^e, which depend on the recursive estimates a_t, and may therefore be highly persistent.

The key idea behind the solution we propose is the observation that the lags of η_t are valid instruments, and that, since η_t is a MDS, the asymptotic normality of $T^{-1/2}\sum_{t=j}^{T}\eta_t\eta_{t-j}$ can be established under mild conditions based on standard limit theory, see e.g., Hamilton (1994) or White (1984).[8] So, by suitable selection of instruments and the use of the $\mathcal{AR}$ statistic, we have turned a difficult problem into a trivial one.

The above principle can be generalized to cover alternative assumptions on the time dependence of the disturbances, η_t. In particular, suppose we wish to relax the assumption that the disturbance η_t is an MDS. Clearly, some structure must be placed on the autocorrelation of η_t for the model to be identified, as it is the case for any dynamic model that does not contain only strictly exogenous regressors, see, for example, Cochrane (2007a) and Beyer and Farmer (2007). For instance, suppose we wish to replace $E_{t-1}\eta_t = 0$ by the weaker condition that the shock is autoregressive of order 1, denoted AR(1), which is common in applied work: $\eta_t = \kappa\eta_{t-1} + \varepsilon_t$, where $E_{t-1}\varepsilon_t = 0$. The $\mathcal{AR}$ statistic (12) can be easily adapted to this alternative specification as follows. We first obtain $\eta_t^0 = h\left(\mathcal{Y}_t, \theta_0\right)$ as before, we then regress it on its lags $\eta_{t-1}^0, ..., \eta_{t-m}^0$ and any n additional instruments $z_{2,t}$, and we finally derive the Wald statistic for testing the hypothesis that all the coefficients, except that on η_{t-1}^0, are equal to zero. Under mild assumptions (e.g., $\varepsilon_t, z_{2,t}$ are stationary and ergodic with finite fourth moments and $|\kappa| < 1$), the asymptotic distribution of this statistic is $\chi^2(k)$, with $k = m + n - 1$.

[8] This idea is motivated by the recent work of Gorodnichenko and Ng (2007), who used a similar approach to develop methods of inference that are robust to misspecification in the way in which data has been detrended.

Confidence sets for the parameters θ with confidence level φ can be obtained by 'inverting' the $\mathcal{AR}$ test in the usual way. This involves testing every value of $\theta_0 \in \Theta$ by comparing $\mathcal{AR}(\theta_0)$ to the φ quantile of the $\chi^2(k)$ distribution, or alternatively, by comparing the p-value associated with $\mathcal{AR}(\theta_0)$ to $1-\varphi$, and collecting all the values of θ_0 that are not rejected by this test. This can be achieved by grid search over the parameter space Θ. The computational cost of this procedure may appear to be very high when there are several parameters, since the number of computations rises exponentially in the dimension of θ for any degree of accuracy. However, in practice, joint confidence sets on the entire parameter vector are rarely of interest. One-dimensional confidence intervals, or confidence sets on subsets (e.g., pairs) of the parameters can be straightforwardly obtained by projection methods without computing the joint confidence set for the entire parameter vector, see the Appendix for details.

The minimum value of the $\mathcal{AR}$ statistic, $\min_\theta \mathcal{AR}(\theta)$, serves as a formal test of the fit of the model to the data. When $\min_\theta \mathcal{AR}(\theta)$ exceeds the critical value given by the φ quantile of the $\chi^2(k)$ distribution, there is no value of the parameters $\theta \in \Theta$ for which the exclusion restrictions are consistent with the data, and hence the φ-level confidence set for θ is empty. In that case, we conclude that the model does not fit the data at the $(1-\varphi)$ level of significance, see Stock and Wright (2000).

Finally, the minimizer $\widehat{\theta}$ of $\mathcal{AR}(\theta)$ can be used as a point estimate for θ, because it has an interesting interpretation: it is the "least-objectionable" or "least-rejected" value of the parameters under the $\mathcal{AR}$ criterion, since it is the value that results in the highest p-value associated with the $\mathcal{AR}$ statistic. Moreover, since $\mathcal{AR}(\theta)$ is also a continuously updated GMM objective function, $\widehat{\theta}$ can be seen as a Continuously Updated Estimator (CUE), see

Stock and Wright (2000). Note that $\widehat{\theta}$ is also a Hodges-Lehmann estimator, see Hodges and Lehmann (1963).

3.2 Simulations

We evaluate the finite sample size and power properties of the proposed $\mathcal{AR}$ statistic and compare it to the Wald statistic, using simulations of the hybrid NKPC model of inflation, i.e., the model that we use in our empirical application in Section 4. The present section provides a brief summary of our simulation results which illustrate that our proposed method works well. Extensive simulation results are available on request.

The hybrid NKPC model with indexation takes the form

$$\pi_t = \frac{\beta}{1+\beta\varrho}\pi^e_{t+1} + \frac{\varrho}{1+\beta\varrho}\pi_{t-1} + \frac{\lambda}{1+\beta\varrho}s_t + \frac{1}{1+\beta\varrho}\eta_t, \tag{13}$$

where π_t denotes inflation, s_t is an observable forcing variable, and η_t is a MDS. Details of the model are given in the next section. The observable forcing variable s_t is assumed to follow a second-order autoregressive process $s_t = \rho_1 s_{t-1} + \rho_2 s_{t-2} + v_t$, where the shocks η_t, v_t are independently and jointly normally distributed with zero mean, and second moment $E\eta_t^2 = \sigma_\eta^2$, $E\eta_t v_t = \sigma_{\eta v}$ and $Ev_t^2 = 1$. The rational expectation of π_{t+1} is given by

$$E\left(\pi_{t+1}|\pi_{t-1}, ..., s_t, s_{t-1}, ...\right) = \alpha_1\pi_{t-1} + \alpha_2 s_t + \alpha_3 s_{t-1} \tag{14}$$

where the parameters $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ are functions of the structural parameters, see Mavroeidis (2005). We assume the PLM is given by (14), and α is estimated by CGLS, so that π^e_{t+1} is given by:[9]

$$\pi^e_{t+1} = a_{1,t-1}\pi_{t-1} + a_{2,t-1}s_t + a_{3,t-1}s_{t-1} = a_{t-1}x_t.$$

Our information assumptions and the fact that $E_{t-1}\eta_t = 0$ imply that the parameters of equation (13) can be estimated by two stage least squares (2SLS) using predetermined variables as instruments. For the Wald statistic, we use the first two lags of π_t and s_t as instruments, while the $\mathcal{AR}$ statistic is computed using two lags of η_t and s_t as instruments. We also allow for an unrestricted constant in the estimation and we impose the restriction that β is known, as is common in applied work. With this restriction, the NKPC can be written in the following linear form $y_t = \varrho w_t + \lambda s_t + \eta_t$ where $y_t = \pi_t - \beta\pi^e_{t+1}$ and $w_t = \pi_{t-1} - \beta\pi_t$ are both endogenous regressors. The parameter values in the DGP are chosen so as to be representative of the estimates reported in the literature, e.g., Galí and Gertler (1999), while the parameters of the forcing variable s_t are calibrated to US data, see Mavroeidis (2005) for details.

We first study the coverage probabilities of confidence intervals derived by inverting the Wald and $\mathcal{AR}$ tests. Table 2 displays the actual coverage probabilities for the Wald test of $H_0 : \varrho = \varrho_0$ at nominal levels φ of 75%, 90%, 95% and 99%. For simplicity, we assume λ is known. The $\mathcal{AR}$-based confidence sets have exact coverage, that is, a φ-level confidence interval for ϱ contains the true value with probability φ, with only slight distortions in small samples. The Wald always undercovers, which means that the usual standard error bands

[9] In excluding π_t from the information set used to forecast π_{t+1} we follow the vast majority of the literature, in order to avoid the simultaneity induced by having π_t on both sides of the model (13). This informational assumption is used to simplify the simulations, and it actually makes no difference to the empirical results on the NKPC reported later.

around the point estimate are too tight.

Figure 3 shows the power curves of the Wald and $\mathcal{AR}$ tests of the hypothesis $H_0 : \varrho = \varrho_0$ at the 5% nominal level of significance for sample sizes $T = 100$ and 200. As we explain in the next section, the parameter ϱ measures the degree of indexation of prices to past inflation. It is evident that the $\mathcal{AR}$ test has good power, especially over the theoretically relevant parameter regions. In particular, it rejects with high probability the null hypothesis of $\varrho = 0$ (no indexation) when indexation is substantial, and it thus can provide reliable evidence on this issue of considerable interest in applied work. The $\mathcal{AR}$ test does not have good power for high values of ϱ against higher alternatives. This means the test has difficulty distinguishing between a high degree of indexation and complete indexation.

Finally, as we discussed in section 2.1, higher values of the gain parameter, which are interpretable as a slower speed of learning, generate more variability of the adaptive forecasts relative to the corresponding RE forecast, and we expect this to have a positive impact on the accuracy of inference on the structural parameters. Figure 4 gives evidence of this effect through a direct comparison of the power function of the $\mathcal{AR}$ statistic at different values of γ and T. Specifically, the figure depicts the contour plots of the power function for the null hypothesis $H_0 : \varrho = 0$ against four alternatives, with respect to γ and T. It is clear that power increases in γ as well as T, and, moreover, that γ plays a role similar to the sample size, in that the same power can be achieved with a lower value of T and a higher value of γ, or vice versa.

4 Application: the new Keynesian Phillips curve

4.1 Model and empirical issues

The NKPC is a purely forward-looking model of inflation dynamics which takes the form

$$\pi_t = \beta E_t \pi_{t+1} + \lambda \widehat{s}_t + \eta_t \tag{15}$$

where π_t denotes inflation, the forcing variable $\widehat{s}_t$ is a measure of real marginal costs in deviation from their steady state and η_t is an unobserved disturbance. The deep parameters of the model are the discount factor β and the degree of price stickiness ϑ, which is the probability that a firm will be unable to change its price in a given period. The slope of the Phillips curve λ is a function of β and ϑ, namely $\lambda = (1-\vartheta)(1-\vartheta\beta)/\vartheta$, see Galí and Gertler (1999).

Many studies report difficulties in fitting model (15) to US data when expectations are modeled as rational; see, for instance, Fuhrer and Moore (1995), Galí and Gertler (1999), Rudd and Whelan (2005, 2006). In particular, the model predicts that the dynamics of inflation should be explained solely by the dynamics of marginal costs, but this does not turn out to be the case with US data. In response to this empirical failure, the purely forward-looking specification in (15) has been extended to include lagged inflation on the right hand side. For example, assuming that a fraction ϱ of prices that cannot be re-optimized are indexed to past inflation, the model becomes

$$\pi_t = \beta E_t (\pi_{t+1} - \varrho \pi_t) + \varrho \pi_{t-1} + \lambda \widehat{s}_t + \eta_t. \tag{16}$$

see Woodford (2003, section 3.2) for details. Notably, the hybrid NKPC (16) nests the pure model (15) when $\varrho = 0$. The other polar case of complete indexation, $\varrho = 1$, is often used in empirical studies, see, e.g., Christiano et al. (2005).[10] Another common approach to introducing additional dynamics to the pure NKPC model (15) is to suppose that the unobservable cost push shock η_t is autocorrelated, as in Clarida et al. (1999).[11] Yet another source of inflation persistence are time delays in the introduction of new prices. With a delay of d quarters, the NKPC (16) becomes

$$\pi_t = \beta E_{t-d} \left(\pi_{t+1} - \varrho \pi_t \right) + \varrho \pi_{t-1} + \lambda E_{t-d} \widehat{s}_t + \eta_t \tag{17}$$

see Woodford (2003, p. 217).

Recently, Milani (2005, 2007) argued that if the assumption of rational expectations is replaced with some form of boundedly rational expectations, the pure NKPC model fits the data without the need to make potentially ad hoc assumptions to generate additional sources of persistence. Specifically, if inflation expectations are formed recursively by CGLS, they will depend on past data more than they would under rational expectations. Using least-squares and Bayesian methods, Milani found that indexation is not statistically significantly different from zero when expectations are formed by CGLS learning.

Since our focus is primarily on the fit of the NKPC, we take a limited-information approach, following Galí and Gertler (1999), Sbordone (2002) and Milani (2005). As Woodford

[10] Galí and Gertler (1999) provided an alternative derivation of the NKPC, based on the idea that some fraction of firms set prices according to a backward-looking rule of thumb. As Woodford (2003, p. 217) notes, the two models have identical implications in the limiting case $\beta = 1$.

[11] There are also studies that use both indexation and autocorrelated shocks, e.g., Smets and Wouters (2007).

(2003) explains, this approach makes weaker assumptions than full-information methods about the determinants of marginal costs, and is therefore more robust to misspecification of other parts of the system. Moreover, unlike Bayesian inference, our approach does not require the specification of the distribution of the shocks, or any priors on the parameters. Of course, weaker assumptions imply fewer identifying restrictions, and thus robustness comes at the cost of lower accuracy of inference. However, our results show, in line with the simulation evidence reported earlier, that our limited-information analysis is powerful enough to uncover new evidence on the empirical fit of the NKPC under learning, and our tests are highly informative about certain parameters of the model.

In our analysis we fix the discount factor β to 0.99. This assumption simplifies inference on the other three key parameters of the model, viz. (ϱ, ϑ) and the gain parameter, which are the more interesting ones. We note that our results remain robust for other values of β commonly advocated in the literature.

Data Our estimation results are based on quarterly US data that cover the period 1960:Q1 to 2007:Q3. Following Galí and Gertler (1999) and Sbordone (2002), we derive our measure of $\widehat{s}_t$ assuming it is proportional to the log of the labor share.[12] The factor of proportionality depends on assumptions about factor markets and is typically calibrated. We set it to the value used by Galí and Gertler (1999), following the method of Sbordone (2002), so that our results are comparable to theirs. Inflation is measured by the first difference in the logarithm of the implicit GDP deflator, which is only available in seasonally adjusted form. An alternative inflation measure, based on the Consumer Price Index (CPI), is used for

[12] For the labor share, we use the data reported by the Bureau of Labor Statistics (series ID: PRS85006173).

robustness checks. The Federal Funds rate is used as an additional instrumental variable.

4.2 Results for the baseline specification

Our baseline specification is the hybrid NKPC (16) with expectations determined by perpetual learning, that is, CGLS. To close the model, we need to specify the PLM agents use to derive their forecasts. We assume the PLM is a vector autoregression of order p, or VAR(p), in inflation and the labor share. This nests the simple specification of a first order autoregression for inflation, used, amongst others, in Milani (2005) and Orphanides and Williams (2005b). Also, under certain conditions on the law of motion of the labor share, it nests the rational expectations equilibrium, as explained in Evans and Honkapohja (2001). This approach is also common in the literature, see Bullard and Eusepi (2005), Milani (2007) and Orphanides and Williams (2005a). In our baseline specification, we set the order of the VAR to be $p = 1$, based on the Schwarz criterion, but we also investigate the robustness of the results to larger values of p.

Agents estimate the coefficients of the PLM recursively using the CGLS with gain parameter γ. The computation of the $\mathcal{AR}$ statistic requires the specification of the initial conditions of the learning algorithm. Following common practice in the literature (e.g., Milani 2007), we calibrate these to pre-sample data, starting in 1955:Q1.[13] Carceles-Poveda and Giannitsarou (2007) discuss alternative initialization schemes and argue that the choice of initialization is unimportant for CGLS. However, since our inference is conditional on the calibrated values of the initial conditions, we investigate the robustness of our results to

[13]Specifically, we use the formulae in Carceles-Poveda and Giannitsarou (2007, Equations 98 and 99).

alternative calibrations, and we find that our results are indeed robust.[14]

Our identification assumption in the baseline model is that the disturbance term η_t is uncorrelated with its own lags and any other predetermined variables. This is identical to the assumption used in Galí and Gertler (1999), Sbordone (2002) and Milani (2005),[15] but it can be easily relaxed to allow for exogenous persistence in the cost push shock, e.g., an autoregressive process of order q. We investigate the robustness of the results to these alternatives.

We collect the parameters of interest in a vector $\theta = (\vartheta, \varrho, \gamma)'$ and use the notation $\pi^e_{t+1}(\gamma)$ to denote explicitly the dependence of π^e_{t+1} on γ. We allow the parameters ϑ and ϱ to take values that are consistent with the underlying theory, namely $0 \leq \vartheta \leq 1$ and $0 \leq \varrho \leq 1$. In principle, γ could take any value between zero and one but we put an upper bound at 0.1. This is motivated by the assumptions in Milani (2007), but our results are robust to using higher upper bounds on γ.[16] We exclude zero, since π^e_{t+1} cannot be computed at zero using the CGLS algorithm, as explained in Carceles-Poveda and Giannitsarou (2007), though it can be argued that, provided that $\beta < 1$, small values of γ can be interpreted as small deviations of $\pi^e_{t+1}(\gamma)$ from the rational expectation of π_{t+1}; see Milani (2007) or Orphanides and Williams (2005b).

We compute the $\mathcal{AR}$ statistic at θ using formula (12) corrected for an unrestricted con-

[14] There are alternative approaches to the treatment of the initial conditions, such as viewing them as estimable parameters, or as random variables drawn from some distribution. We investigate these alternatives in related work.

[15] In their estimation, Galí and Gertler (1999) assumed rational expectations and replaced π^e_{t+1} by its realization π_{t+1}, thus causing the residual in the estimated model to be a moving average of order 1. This is consistent with the disturbance in the NKPC being serially uncorrelated.

[16] Milani's prior distribution restricts the gain parameter to be less than 0.1 with probability 0.999. Most other studies fix or calibrate the gain parameters to values well below 0.1, and typically around 0.02. Our results are robust to using an upper bound of 0.2.

stant, where the residuals η_t are given by the following function of the data and the parameters:

$$h_t(\theta) = \pi_t - \beta\pi^e_{t+1|t}(\gamma) - \varrho(\pi_{t-1} - \beta\pi_t) - \frac{(1-\vartheta)(1-\beta\vartheta)}{\vartheta}\widehat{s}_t. \tag{18}$$

We employ four lags of the residuals, of the labor share and of the Federal Funds rate as instruments, and use White's (1980) heteroskedasticity consistent estimator for the variance in the $\mathcal{AR}$ statistic, for robustness against potential heteroskedasticity.[17]

We start by assessing the fit of the baseline model (16). As we explained in section 3, the p-value corresponding to the minimum value of $\mathcal{AR}(\theta)$ serves as a measure and formal test of the model's fit, at any desired level of significance, while the minimizer of $\mathcal{AR}(\theta)$ is the least-rejected or best-fitting value of the parameters. Our main finding is that the baseline model does not fit the data. The p-value associated with the least rejected value of θ is 0.0001, indicating rejection of the model at the 0.01% level of significance. The failure of the baseline model to match the data is robust to alternative specifications of π^e_{t+1}. Table 3 reports such robustness checks and shows that, whether π_t is included in the forecast of π_{t+1} or not, and whether higher order VAR specifications are used in the PLM, the p-value of the minimum $\mathcal{AR}$ statistic is very small, remaining almost always below the 1% level of significance. As a further robustness check against potential misspecification of agents' forecasting model, we report the fit of the baseline model when π^e_{t+1} is measured using data from the Survey of Professional Forecasters, or real-time Greenbook forecasts.[18] We also consider the case in which survey data enter the PLM as an additional predictor, in order to

[17]Heteroskedasticity may be induced by, amongst other things, time-variation in the volatility of the shocks, given the evidence reported in the literature, see Sims and Zha (2006).

[18]The Greenbook data were kindly provided by Athanasios Orphanides, and are the data used in Orphanides (2004). The survey data are the median one-year-ahead forecasts of inflation compiled by the Philadelphia Fed.

incorporate additional information that agents may have used and that is not captured by a VAR. The results reported in Table 3 show that the baseline model remains resoundingly rejected. Our conclusion is also robust to using the CPI-based measure of inflation.

A possible cause of misspecification is variation in the parameters of the model over time, such as a structural break. Parameter instability will result in a violation of the moment conditions of a model that incorrectly assumes the parameters are constant over the entire sample. Perhaps the simplest way to account for variation in the parameters is to check the fit of the model over subsamples. Looking at various subsamples, the results remain the same. In all cases the model is rejected at well below the 5% level.

The bad fit of the model is also manifest in residual autocorrelation. Figure 5 plots the correlogram of the residuals of the baseline model for lags up to twelve quarters. It is apparent that the residuals exhibit autocorrelation, especially at lag four. Moreover, the structure of autocorrelation is such that it cannot be captured by modelling the shocks as AR(1). Table 4 reports tests of the fit of the model with alternative assumptions about the autocorrelation of η_t in the NKPC (16): the $\mathcal{AR}$ statistic remains significant at the 1% level unless η_t is modelled as AR(4). Autoregressive moving average (ARMA) models for the shock η_t have been used as alternatives to the autoregressive specification; see Smets and Wouters (2007). However, a parsimonious specification, such as an ARMA(1,1), would not be flexible enough to capture the serial correlation pattern of the residuals.[19] The type of model for η_t that would be required to do so is not one that we have seen used in applied work. In the next subsection, therefore, we consider time delays in price changes as an alternative source

[19]Note that Smets and Wouters (2007) use a different version of the NKPC such that their specification of the shocks is not directly comparable to ours.

of persistence in inflation, see equation (17), which is more appealing from a theoretical perspective than arbitrary assumptions about the autocorrelation of the errors.

4.3 Results for a model with time delays in price changes

We now turn to the NKPC with indexation and time delays of d quarters, which is given by equation (17). First, we observe that this model is still unable to fit the data over the full sample, see the estimates reported in Table 5. However, it turns out that a model with $d = 4$ and AR(1) shocks fits the data when the parameters are allowed to be different over subperiods. In fact, a model estimated over the entire sample that allows for changes only in the gain parameter suffices to fit the data. The gain parameter is interpretable as measuring the speed of learning, because higher gains are associated with faster discounting of past data in the estimation of the PLM. One may expect to see higher discounting of past observations over periods of instability, during which the use of constant parameter reduced-form models for forecasting is susceptible to the Lucas (1976) critique.

The above considerations motivate us to consider three periods of similar length across which the gain parameter may be different: 1960:Q1-1973:Q3, 1973:Q4-1987:Q3 and 1987:Q4-2007:Q3. The second period starts at the onset of the first oil price shock, covers the great inflation of the seventies and the subsequent disinflation of the early eighties, and ends in 1987:Q3, when Greenspan became the chairman of the Fed. The first and third periods are characterized by relative macroeconomic stability. The gain parameter is allowed to be different across periods, but constant within each period:

$$\gamma_t = \begin{cases} \gamma_1, & \text{before 1973:Q4} \\ \gamma_2, & \text{1973:Q4 to 1987:Q3} \\ \gamma_3, & \text{after 1987:Q3.} \end{cases} \tag{19}$$

We estimate the NKPC in (17) with indexation, a time delay of four quarters, $d = 4$, and an AR(1) shock, allowing for breaks in the gain parameter as in (19). The p-value associated with $\min \mathcal{AR}(\theta)$ is 0.31, indicating non-rejection of this specification at conventional significance levels. Thus, confidence sets on θ derived by inverting the $\mathcal{AR}$ statistic are non-empty at conventional levels. In fact, we find that the estimates of the gain parameters in the pre-1974 and post-1987 periods, $\widehat{\gamma}_1$ and $\widehat{\gamma}_3$, are identical and testing the restriction that $\gamma_1 = \gamma_3$ using the $\mathcal{AR}$ statistic accordingly yields a p-value of 0.31, so we impose this restriction in the ensuing analysis. The results are reported in Figure 6 and Table 6.

As expected, the gain parameter is different across the three periods. Specifically, the gain parameter is high during the volatile period of 1974 to 1987, and is small before and after that period. Importantly, the difference between the estimates $\widehat{\gamma}_1 = \widehat{\gamma}_3$ and $\widehat{\gamma}_2$ is statistically significant at the 5% level, with the p-value associated with the hypothesis $\gamma_1 = \gamma_2 = \gamma_3$ being 0.026. This is also evident from the confidence set on (γ_1, γ_2) reported in the left panel of Figure 6, the shaded areas of which contain φ-level confidence sets for (γ_1, γ_2) derived by the projection method. Even though the confidence sets are wide, showing that the gain parameters cannot be estimated very precisely, they are still informative about a break in γ, since the 95%-level confidence set does not cross the 45° line.

We now look at the estimates of the two structural parameters ϑ and ϱ in Table 6 and in the right panel of Figure 6. Consistently with Milani (2007), the indexation parameter ϱ is not significantly different from zero. Its point estimate of 0.13 is low compared to other studies, but the 95% confidence interval associated with it is wide, viz. $[0, 0.46]$. Still, the estimates are precise enough for a model with full indexation to be rejected at the 5% significance level. The parameter governing the degree of price stickiness, ϑ, is notably very imprecisely estimated. The point estimate of 0.62 implies that prices remain fixed for two and a half quarters on average. However, virtually all the parameter estimates reported in the literature fit within the 90% confidence interval $[0.4, 1]$, and we cannot reject the null hypothesis that the Phillips curve is completely flat, $\vartheta = 1$.

The above inference can be sharpened considerably by using more instruments, without altering any of the conclusions reached. When we use two additional lags of the residuals of the model as instruments, the confidence sets and intervals are generally tighter, and we can reject the null hypothesis of a constant gain across periods at the 0.05% level of significance. Only the confidence interval for the price stickiness parameter ϑ does not get much tighter, indicating that this parameter is rather weakly identified. In fact, this result is similar to the evidence obtained on the degree of price stickiness under rational expectations, see Kleibergen and Mavroeidis (2008).

4.4 Discussion

One important message of our analysis is that standard versions of the NKPC are unable to fit the dynamics in inflation, and this is as true under learning as it is under rational expectations, the latter shown, amongst others, by Rudd and Whelan (2005, 2006). In particular, the model fails to account for the pattern of autocorrelation in inflation. To improve the model's ability to fit the observed dynamics in inflation, we considered a simple extension of the model that allows for time delays in price changes. Time delays in price changes may be a relevant feature of most markets, but a delay of four quarters, which is needed for the NKPC to fit the data, may seem unrealistically long. It seems plausible that such dependence is the result of wage contracts being negotiated on an annual basis, so future work on modelling wage and price setting behavior jointly may yield a more appealing explanation of this feature of the data.

Another important result concerns the nature of learning dynamics. Unlike earlier work, we find evidence against a model of least squares learning with small and constant gain. In particular, our results show that the gain parameter varies over time and is higher in periods of macroeconomic instability. Our attempt to model this time variation in the gain parameter is rather limited, of course, and is not intended as a structural alternative to CGLS. However, it is sufficient to provide reduced-form evidence against a constant gain specification of learning dynamics.

5 Conclusion

The objective of this paper was to study classical inference on the structural parameters of models with adaptive learning. We made three contributions. First, we showed that standard tests on the structural parameters are unreliable, and we provided explanations for this result. Specifically, we argued that this can be attributed to weak identification and persistent dynamics induced by learning. Weak identification and persistence in the data are two of the most important issues that have preoccupied econometricians over the past two decades.

Second, we provided a solution to the problem of inference. Our proposed method is based on the Anderson and Rubin (1949) principle of testing the identification restrictions of the model. The method is quite general and fully robust to the above problems, and its implementation is straightforward. Simulations conducted for the new Keynesian Phillips curve model with indexation showed that our method is reliable and has good power in finite samples. Our empirical application confirmed that in practice.

Third, we tested the hybrid NKPC model of inflation with indexation under adaptive learning, and found that standard versions of the model do not fit the observed dynamics in US inflation over the postwar period. We also found evidence against constant gain least squares specifications of learning dynamics. When we allow for time delays in price changes, an autoregressive cost-push shock and shifts in the gain parameter over time, the model can fit the data without indexation.

Our empirical results suggest that the gain parameter, which characterizes the ability of the learning algorithm to track structural changes, varies in a way that depends on macroeconomic conditions. Specifically, tracking seems to be more pronounced over a period of macroeconomic instability. Thus, we believe that it is appropriate to model learning using a gain sequence that is endogenous, time-varying and state-dependent, as for instance in Marcet and Nicolini (2003). Such alternatives can be studied easily using the econometric method that we propose in this paper.

A Appendix

Proof of proposition 2. Solving equation (9) in terms of $\{\eta_t\}_{t=1}^{T}$ and a_0, we obtain:

$$a_t - \alpha = (1-(1-\beta)\gamma)^t (a_0-\alpha) + \gamma \sum_{i=0}^{t-1} (1-(1-\beta)\gamma)^i \eta_{t-i}.$$

Substituting for β and γ using $\gamma = \psi/\sqrt{T}$ and $1-(1-\beta)\gamma = \exp(\phi/T)$, this can be written as:

$$a_t - \alpha = e^{\phi t/T}(a_0-\alpha) + \frac{\psi}{\sqrt{T}} \sum_{i=0}^{t-1} e^{\phi i/T} \eta_{t-i}.$$

Let the standard Brownian motion W such that, for $0 \leq r \leq 1$, $T^{-1/2}\sum_{t=1}^{[Tr]} \eta_t \Rightarrow \sigma_\eta W(r)$ as $T \to \infty$, then

$$\frac{\psi}{\sqrt{T}} \sum_{i=0}^{[Tr]-1} e^{\phi i/T} \eta_{[Tr]-i} \Rightarrow \psi \sigma_\eta J_\phi(r),$$

see Phillips (1987, Lemma 1), where J_ϕ is an Ornstein-Uhlenbeck diffusion with $J_\phi(0) = 0$ and parameter ϕ, driven by the Brownian motion $W(r)$. Moreover, since $e^{\phi r} - e^{\phi t/T} \to 0$ as $T \to \infty$ uniformly in $0 \leq r \leq 1$, equation (10) follows by Slutsky's formula for weak convergence.

Now, we turn to the OLS estimators:

$$\begin{bmatrix} \widehat{\beta}-\beta \\ \widehat{\delta}-\delta \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T} a_{t-1}^2 & \sum_{t=1}^{T} a_{t-1} \\ \sum_{t=1}^{T} a_{t-1} & T \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} a_{t-1}\eta_t \\ \sum_{t=1}^{T} \eta_t \end{bmatrix}, \quad \text{or}$$

$$\begin{bmatrix} \sqrt{T}\left(\widehat{\beta}-\beta\right) \\ \sqrt{T}\left(\widehat{\delta}-\delta\right) \end{bmatrix} = \left(\begin{bmatrix} T^{-1}\sum_{t=1}^{T} a_{t-1}^2 & T^{-1}\sum_{t=1}^{T} a_{t-1} \\ T^{-1}\sum_{t=1}^{T} a_{t-1} & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} a_{t-1}\eta_t \\ T^{-1/2}\sum_{t=1}^{T} \eta_t \end{bmatrix} \quad (20)$$

Since $K_{\psi,\phi}(r)$ is adapted to $W(r)$, it follows that $\sum_{t=1}^{T} a_{t-1} \frac{\eta_t}{\sqrt{T}} \Rightarrow \sigma_\eta \int_0^1 K_{\psi,\phi}(r)\, dW(r)$. Moreover, application of the continuous mapping theorem shows that $T^{-1} \sum_{t=1}^{T} a_{t-1} \Rightarrow \int_0^1 K_{\psi,\phi}(r)\, dr$ and $T^{-1} \sum_{t=1}^{T} a_{t-1}^2 \Rightarrow \int_0^1 K_{\psi,\phi}^2(r)\, dr$, and hence, the result (11) follows. ■

Computation of confidence sets The $\mathcal{AR}$ statistic for the final model in section 4 depends on the two pairs of parameters $\theta^1 = (\vartheta, \varrho)$ and $\theta^2 = (\gamma_1, \gamma_2)$. Projection-based confidence sets for each pair, θ^i, $i = 1, 2$, are obtained by comparing the statistic $\min_{\theta^j \in \Theta^j} \mathcal{AR}(\theta_0^i, \theta^j)$ to the appropriate critical value of the $\chi^2(k)$ distribution for each $\theta_0^i \in \Theta^i$. Since the $\mathcal{AR}$ criterion function is continuously differentiable with respect to (ϑ, ϱ), but the parameter space is compact, minimization with respect to (ϑ, ϱ) is performed using a sequential quadratic programming algorithm. The criterion function is not smooth in (γ_1, γ_2) due to the projection facility, so optimization with respect to (γ_1, γ_2) is performed by grid search or simulated annealing.

References

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics 20*, 46–63.

Andrews, D. W. and J. H. Stock (2005). Inference with weak instruments. NBER Technical Working Papers 0313, National Bureau of Economic Research, Inc.

Beyer, A. and R. E. A. Farmer (2007). Testing for indeterminacy: An application to U.S. monetary policy: Comment. *American Economic Review 97*(1), 524–529.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association 90*(430), 443–450.

Bray, M. M. and N. E. Savin (1986). Rational expectations equilibria, learning, and model specification. *Econometrica 54*(5), 1129–1160.

Bullard, J. B. and S. Eusepi (2005). Did the Great Inflation occur despite policymaker commitment to a Taylor rule? *Review of Economic Dynamics 8*, 3244–359.

Canova, F. and L. Sala (2009). Back to square one: identification issues in DSGE models. *Journal of Monetary Economics.* forthcoming.

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

Chan, N. H. and C. Z. Wei (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics 15*(3), 1050–63.

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

Clarida, R., J. Galí, and M. Gertler (1999). The science of monetary policy: A New Keynesian perspective. *Journal of Economic Literature 37*(4), 1661–1707.

Cochrane, J. H. (2007a). Identification with Taylor rules: A critical review. NBER Working Papers 13410, National Bureau of Economic Research, Inc.

Cochrane, J. H. (2007b). Inflation determination with Taylor rules: A critical review. NBER Working Papers 13409, National Bureau of Economic Research, Inc.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica 65*(6), 1365–1387.

Dufour, J.-M. (2003). Identification, weak instruments and statistical inference in econometrics. *Canadian Journal of Economics 36*(4), 767–808. Presidential Address to the Canadian Economics Association.

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

Evans, G. W. and S. Honkapohja (2008). Expectations, learning and monetary policy: An overview of recent rersearch. Discussion Paper 6640, CEPR.

Fourgeaud, C., C. Gourieroux, and J. Pradel (1986). Learning procedures and convergence to rationality. *Econometrica 54*(4), 845–68.

Fuhrer, J. C. and G. R. Moore (1995). Inflation persistence. *Quarterly Journal of Economics 110*, 127–159.

Galí, J. and M. Gertler (1999). Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics 44*, 195–222.

Gorodnichenko, Y. and S. Ng (2007). Estimation of DSGE models when the data are persistent. Technical report. Presented at NBER Summer Institute.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hodges, J. L. and E. L. Lehmann (1963). Estimates of location based on rank tests. *Annals of Mathematical Statistics 34*(2), 598–611.

Judge, G., R. Hill, W. Griffiths, H. Lutkepohl, and T.-C. Lee (1985). *The Theory and Practice of Econometrics*. New York, U.S.A.: Wiley.

Kleibergen, F. and S. Mavroeidis (2008). Weak instrument robust tests in GMM and the new Keynesian Phillips curve. Brown University.

Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *American Economic Review 63*(3), 326–334.

Lucas, R. E. (1976). Econometric policy evaluation: a critique. In K. Brunner and A. Meltzer (Eds.), *The Philips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy. Amsterdam: North-Holland.

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

Mavroeidis, S. (2005). Identification issues in forward-looking models estimated by GMM with an application to the Phillips Curve. *Journal of Money Credit and Banking 37*(3), 421–449.

Milani, F. (2005). Adaptive learning and inflation persistence. Working Papers 050607, University of California-Irvine, Department of Economics.

Milani, F. (2007). Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics 54*(7), 2065–2082.

Nicholls, D. F. and A. R. Pagan (1983). Heteroscedasticity in models with lagged dependent variables. *Econometrica 51*(4), 1233–42.

Orphanides, A. (2004). Monetary policy rules, macroeconomic stability, and inflation: A view from the trenches. *Journal of Money, Credit and Banking 36*(2), 151–75.

Orphanides, A. and J. C. Williams (2004). Imperfect knowledge, inflation expectations, and monetary policy. In B. Bernanke and M. Woodford (Eds.), *The Inflation Targeting Debate*. University of Chicago Press.

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

Orphanides, A. and J. C. Williams (2005b). Inflation scares and forecast-based monetary policy. *Review of Economic Dynamics 8*(2), 498–527.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika 74*(3), 535–547.

Primiceri, G. E. (2006). Why inflation rose and fell: Policymakers' beliefs and US postwar stabilization policy. *Quarterly Journal of Economics 121*(3), 867–901.

Rudd, J. and K. Whelan (2005). New tests of the New-Keynesian Phillips curve. *Journal of Monetary Economics 52*(6), 1167–1181.

Rudd, J. and K. Whelan (2006). Can rational expectations sticky-price models explain inflation dynamics? *American Economic Review 96*(1), 303–320.

Sargent, T. J. (1993). *Bounded Rationality in Macroeconomics*. Oxford: Clarendon Press.

Sbordone, A. M. (2002). Prices and unit labor costs: a new test of price stickiness. *Journal of Monetary Economics 49*, 265–292.

Schorfheide, F. (2005). Learning and monetary policy shifts. *Review of Economic Dynamics 8*, 392–419.

Sims, C. A. and T. Zha (2006). Were there regime switches in U.S. monetary policy? *American Economic Review 96*(1), 54–81.

Smets, F. and R. Wouters (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review 97*(3), 586–606.

Staiger, D. and J. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica 68*(5), 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo (2002). GMM, weak instruments, and weak identification. *Journal of Business and Economic Statistics 20*, 518–530.

Stock, J.H. and M. Yogo (2005). Testing for Weak Instruments in Linear IV Regression. In D.W.K. Andrews and J.H. Stock (Ed.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 80–108. Cambridge: Cambridge University Press. Prepared for Festschrift in honor of Thomas Rothenberg.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–38.

White, H. (1984). *Asymptotic Theory for Econometricians*. New York: Academic Press.

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

β \ γ	0.01	0.05	0.1
0.90	$20,000$	$3,000$	$1,000$
0.95	$40,000$	$6,000$	$4,000$
0.99	$100,000$	$40,000$	$10,000$

Table 1: Estimates of the minimum sample size T that is needed for a 5% nominal level t-test on β to be rejected no more than 10% of the time under the null hypothesis. The model is $y_t = \beta y_t^e + \delta + \eta_t$ with CGLS learning and gain parameter γ. η_t is Gaussian white noise with unit variance, $\delta = 0$ and learning is initialized using a pre-sample of 1000 observations. T is incremented by 10^n up to 10^{n+1}, then by 10^{n+1} up to 10^{n+2} and so on, starting with $n = 2$. The number of Monte Carlo replications is 2000.

T \ φ	Wald 75%	Wald 90%	Wald 95%	Wald 99%	$\mathcal{AR}$ 75%	$\mathcal{AR}$ 90%	$\mathcal{AR}$ 95%	$\mathcal{AR}$ 99%
100	48.7**	63.0**	70.4**	82.0**	73.1**	88.5**	94.0**	98.6**
200	56.0**	71.2**	78.7**	89.2**	74.1*	89.4	94.4*	98.9
400	59.6**	75.5**	82.6**	92.2**	74.5	89.7	94.9	98.9
600	60.2**	76.7**	84.3**	93.1**	75.0	89.9	94.9	99.0
800	60.8**	78.3**	85.4**	94.5**	75.2	89.6	94.5*	99.0
1000	61.5**	78.2**	85.7**	94.5**	75.0	90.0	94.9	99.0
10000	66.3**	83.4**	90.4**	97.1**	75.6	90.3	95.1	99.0

Table 2: Coverage probabilities of the Wald and the $\mathcal{AR}$-based confidence sets with confidence levels 75%, 90%, 95% and 99% for the null hypothesis that ϱ is equal to its true value. The parameter values in the DGP are $\beta = 0.99$, $\gamma = 0.01$, $\varrho = 0.65$, $\lambda = 0.15$, $\sigma_\varepsilon = 3$, $\sigma_{\varepsilon v} = 0.1$, $\rho_1 = 0.9$ and $\rho_2 = 0$. The number of Monte Carlo replications is $M = 10000$. One or two asterisks indicate significance of the coverage probability at the 5% and 1% level, respectively, as measured by its asymptotic χ^2 distribution.

model specification	min $\mathcal{AR}(\theta)$	p-value*
baseline	39.90	0.0001
π_t included in π_{t+1}^e	31.31	0.0018
PLM is VAR(2)	27.16	0.0073
PLM is VAR(3)	25.84	0.0113
PLM is VAR(4)	26.84	0.0082
π_{t+1}^e from Greenbook†	41.65	0.0000
π_{t+1}^e from Survey of Prof. Forecasters (SPF)‡	59.00	0.0000
PLM with SPF as additional predictor‡	38.24	0.0001
inflation measure based on CPI	42.47	0.0000

*p-value is based on $\chi^2(12)$ distribution.

†available sample: 1967:Q1-1995:Q4.

‡available sample: 1970:Q2-2007:Q3.

Table 3: Fit of the baseline NKPC with indexation.

q	$\min \mathcal{AR}(\theta)$	p-value*
1	33.69	0.0004
2	33.22	0.0003
3	33.18	0.0001
4	12.95	0.1135

shocks are $\eta_t \sim AR(q)$.

*p-value is based on $\chi^2(12-q)$.

Table 4: Fit of the NKPC with indexation and autocorrelated shocks

	$\eta_t \sim$ MDS		$\eta_t \sim AR(1)$	
d	$\min \mathcal{AR}(\theta)$	p-value*	$\min \mathcal{AR}(\theta)$	p-value†
1	34.60	0.0005	32.33	0.0007
2	43.45	0.0000	37.20	0.0001
3	41.71	0.0000	36.19	0.0002
4	33.29	0.0009	21.72	0.0266

*p-value is based on $\chi^2(12)$.

†p-value is based on $\chi^2(11)$.

Table 5: Fit of the NKPC with d-quarter delays in price changes

parameter	estimate	95% CI	90% CI
ϑ	0.62	$[0.36, 1.00]$	$[0.40, 1.00]$
ϱ	0.13	$[0.00, 0.46]$	$[0.00, 0.40]$
γ_1	0.01	$[0.01, 0.06]$	$[0.01, 0.04]$
γ_2	0.10	$[0.03, 0.10]$	$[0.04, 0.10]$
$\min \mathcal{AR}(\widehat{\theta})$: 12.77, p-value: 0.31, based on $\chi^2\,(11)$			

Table 6: The NKPC with time delays of 4 quarters in price changes, autocorrelated shocks and shifts in the gain parameter: γ_2 is the value of the gain parameter in the period 1974:Q1-1987:Q3, and γ_1 is the value before and after that period. The confidence intervals (CI) are derived by the projection method. Instruments include fours lags of the shocks, the labor share and the Fed funds rate, and the estimation sample is 1960:Q1-2007:Q3.

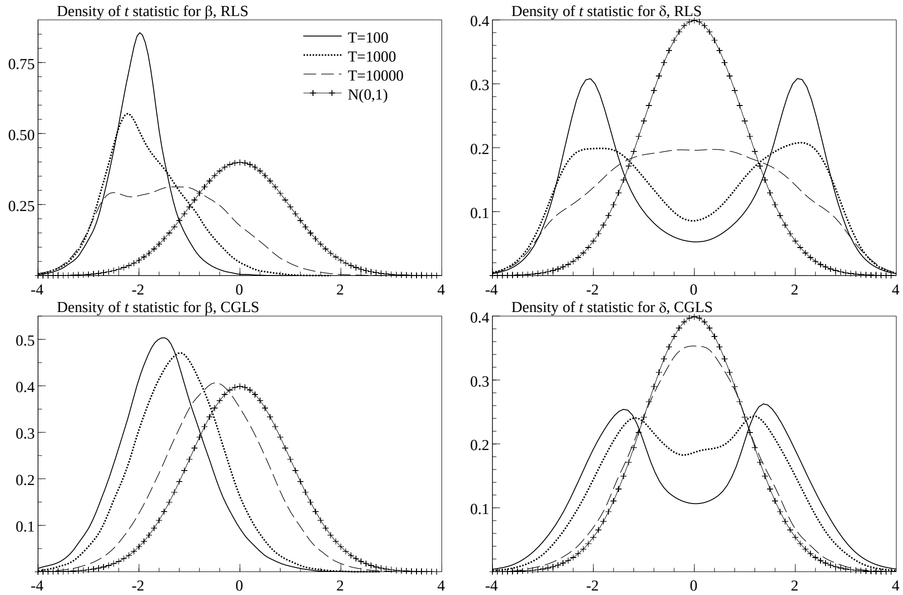

Figure 1: Densities of t statistics under the null hypothesis for the coefficients of model $y_t = \beta y_t^e + \delta + \eta_t$, $y_t^e = y_{t-1}^e + \gamma_t \left(y_{t-1} - y_{t-1}^e\right)$ for samples of size $T = 100, 1000, 10000$. η_t is Gaussian white noise with unit variance, $\beta = 0.9$ and $\delta = 0$. RLS corresponds to $\gamma_t = 1/t$, CGLS to $\gamma_t = 0.02$, and $y_0^e = 0$. The number of MC replications is 30000.

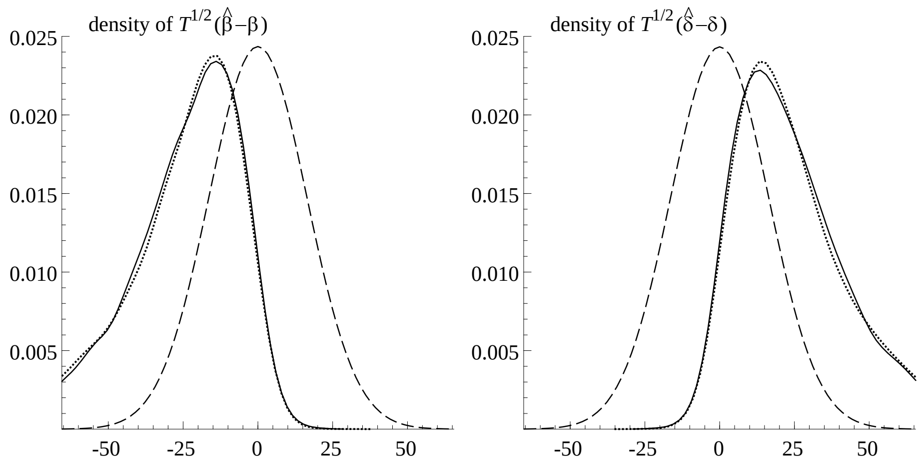

Figure 2: Densities of the OLS estimators for β and δ in a sample of size $T = 100$ (solid lines), local asymptotic approximations given by expression (11) (dotted lines) and normal asymptotic approximation (dashed lines). The model is $y_t = \beta y_t^e + \delta + \eta_t$ with CGLS with parameter $\gamma = 0.02$, and $\beta = 0.99$, $\delta = 0$, and learning is initialized at $a_0 = 1$. The number of MC replications is 10000.

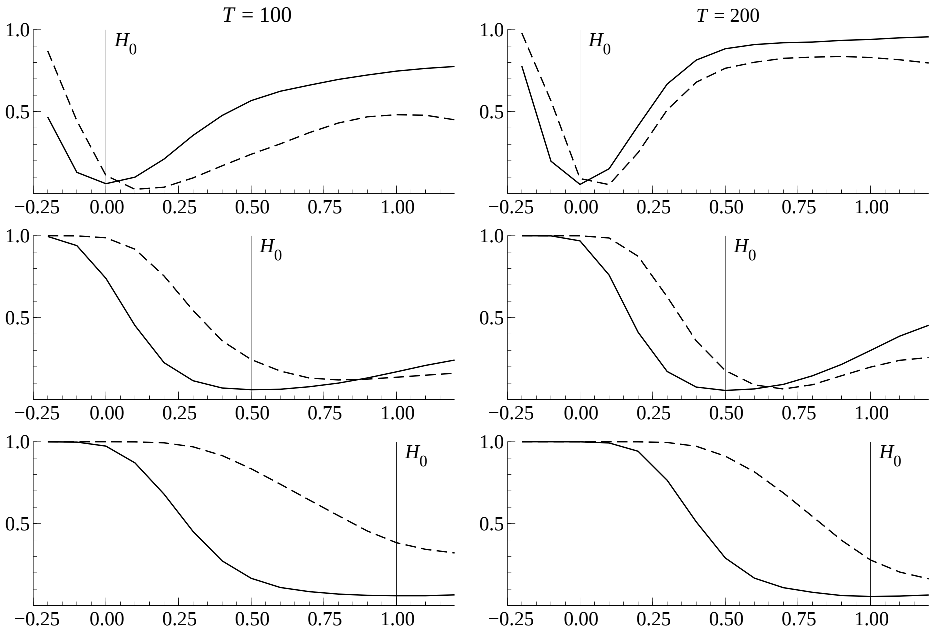

Figure 3: Power curves of the Wald (dotted line) and $\mathcal{AR}$ (solid line) tests for $T = 100$ (left column), and $T = 200$ (right column). The null hypothesis is $H_0 : \varrho = \varrho_0$, where $\varrho_0 = 0$ (top row), $\varrho_0 = 0.5$ (middle row), and $\varrho_0 = 1$ (bottom row), and it is superimposed by means of a vertical line. The value of ϱ under the alternative is shown by the abscissa. The other parameter values in the DGP are: $\beta = 0.99$, $\gamma = 0.01$, $\lambda = 0.15$, $\sigma_\varepsilon = 3$, $\sigma_{\varepsilon v} = 0.1$, $\rho_1 = 0.9$, $\rho_2 = 0$. The number of MC replications is $M = 10000$ and the nominal level of significance is 5%.

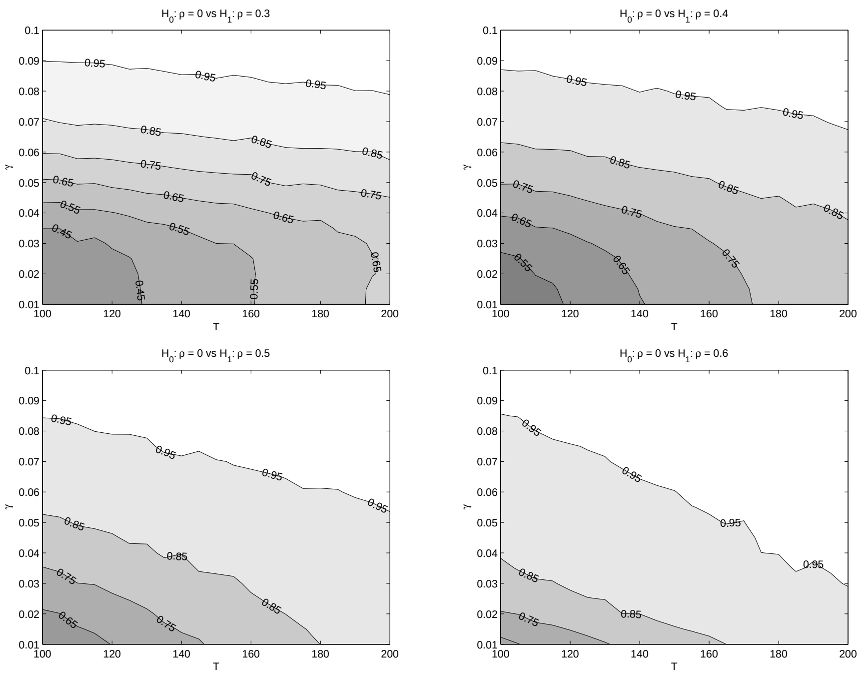

Figure 4: Power of the 5% level $\mathcal{AR}$ test of the null hypothesis $H_0 : \varrho = 0$ against the four alternatives: $\varrho = 0.3$, 0.4, 0.5 and 0.6. Each panel shows contours of the power function in terms of γ (the ordinate) and T (the abscissa). Power increases in the north-easterly direction in each of the four panels. The other parameter values are: $\beta = 0.99$, $\lambda = 0.15$, $\sigma_\varepsilon = 3$, $\sigma_{\varepsilon v} = 0.1$, $\rho_1 = 0.9$, $\rho_2 = 0$. The number of MC replications is $M = 10000$.

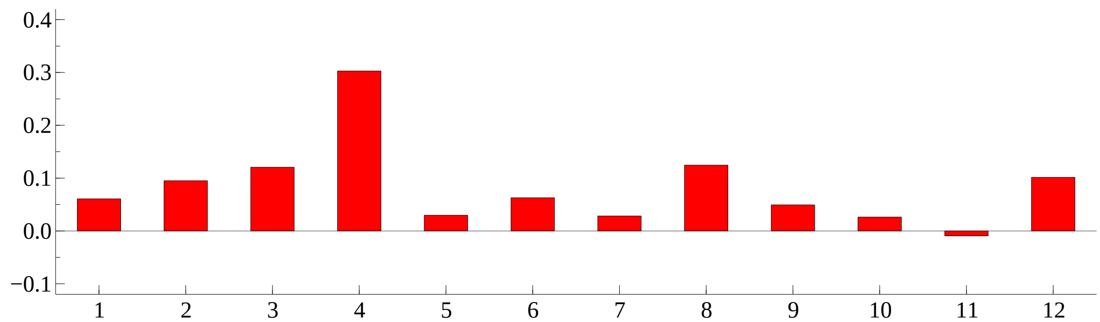

Figure 5: Correlogram of the residuals of the baseline NKPC model with indexation.

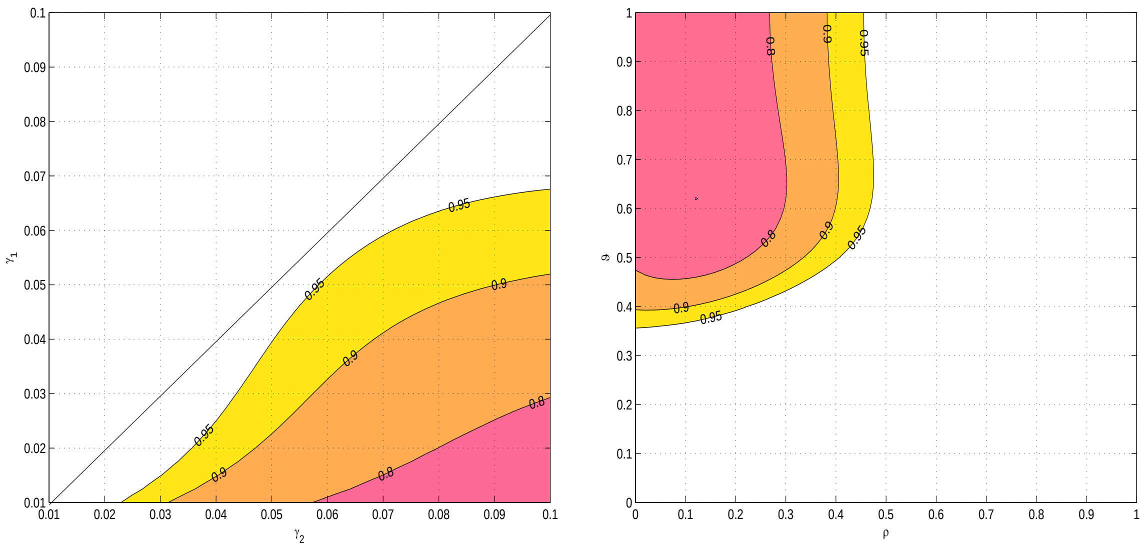

Figure 6: Two-dimensional confidence sets for (γ_1, γ_2), on the left, and (ϑ, ϱ), on the right, based on the Anderson-Rubin statistic. The model is the NKPC with indexation, four-quarter time delay in price changes and AR(1) shock. Adaptive learning is CGLS with gain parameter γ_1 over the periods 1960:Q1-1973:Q3 and 1987:Q4-2007:Q3, and γ_2 over the period 1973:Q4-1987:Q3. Instruments include four lags of the shocks, the labor share and the Fed Funds rate, and White's HC estimator is used.