Econometric Theory, **34**, 2018, 68–111.
doi:10.1017/S0266466616000529

ESTIMATING STRUCTURAL PARAMETERS IN REGRESSION MODELS WITH ADAPTIVE LEARNING

NORBERT CHRISTOPEIT
University of Bonn

MICHAEL MASSMANN
WHU – Otto Beisheim School of Management and Vrije Universiteit Amsterdam

This paper examines the ordinary least squares (OLS) estimator of the structural parameters in a simple macroeconomic model in which agents are boundedly rational and use an adaptive learning rule to form expectations of the endogenous variable. The popularity of learning models has recently increased amongst applied economists and policy makers who seek to estimate them empirically. Yet the econometrics of learning models is largely uncharted territory. We consider two prominent learning algorithms, namely constant gain and decreasing gain learning. For each of the two learning rules, our analysis proceeds in two stages. First, the paper derives the asymptotic properties of agents' expectations. At the second stage, the paper derives the asymptotics of OLS in the structural model, taking the first stage learning dynamics as given. In the case of constant gain learning, the structural model effectively amounts to a stationary, cointegrating, or co-explosiveness regression. With decreasing gain learning, the regressors are asymptotically collinear such that OLS does not satisfy, in general, the Grenander conditions for consistent estimability. Nevertheless, this paper shows that the OLS estimator remains consistent in all models considered. It also shows, however, that its asymptotic distribution, and hence any inference based upon it, may be nonstandard.

1. INTRODUCTION

The purpose of this paper is to investigate the properties of the ordinary least squares estimator of the structural parameters in a simple macroeconomic model with adaptive learning. In particular, we will consider a special case of the following stereotypical model that has received substantial attention in the economic

This paper replaces and extends Christopeit and Massmann (2010, 2013). The authors would like to thank seminar participants at the Universities of Cologne, Oxford, St. Andrews and Birmingham, at the Tinbergen Institutes in Amsterdam and Rotterdam, at ESEM 2013 and at CFE 2015. In particular, the discussions with and suggestions of Karim Abadir, Klaus Adam, Jörg Breitung, George Evans, Cars Hommes, Offer Lieberman, Tassos Magdalinos, Roderick McCrorie, Kaushik Mitra and Bent Nielsen are greatly appreciated. All remaining errors are of course ours.
Address correspondence to Michael Massmann, WHU – Otto Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany; e-mail: michael.massmann@whu.edu.

© Cambridge University Press 2017

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

theory literature:

$$y_t = \beta y^e_{t|t-1} + \delta x_t + \varepsilon_t, \quad t = 1, 2, \ldots \tag{1.1}$$

where $y^e_{t|t-1}$ denotes agents' expectations about y_t based on the information available at time $t-1$, the driving variable x_t is exogenous, and the error terms ε_t are independent and identically distributed (i.i.d.). Models of this type have a long tradition in economics. For instance, the classical cobweb model fits into this form, see e.g., Bray and Savin (1986) and Fourgeaud, Gourieroux, and Pradel (1986), as does the Lucas (1973) aggregate supply model and the New Keynesian Phillips curve (NKPC), cf. Roberts (1995) and Clarida, Galí, and Gertler (2000). In some settings, (1.1) is the solution of a system of behavioural equations, with the coefficients β and δ being functions of deep parameters. What is common to most specifications, however, is that the slope parameter δ generally turns out to be unrestricted while the autoregressive coefficient β either lies within the unit interval, as in the NKPC, or is negative, as in the cobweb model. For the purposes of this paper, we will refer to the coefficients in (1.1) as structural parameters as they can often be given an economic interpretation: For instance, in the NKPC, β is the discount factor and δ the slope of the Phillips curve.

With a view to analysing the econometrics of model (1.1) consider its economic aspects first: Of central interest in the economic literature is the way in which the expectations $y^e_{t|t-1}$ are modelled. The traditional approach is via rational expectations, cf. Muth (1961) or Sargent (2008), which assumes that agents, when forming expectations $y^e_{t|t-1}$, have complete knowledge of the model as well as of the past $\mathcal{F}_{t-1} = \sigma(y_s, s \leq t-1; x_s, s \leq t)$ and make best use of it, i.e., set $y^e_{t|t-1} = \mathbf{E}(y_t|\mathcal{F}_{t-1})$. Taking conditional expectations in (1.1) yields $\mathbf{E}(y_t|\mathcal{F}_{t-1}) = \alpha x_t$ with

$$\alpha = \frac{\delta}{1-\beta}. \tag{1.2}$$

The so-called rational expectations equilibrium (REE) model is thus

$$y_t = \alpha x_t + \varepsilon_t. \tag{1.3}$$

Obviously, under the assumption of rational expectations, only α is identified; not, however, δ and β separately.

More recently, economic agents are frequently assumed to be boundedly rational and to form their expectations via adaptive learning, see Sargent (1993, 1999), Evans and Honkapohja (2001), Hommes (2002), Preston (2005) or Gaspar, Smets, and Vestin (2010). The basic idea underlying all adaptive learning procedures is that agents employ an auxiliary model, or so-called perceived law of motion, to form their expectations $y^e_{t|t-1}$. One way to specify this auxiliary model is to assume that its functional form corresponds to that of the REE in (1.3). Agents are then presumed not to know the parameter α and replace it by some estimate a_{t-1}, based on information $\mathcal{F}_{t-1}$. Typically, they estimate the parameter α by a recursive procedure of the form:

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$a_t = a_{t-1} + \gamma_t \frac{x_t}{r_t}(y_t - a_{t-1}x_t) \tag{1.4a}$$

$$r_t = r_{t-1} + \gamma_t\left(x_t^2 - r_{t-1}\right), \tag{1.4b}$$

where γ_t is some weighting, or gain, sequence. This updating algorithm can be viewed as generalising the recursive least squares estimator of α, which has $\gamma_t = 1/t$ and whose r_t is the sample second moment of x_t. With the learning scheme in (1.4), agents' expectation will be given by

$$y^e_{t|t-1} = a_{t-1}x_t,$$

and the resulting so-called actual law of motion, or data generating process (DGP), is

$$y_t = \beta a_{t-1}x_t + \delta x_t + \varepsilon_t. \tag{1.5}$$

It is thus plain that, in models with adaptive learning, the expectational term $y^e_{t|t-1}$ creates a forecast feedback, resulting in a self-referential, and thus highly complex, DGP. Moreover, the stochastic behaviour of the DGP depends crucially on the specification of the gain sequence γ_t.

The analysis of the recursion in (1.4) is a problem of *stochastic approximation*. Since the seminal paper of Robbins and Monro (1951), stochastic approximation has become an important topic in the statistics literature and, from the late 1970s onwards, in stochastic system as well as control theory. Various approaches to analysing such algorithms have been developed, the most prominent of which are those based on the convergence of non-negative almost supermartingales or of stochastic Lyapunov functions (cf. Robbins and Siegmund (1971) and Lai (1989), respectively), the ODE approach of Ljung (1977), and the weak convergence method of Kushner (1984). Of interest in their own right are also the functional analytic procedure of Walk and Zsidó (1989) and the averaging analysis of Solo and Kong (1995). For recent surveys of stochastic approximation algorithms and their applications, see Lai (2003) and Kushner (2010).

Empirical models with learning have recently gained popularity amongst researchers and policy makers; see for instance the New Keynesian Phillips curve models estimated by Milani (2007) and Chevillon, Massmann, and Mavroeidis (2010), the European Central Bank's New Multi-Country Model by Dieppe, González Pandiella, Hall, and Willman (2011), the inflation model by Malmendier and Nagel (2016) and the model of stock market volatility by Adam, Marcet, and Nicolini (2016). Yet not much is known about the econometrics of adaptive learning models. In this paper, we contribute to filling this gap by investigating the asymptotic behaviour of the OLS estimator of the structural parameters β and δ in (1.5), henceforth denoted by $\widehat{\beta}$ and $\widehat{\delta}$, respectively. We will refer to this issue as the *external estimation problem* (*EEP*). In particular, the EEP concerns the question of the weak consistency of the OLS estimator and of its asymptotic distribution, the former in fact being a natural by-product of the existence of the

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

latter.[1] Note that the model in (1.5) is a linear regression model with predetermined stochastic regressors. There is a rich literature on the properties of OLS estimators in this model class, yet even the most powerful available results cannot be applied to particular models considered in this paper. This problem is particularly acute for the question of consistency and will be commented on further below. With the properties of $\widehat{\beta}$ and $\widehat{\delta}$ in (1.5) thus being generally unsettled, it is as yet unclear whether empirical implementations of it are built on sound econometric principles. Examining the OLS estimator of β and δ is the main purpose of this paper, and the aim is to provide a response to the contention of Marcet and Sargent (1988, p. 171):

> *"It is open and problematic whether* [*a learning system*] *can ever be expected to yield econometric models that can be applied. ...* [*Our results*] *probably imply that these parameters cannot be consistently estimated."*

With a view to examining the EEP we will first have to address the asymptotic properties of agents' forecast a_t of α, see (1.4a) above, an issue usually referred to as the *internal forecasting problem* (*IFP*). Two aspects of this issue are of interest: (*i*) Does a_t converge to the REE α? (*ii*) If so, at which rate does a_t converges? The first of these questions has been conclusively answered in the literature on stochastic approximation; a good account of seminal results can be found in Benveniste, Métivier, and Priouret (1990) and Kottmann (1990). In particular, the literature generally distinguishes between two basic approaches of specifying the gain sequence: constant gain learning, i.e., with $\gamma_t = \gamma$, and decreasing gain learning, for which $\gamma_t \to 0$. It can be shown that, in the former case, a_t does not in general converge to α. Agents are thus said to learn perpetually. As opposed to that, agents are fully rational asymptotically in the case of decreasing gain learning since the convergence $a_t \to \alpha$ does hold with probability one under suitable summability assumptions on γ_t, see below. This result presumes that $\beta < 1$ since it can be shown that, if $\beta \geq 1$, a_t diverges, see e.g., Christopeit and Massmann (2010) for details. Indeed, while most of the models in the literature presume that $\beta \in (0, 1)$, there are some that consider negative values; see, for example, Evans, Honkapohja, Sargent, and Williams (2013) and Brock and Hommes (1997) who analyse cobweb-type models with $-0.5 < \beta < 1$ and $\beta < -1$, respectively. With the question of convergence under decreasing gain settled, it is surprising that the second of the two questions above has to date only received scarce attention in the literature. Results on the convergence rate for a particular parameterisation have been derived by Marcet and Sargent (1995) and for agents employing a Bayesian estimation approach by Vives (1993). Yet the asymptotic behaviour of a_t turns out to be crucial for the properties of the OLS estimator in the EEP and so we provide a complete treatment of this issue in the present paper.

As to the EEP, the few existing results on the asymptotic behaviour of $\widehat{\beta}$ and $\widehat{\delta}$ pertain mainly to the case of constant gain learning, see e.g., Chevillon et al. (2010) and Adam et al. (2016). In this case, the regressor a_t is an autoregressive

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

process with constant coefficients and, depending on the value of β and γ, is either stationary ergodic, a random walk (possibly with a negative unit root) with drift, or explosive, given suitable input sequences x_t. We will appeal to results available in the literature to derive the asymptotic properties of $\widehat{\beta}$ and $\widehat{\delta}$; see Lai and Wei (1985) for the stationary ergodic case, Chan and Wei (1988) for the (negative) unit root case, and Phillips and Magdalinos (2008) as well as Wang and Yu (2015) for the explosive case. On the other hand, when learning is of decreasing gain type such that $a_t \to \alpha$ with probability one, then the regressors in (1.5) will be asymptotically collinear: $y_t \sim \beta\alpha x_t + \delta x_t + \varepsilon_t$, i.e., the asymptotic moment matrix

$$M = \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{pmatrix} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t^2 \tag{1.6}$$

will be singular. This violates one of the classical "Grenander conditions", i.e., the condition that the regressor sample second moment matrix, suitably scaled, converges to a positive definite limit; see Grenander and Rosenblatt (1957). In the econometrics literature the singularity of M is generally referred to as absence of strong asymptotic identification, see e.g., Davidson and MacKinnon (1993) or Newey and McFadden (1994). Given that the Grenander conditions are only sufficient but not necessary for consistent estimability, it is not clear a priori whether the OLS estimators $\widehat{\beta}$ and $\widehat{\delta}$ possess desirable asymptotic properties. One leading example in which the asymptotic singularity of M does not preclude the consistency and asymptotic normality of OLS is that of a linear model with slowly varying regressors, see Phillips (2007). The regressors in that paper, however, are deterministic while in our model a_t is stochastic. Indeed, the available results on the consistency of OLS in linear models with stochastic regressors, see, for instance, Christopeit and Helmes (1980), Lai and Wei (1982a, 1982b) and Christopeit (1986), are not met in some of our models. Nevertheless, in all of them, the OLS estimator turns out to have an asymptotic distribution and to be weakly consistent, as will be discussed below.

The present paper hence investigates two issues. First, as part of the IFP, the influence of the choice of gain sequence γ_t on the asymptotics of a_t is investigated. Secondly, the effect of γ_t on the properties of the OLS estimators $\widehat{\beta}$ and $\widehat{\delta}$ is explored in the context of the EEP. Specifically, we consider models such as (1.4)–(1.5) with either a constant gain

$$\gamma_t = \gamma, \tag{1.7}$$

or with a decreasing gain sequence

$$\gamma_t = \frac{\gamma}{t}, \tag{1.8}$$

where $\gamma > 0$. To see the motivation for this choice, note that, for a decreasing gain sequence, a set of summability conditions guaranteeing the convergence of a_t to α is $\sum_t \gamma_t = \infty$ but $\sum_t \gamma_t^2 \ln^2 t < \infty$, cf. Kottmann (1990). These conditions would

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

suggest considering sequences $\gamma_t = \gamma / t^{\eta}$, with $\eta \in \left(\frac{1}{2}, 1\right]$. Yet it is to be expected that the case $\eta = 1$ exhibits a behaviour that is furthest removed from that in the constant gain case, obtained by setting $\eta = 0$. Our choice of gain sequences can hence be seen as covering the two extremes on a behavioural continuum without introducing a further parameter, namely η, that would either have to be estimated or assumed known in the EEP. Moreover, recall also that $\gamma_t = 1/t$ yields recursive least squares learning while constant gain learning is known to be equivalent to exponential smoothing, both of which are, due to their intuitive appeal, of interest per se. Indeed, the economic literature often uses recursive least squares for modelling the learning of agents in stable regimes, as it implies that a forecast is a linear combination of past data, each of which receives the same weight, no matter how far in the past it was observed. Exponential smoothing, on the other hand, means that the weight given to past observations declines exponentially, making it predestined to be used by agents in unstable regimes. For instance, Marcet and Nicolini (2003) combine both types of learning in a switching model of hyperinflation, while Evans and Ramey (2006) discuss their relationship to the Lucas critique.

In order to delineate the influence of the choice of γ_t on agents' forecasts on the one hand and on the behaviour of the OLS estimators $\widehat{\beta}$ and $\widehat{\delta}$ on the other hand, without obscuring it by the influence of the regressors x_t, we consider in the following the very simplest scenario of (1.4)–(1.5) in which $x_t = x$ is a constant. This toy model will lay bare the difficulties arising from the feedback of the expectation formation process to the data generating process since even in this scenario the analysis will turn out to be nontrivial. Here, $r_t = x^2$ is the stationary solution of (1.4b) such that, for any starting value r_0, $\lim_{t\to\infty} r_t = x^2$. For constant gain learning, this is true whenever $\gamma \in (0, 1)$, while for decreasing gain it holds for all $\gamma > 0$. Without loss of generality, we may then assume that $x = 1$, any other value of x merely requiring that the variance of ε_t be changed from σ^2 to σ^2/x^2. Working with this stationary solution for r_t, the recursion in (1.4) reduces to

$$a_t = a_{t-1} + \gamma_t (y_t - a_{t-1}) \tag{1.9}$$

while the structural equation in (1.5) becomes

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t. \tag{1.10}$$

Note that by inserting (1.10) into (1.9) the dynamics of a_t may also be written as

$$a_t = (1 - c_t)\, a_{t-1} + \gamma_t (\delta + \varepsilon_t), \tag{1.11}$$

where we have defined

$$c_t = (1 - \beta)\, \gamma_t. \tag{1.12}$$

Under constant gain learning, c_t is equal to

$$c = (1 - \beta)\gamma, \tag{1.13}$$

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

cf. (1.7), while it is equal to $c_t = c/t$ under decreasing gain, see (1.8).

The model we consider in this paper is hence (1.10)–(1.12). Time-varying regressor sequences x_t are beyond the scope of the present paper. Similarly, replacing $y^e_{t|t-1}$ in the economic model (1.1) by forward-looking expectations $y^e_{t+1|t}$, i.e., expectations of time $t+1$ formed by agents at time t, introduces economic as well as econometric complications. The repercussions likely to arise from both extensions are discussed below.

The main contribution of the paper to the literature is that it derives the asymptotics of the OLS estimator $\widehat{\delta}$ and $\widehat{\beta}$ in the setting of model (1.10)–(1.12) under both constant and decreasing gain learning. This has not been attempted before. The paper may thus provide a theoretical justification for estimating this type of model empirically, which is becoming more and more popular in the profession. As a second major contribution, this paper presents a complete analysis of the asymptotics of agents' expectations a_t in the said model setting, and in particular of the rate at which a_t converges to the REE, if at all. Again, this is a novel undertaking, only comparable to the aforementioned results by Marcet and Sargent (1995) and Vives (1993), and will be of interest in the economic theory literature.

A first conclusion of our analysis concerns the interplay of IFP and EEP. It turns out that there is some sort of trade-off between the asymptotic behaviour of a_t on the one hand and that of the OLS estimators on the other, in the sense that convergence of agents' expectations to the REE is likely to have detrimental effects on the convergence of the OLS estimator, and vice versa. Our second conclusion is that the OLS estimator of the structural parameters β and δ is consistent albeit with highly nonstandard asymptotic distributions in all but one scenario. This result hence provides an answer to Marcet and Sargent's speculation quoted above, in that the learning systems considered in this paper do yield useful, albeit complex, econometric models.

The paper is structured as follows: Our results pertaining to constant gain learning are presented in Section 2 whilst those relating to decreasing gain learning are described in Section 3. Subsidiary issues and forward-looking expectations are discussed in Section 4 before Section 5 concludes. Proofs of the constant gain results are relegated to Appendix 1 and those of the decreasing gain results to Appendix A.3.3. Some auxiliary results as well as a discussion of time-varying regressors and of consistency are provided in an online supplement associated with this article, available at Cambridge Journals Online (journals.cambridge.org/ect).

2. CONSTANT GAIN LEARNING

In this section, we consider the model in (1.10)–(1.12) with a constant gain sequence $\gamma_t = \gamma > 0$. In particular, the model of interest comprises the structural equation

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t, \tag{2.1}$$

while the recursion in (1.11) becomes

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$a_t = (1-c)\,a_{t-1} + \gamma\,(\delta + \varepsilon_t), \tag{2.2}$$

with c as in (1.13). Note that, for $\beta \neq 1$, $\gamma\,\delta = c\alpha$. Although it is common in the economics literature to assume that $\gamma \in (0,1)$ and that $\beta < 1$, from a mathematical point of view, any value of β and any positive value of γ can and will in the following be admitted.

We will make the following maintained assumptions on the error term ε_t and on the initial value a_0 throughout the entire paper.

Assumption MA1. The ε_t are i.i.d. with mean 0 and variance σ^2.

Assumption MA2. The initial value a_0 may be deterministic or stochastic. In the latter case, it is in L^2 and independent of ε_t, $t \geq 1$.

It will turn out that the parameter c is decisive for the behaviour of a_t in Section 2.1 as well as for that of the OLS estimators $\widehat{\beta}$ and $\widehat{\delta}$ in Section 2.2.

2.1. IFP

Let us first consider the internal forecasting problem (IFP), i.e., the asymptotic behaviour of agents' forecast a_t in (2.2). The results will be used in the analysis of the external estimation problem (EEP) in Section 2.2 below, yet they are also of interest per se. Depending on the value of c, three different types of autoregressive processes arise:

(*i*) If $0 < c < 2$, the autoregressive coefficient $\rho = 1 - c$ in (2.2) satisfies $|\rho| < 1$. Therefore we are in the classical scenario of stable autoregressive processes.

(*ii*) If $c = 0$ then $\rho = 1$ and (2.2) is a random walk with drift. As opposed to that, if $c = 2$ then $\rho = -1$ and (2.2) is an alternating random walk with drift. Both are hence unit root processes with drift. The latter scenario of a negative unit root has not received much attention in the literature, the main reference being Chan and Wei (1988). Indeed, the behaviour of a_t under a negative unit root turns out below to be very different from the case with a positive unit root.

(*iii*) When $c < 0$ or $c > 2$, we have $|\rho| > 1$, and (2.2) is an explosive autoregressive process.

The asymptotic behaviour of a_t for the various cases is summarised in the following theorem. The proof is given in Appendix A.2.

THEOREM 1.

(*i*) *If* $0 < c < 2$ *then* a_t *converges in distribution to the law* $\mathcal{L}$ *of the stationary solution, i.e., to the invariant distribution. This is nondegenerate with mean* α *and positive variance.*

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

(*ii*) *If* $c = 0$ *then* a_t *is a random walk with drift* $\delta\gamma$*:* $a_t = \gamma\,\delta t + o(t)$ *a.s. If, instead,* $c = 2$ *then* a_t *is an alternating random walk with drift* 2α *and*

$$\frac{1}{\sigma\,\gamma\,\sqrt{t}} a_t \xrightarrow{d} \mathcal{N}(0,1). \tag{2.3}$$

(*iii*) *If* $c < 0$ *or* $c > 2$ *then* $(1-c)^{-t}\,a_t$ *converges with probability one and in* L^2 *to a limit with continuous distribution function and mean* $\mathbf{E}a_0 - \alpha$.

It is hence plain that, in none of the three cases, a_t converges to the REE α in any probabilistic sense. Agents will thus not be rational in the limit but learn ad infinitum.

As mentioned above, the behaviour of a_t depends on the value of $c = (1-\beta)\gamma$. While β is estimated as part of the EEP the value of γ is assumed known, the leading scenario being $\gamma = 1$ in the case of recursive least squares learning. A known γ is of course synonymous with the investigator specifying how quickly agents discount past observations in constructing their forecasts of α. Issues resulting from assuming that γ is unknown are discussed in Section 4.

2.2. EEP

We now turn to the external estimation problem (EEP) and to the problem of estimating the unknown parameter vector $\theta = (\delta, \beta)'$ in the structural equation (2.1). It turns out that prominent sufficient conditions for the consistency of OLS available in the literature do not generally apply, as will be discussed in Section D of the Online Appendix. Yet the following observation on the relationship between the EEP on the one side and OLS estimation in a first-order autoregressive model on the other side will allow us to derive the limiting distribution of the OLS estimator by making use of the asymptotics of a_t, as summarised in Theorem 1, and of results available in the time series literature. In particular, note that the OLS estimator $\widehat{\theta}_T = (\widehat{\delta}_T, \widehat{\beta}_T)'$ of θ based on T observations is traditionally written in the form

$$\widehat{\theta}_T - \theta = M_T^{-1} u_T, \tag{2.4}$$

where the second moment matrix M_T is given by

$$M_T = \begin{pmatrix} T & \sum_{t=1}^T a_{t-1} \\ \sum_{t=1}^T a_{t-1} & \sum_{t=1}^T a_{t-1}^2 \end{pmatrix} \tag{2.5}$$

and

$$u_T = \begin{pmatrix} \sum_{t=1}^T \varepsilon_t \\ \sum_{t=1}^T a_{t-1}\varepsilon_t \end{pmatrix}, \tag{2.6}$$

with a_t generated by (2.2). Yet an alternative way of analysing $\widehat{\theta}_T - \theta$ is to examine in lieu of equation (2.1) the following AR(1)-model:

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$a_t^* = \delta^* + \beta^* a_{t-1}^* + \gamma\, \varepsilon_t, \tag{2.7}$$

where, generally, the sequence a_t^* as well as the parameter vector $\theta^* = (\delta^*, \beta^*)'$ differ from, respectively, a_t and θ above. The OLS estimator of θ^* is given by

$$\widehat{\theta}_T^* - \theta^* = M_T^{*-1} u_T^*, \tag{2.8}$$

where M_T^* and u_t^* are defined analogously to (2.5) and (2.6), with a_t replaced by a_t^*.

If $\delta^* = \gamma\,\delta$ and $\beta^* = 1 - c$ then, if started with the same initial value $a_0 = a_0^*$, it is plain that $a_t = a_t^*$ for all t, that $M_T = M_T^*$ and that $u_T = \gamma^{-1} u_T^*$. Hence it follows from (2.4) and (2.8) that

$$\widehat{\theta}_T - \theta = \gamma^{-1}\left(\widehat{\theta}_T^* - \theta^*\right). \tag{2.9}$$

In summary, the asymptotic behaviour of the OLS estimator in (2.1) is, up to a factor of proportionality, equal to that of the OLS estimator in the AR(1)-model (2.7). This equivalence will be exploited in the proof of Theorem 2, see Appendix A.3.

THEOREM 2. *The OLS estimator $\widehat{\theta}_T = (\widehat{\delta}_T, \widehat{\beta}_T)'$ for the parameters $\theta = (\delta, \beta)'$ in model (2.1) has the following asymptotic behaviour.*

(*i*) *If $0 < c < 2$ then $\widehat{\theta}_T - \theta$ is asymptotically normal at rate $\sqrt{T}$:*

$$\sqrt{T}(\widehat{\theta}_T - \theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \overline{M}^{-1}\right), \tag{2.10}$$

where

$$\overline{M} = \begin{pmatrix} 1 & \alpha \\ \alpha & \sigma_a^2 + \alpha^2 \end{pmatrix},$$

and $\sigma_a^2 = \sigma^2 \gamma^2 / \left(1 - (1-c)^2\right)$ is the variance of the invariant distribution of a_t.

(*ii*) *If $c = 0$ and $\delta \neq 0$ then*

$$\begin{pmatrix} \sqrt{T}(\widehat{\delta}_T - \delta) \\ T^{3/2}(\widehat{\beta}_T - 1) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \Omega\right),$$

where

$$\Omega = \begin{pmatrix} 4 & -\frac{6}{\delta\gamma} \\ -\frac{6}{\delta\gamma} & \frac{12}{\delta^2\gamma^2} \end{pmatrix}. \tag{2.11}$$

If, instead, $c = 0$ and $\delta = 0$ then

$$\begin{pmatrix} \sqrt{T}\widehat{\delta}_T \\ T(\widehat{\beta}_T - 1) \end{pmatrix} \xrightarrow{d} Z,$$

where

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$Z = v^{-1}\begin{pmatrix} \sigma B(1)\int_0^1 B(s)^2 ds - \frac{\sigma}{2}\left(B(1)^2-1\right)\int_0^1 B(s)ds \\ \frac{1}{2\gamma}\left(B(1)^2-1\right) - \frac{1}{\gamma}B(1)\int_0^1 B(s)ds \end{pmatrix},$$

$$v = \int_0^1 B(s)^2 ds - \left(\int_0^1 B(s)ds\right)^2.$$

If, however, $c = 2$ *then*

$$\begin{pmatrix} \sqrt{T}(\widehat{\delta}_T - \delta) \\ T(\widehat{\beta}_T - \beta) \end{pmatrix} \xrightarrow{d} \frac{1}{\int_0^1 \widetilde{B}(s)^2 ds}\begin{pmatrix} \sigma B(1)\int_0^1 \widetilde{B}(s)^2 ds \\ -\frac{1}{2\gamma}\left[\widetilde{B}^2(1)-1\right] \end{pmatrix},$$

where B *and* $\widetilde{B}$ *are two independent standard Brownian motions.*

(*iii*) *If* $c < 0$ *or* $c > 2$ *then*

$$\sqrt{T}(\widehat{\delta}_T - \delta) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

$$(1-c)^T(\widehat{\beta}_T - \beta) \xrightarrow{d} \frac{1}{\gamma}\frac{\left[(1-c)^2-1\right]u}{v + (1-c)c\alpha/\left[(1-c)-1\right]}, \tag{2.12}$$

where u *and* v *are independent* L^2*-variables, and the two estimators are asymptotically independent. Moreover,* v *has a continuous distribution function so that the limit in* (2.12) *is well-defined.*

Some of the theorem's statements make use, of course, of standard textbook results, see e.g., the stable case (*i*) and the unit root case (*ii*) with $c = 0$ and $\delta = 0$. Also, the asymptotic normality of $\widehat{\beta}_T$ at rate $T^{3/2}$ in the unit root case (*ii*) with $c = 0$ and $\delta \neq 0$ corresponds to the results derived by West (1988). The explosive case (*iii*) is based on a recent paper by Wang and Yu (2015), see also Remark B.9 in Appendix B.4. For the alternating unit root case (*ii*) with $c = 2$, however, we obtain a nonstandard limit distribution of the Dickey–Fuller type that, to the best of our knowledge, has not yet been documented in the literature. Note that it concerns the joint limiting distribution of the OLS estimator of the intercept and slope parameters in (2.2) and its proof builds on the results derived by Chan and Wei (1988). A negative unit root has also been considered by, for instance, Abadir (1993), albeit in the context of the *uncentred* normalised autocorrelation coefficient.

The important aspect of Theorem 2 is that the behaviour of the OLS estimator depends crucially on the value of $c = (1-\beta)\gamma$, i.e., on the value of the autoregressive coefficient β and of the gain parameter γ in the structural model. The juxtaposition of the results for different c is meant to reinforce that conclusion. Importantly, the OLS estimators of δ and β are consistent or even superconsistent. Moreover, inference on both parameters is feasible in all scenarios, provided one takes account of the nonstandard distributions resulting in cases (*ii*) and (*iii*). Finally, the usual OLS estimator for σ^2 is also consistent, as will be discussed in Section 4.

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

3. DECREASING GAIN LEARNING

In this section, we consider the model in (1.10)–(1.12) with the gain sequence specified as $\gamma_t = \gamma / t$ for some constant $\gamma > 0$. In view of the discussion in the introduction, we consider the case of $\beta < 1$ so as to ensure that $a_t \to \alpha$. The model is hence given by:

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t, \tag{3.1}$$

$$a_t = \left(1 - \frac{c}{t}\right) a_{t-1} + \frac{\gamma}{t}(\delta + \varepsilon_t), \tag{3.2}$$

where c is, again, as in (1.13). Note that the condition $\beta < 1$ corresponds to $c > 0$. The value c will again turn out to be crucial for the behaviour of a_t as well as of $\widehat{\beta}$ and $\widehat{\delta}$. Recall also the maintained Assumptions MA1 and MA2 mentioned in Section 2.

We will now first discuss the internal forecasting problem (IFP) in the context of model (3.1)–(3.2) before turning to the external estimation problem (EEP) in Section 3.2.

3.1. IFP

As mentioned in the introduction, the mere convergence of a_t to α follows easily from well-known results on recursive algorithms, see for instance Christopeit and Massmann (2010, Section 3.1). In particular, Kottmann's (1990) sufficient conditions are met since $\beta < 1$ by assumption and our choice of gain sequence $\gamma_t = \gamma / t$ satisfies the summability conditions $\sum_t \gamma_t = \infty$ and $\sum_t \gamma_t^2 \ln^2 t < \infty$. However, for our analysis of the asymptotic distribution of the OLS estimator in Section 3.2, we will need the exact rates of convergence of a_t.

The IFP in the context of the model in (3.1)–(3.2) is strikingly nonstandard. This can be gleaned from the following characteristics of the process a_t: First, a_t is autoregressive of first order with a time-varying coefficient which is intrinsically local-to-unity. The behaviour of models of this kind has been analysed by Phillips (1987), with recent extensions to autoregressive coefficients of the type $1 - c/\tau_t$ with $\tau_t = o(t)$ by Phillips and Magdalinos (2007). Next, the impact of the intercept δ and of the disturbance ε_t on a_t tends to zero for large t. In the limit, a_t is thus constant. As a result, the process a_t is highly nonstationary, as reflected in, for instance, the fact that its variance decreases to zero at a rate which increases with c. Finally, a_t is generated by what Solo and Kong (1995) call a long memory algorithm.

The following theorem describes the asymptotic distribution of a_t in general and its convergence rates in particular. Its proof will be given in Appendix B.2

THEOREM 3.

(*i*) *If* $c > 1/2$ *then* $a_t - \alpha$ *is asymptotically normal at rate* $\sqrt{t}$:

$$\frac{\sqrt{2c-1}}{\sigma\gamma}\sqrt{t}\,(a_t - \alpha) \xrightarrow{d} \mathcal{N}(0,1).$$

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

(*ii*) *If* $c = 1/2$ *then* $a_t - \alpha$ *is asymptotically normal at rate* $\sqrt{t/\ln t}$:

$$\frac{1}{\sigma\gamma}\sqrt{\frac{t}{\ln t}}\,(a_t - \alpha) \xrightarrow{d} \mathcal{N}(0,1).$$

(*iii*) *If* $c < 1/2$ *then*

$$\lim_{t\to\infty} t^c\,(a_t - \alpha) = u$$

a.s. and in L^2*, where* u *has a continuous distribution function.*

It is plain that, as c decreases, the convergence of a_t to α gets progressively slower. The value $c = 1/2$ can be interpreted as a boundary separating 'good' asymptotic behaviour of a_t from 'poor' behaviour, in the sense of speed of convergence. To gain an intuition[2] for this threshold, consider the simple case of $\gamma = 1$, as in recursive least squares learning, such that $\beta = 1 - c$. Abstracting from the error term ε_t, the structural equation in (3.1) is given by $y_t = \delta + \beta y_t^e$, since $y_t^e = a_{t-1}$. Denoting the REE by $\bar{y} = \alpha = \delta/(1-\beta)$, the structural equation can be written as

$$y_t = \bar{y} + \beta(y_t^e - \bar{y}).$$

It thus follows that, at the one extreme, i.e., when $\beta = 1$, $y_t = y_t^e$ and the value of the process is always equal to what agents believe it to be. At the other extreme, i.e., when $\beta = 0$, $y_t = \bar{y}$ and the process stays put at the REE. If y_t takes an intermediate value but is closer to y_t^e than to $\bar{y}$, such that $\beta > 1/2 \Leftrightarrow c < 1/2$, then the speed at which it converges towards $\bar{y}$ is fast. Conversely, when $\beta < 1/2 \Leftrightarrow c > 1/2$, then y_t is already close to $\bar{y}$ and hence convergence towards it is slow. Indeed, the threshold of $1/2$ is reminiscent of a similar boundary discussed in Evans, Honkapohja, Sargent, and Williams (2013). In view of the aforementioned trade-off between the behaviour of a_t and that of $\widehat{\beta}_T$, one should expect the converse for the performance of $\widehat{\beta}_T$. This will indeed be seen in Theorem 4.

The result of $\sqrt{t}$-convergence in part (*i*) of Theorem 3 also follows from Benveniste et al. (1990), see their Theorem 3 on p. 110 and their Theorem 13 on p. 332, and has been used by Marcet and Sargent (1995) and Evans and Honkapohja (2001, Theorem 7.10), respectively. Indeed, Benveniste et al. (1990, Theorem 3, p. 110) explicitly state that a particular eigenvalue must be smaller than $-1/2$, a condition that translates into $c > 1/2$ in our context. Parts (*ii*) and (*iii*) of our Theorem 3 are, to the best of our knowledge, new.

For $c < 1/2$, the asymptotic distribution of a_t is, in general, not normal. Yet, making the explicit assumption that the error terms are Gaussian, we have the following corollary.

COROLLARY 1. *Consider part* (*iii*) *of Theorem 3. If* ε_t *is normally distributed then* $a_t - \alpha$ *is asymptotically normal at rate* t^c.

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

3.2. EEP

We now turn to the estimation of the structural parameters in (3.1). Our focus will be on the slope coefficient β from which the properties of the estimator of δ follow immediately, see the discussion at the end of this section. As mentioned earlier, the structural equation is a simple linear regression model with predetermined stochastic regressors, with error terms that are i.i.d. The OLS estimator of β can be written as

$$\widehat{\beta}_T - \beta = \frac{u_T}{A_T}, \tag{3.3}$$

with

$$u_T = \sum_{t=1}^{T} \left(a_{t-1} - \bar{a}_T^-\right) \varepsilon_t \tag{3.4}$$

$$A_T = \sum_{t=1}^{T} \left(a_{t-1} - \bar{a}_T^-\right)^2. \tag{3.5}$$

and $\bar{a}_T^- = \frac{1}{T}\sum_{t=1}^{T} a_{t-1}$.

Although OLS estimation in this model class has a very long tradition in statistics and econometrics it is not clear a priori whether or not $\widehat{\beta}$ is consistent in the present setting. As already discussed in the Section 1, one of the classical Grenander conditions for (3.1)–(3.2) is violated. Although this does not necessarily preclude consistent OLS estimation, even the currently best available and, in a sense, minimal sufficient conditions for consistency are not met for certain parameter constellations. This topic is discussed further in Section D of the Online Supplement.

It is plain that in examining the asymptotic properties of the OLS estimator in (3.3) no recourse can be taken to existing results. We therefore resort to analysing our model from first principles, making use of the behaviour of the predetermined regressor a_t as derived in Theorem 3 above. In particular, Theorem 4 below deals with the cases $c > 1/2$ and $c < 1/2$ and will be proved in Appendix B.3. For the boundary case $c = 1/2$ there is strong evidence that the OLS estimator does not converge to a standard limit distribution, see the discussion in Appendix B.3.3.

THEOREM 4.

(*i*) *If* $c > 1/2$ *then*

$$\sqrt{A_T}\left(\widehat{\beta}_T - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right).$$

The divergence rate of A_T *is given by*

$$\operatorname*{plim}_{T\to\infty} \frac{A_T}{\ln T} = \frac{\sigma^2\gamma^2}{2c-1}.$$

(*ii*) *If* $c < 1/2$ *then*

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$A_T\left(\widehat{\beta}_T-\beta\right)^2 \xrightarrow{d} \sigma^2\chi_1^2.$$

The divergence rate of A_T is given by

$$\operatorname*{plim}_{T\to\infty}\frac{A_T}{T^{1-2c}}=\nu^2u^2,$$

with u as in Theorem 3(iii) and $\nu^2=c^2/(1-2c)(1-c)^2$.

It is evident that Theorem 4 implies that $\widehat{\beta}_T$ is weakly consistent for $c<1/2$ and $c>1/2$. A more refined analysis of the paths of a_t, based upon a law of the iterated logarithm for martingales by Chow and Teicher (1973), shows that the OLS estimator is even strongly consistent. In fact, the result of strong, and thus weak, consistency carries over to the case $c=1/2$, see Christopeit and Massmann (2016).

Clearly, the divergence of A_T, and hence the convergence of $\widehat{\beta}_T$, increases as c decreases, as was expected given that the convergence of a_t to α gets slower as c becomes smaller, see Theorem 3. This is the aforementioned trade-off between the asymptotic behaviour of a_t and that of $\widehat{\beta}_T$.

It is instructive to note that, modulo a consistent estimator of σ^2, Theorem 4 effectively provides the asymptotic behaviour of the t- and Wald statistics.

As will be argued in the proof of Theorem 4, the asymptotic distribution of $\widehat{\beta}_T$ is in general not Gaussian when $c<1/2$. The reason for this is that the limits of both u_T and A_T depend on a random nuisance parameter. This is similar to the random nuisance parameter appearing in the asymptotic distribution theory of explosive AR(1)-models,

$$z_t=\phi z_{t-1}+\varepsilon_t, \qquad \text{with } |\phi|>1, \tag{3.6}$$

where both the numerator $\sum_{t=1}^T z_{t-1}\varepsilon_t$ and the denominator $\sum_{t=1}^T z_{t-1}^2$ of the OLS estimator tend geometrically fast to some nondegenerate random limit, see e.g., Wang and Yu (2015, Theorem 2.2). With the additional assumption of normally distributed ε_t, however, $\widehat{\beta}_T$ can be shown to have an asymptotic Cauchy distribution, as is proved in Appendix B.3.

COROLLARY 2. *Consider part (ii) of Theorem 4. If the ε_t are normally distributed,*

$$\kappa\nu\sqrt{T^{1-2c}}\left(\widehat{\beta}_T-\beta\right)\xrightarrow{d}\mathcal{C}(\mu),$$

where $\mathcal{C}(\mu)$ is a noncentral Cauchy distribution with noncentrality parameter $\mu=a_0B_0$, with B_0 a constant that depends only on c, and $\kappa=\gamma^2\mathbf{Var}(v)$. For $a_0=\alpha$, $\mu=0$ and $\mathcal{C}(0)$ is the standard Cauchy distribution.

Indeed, this limit distribution is also obtained for the OLS estimator of ϕ in (3.6) when the normalisation is with the deterministic sequence $|\phi|^{-2T}$, cf. White (1958), and Shiryaev and Spokoiny (1997).

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

For the asymptotic properties of the OLS estimators of the intercept δ write $\widehat{\delta}_T$ as $\widehat{\delta}_T - \delta = (\beta - \widehat{\beta}_T)\bar{a}^- + \bar{\varepsilon}$ and multiply it by $\sqrt{A_T}$ to obtain

$$\sqrt{A_T}(\widehat{\delta}_T - \delta) = \sqrt{A_T}(\beta - \widehat{\beta}_T)\bar{a}^- + \sqrt{A_T}\bar{\varepsilon}. \tag{3.7}$$

Since $\operatorname{plim}_{T\to\infty}\sqrt{A_T}\bar{\varepsilon} = 0$ and $\operatorname{plim}_{T\to\infty}\bar{a}^- = \alpha$, $\sqrt{A_T}(\widehat{\delta}_T - \delta)$ and $\sqrt{A_T}(\widehat{\beta}_T - \beta)$, as well as their respective squares, are asymptotically collinear.

4. DISCUSSION

4.1. The Error Variance and the Gain Parameter

Given the consistency of the OLS estimators $\widehat{\delta}_T$ and $\widehat{\beta}_T$ under both constant and decreasing gain learning, it follows immediately that the usual residual-based estimator of σ^2 is also consistent. The proof of the following corollary is provided in Section F of the Online Supplement.

COROLLARY 3. *Define the OLS residual* $\widehat{\varepsilon}_t = y_t - \widehat{\delta}_T - \widehat{\beta}_T a_{t-1}$. *Then* $\widehat{\sigma}^2_T = \frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2 \xrightarrow{p} \sigma^2$.

Regarding the gain parameter γ under constant or decreasing gain learning, see (2.2) and (3.2) respectively, the treatment so far has assumed that it is known. A case in point is to regard the decreasing gain updating equation for a_t as an instance of recursive least squares, since then $\gamma = 1$. Given γ and a starting value a_0, agents' forecast a_t can be endogenously generated by either

$$a_t = a_{t-1} + \gamma\,(y_t - a_{t-1}) \tag{4.1}$$

or

$$a_t = a_{t-1} + \frac{\gamma}{t}\,(y_t - a_{t-1})\,, \tag{4.2}$$

depending on which type of learning is stipulated. As a consequence, the OLS estimates $\widehat{\beta}_T$, $\widehat{\delta}_T$, and $\widehat{\sigma}^2_T$, all being functions of (y_t, a_{t-1}), can be computed. With γ known, c can also be estimated consistently by $\widehat{c} = (1 - \widehat{\beta})\gamma$ and hence the results in Theorems 2 and 4 are feasible. Note, however, that given an estimator $\widehat{c}$ the classifications in Theorems 1–4 are subject to estimation uncertainty.

It is instructive to note that the results stated in Theorems 2 and 4 would also go through if, conversely, γ were unknown yet agents' forecasts a_t were observed. For then γ may be computed from its DGP in (4.1) or (4.2). This type of situation might arise when survey data is used as measurement of agents' forecasts. Indeed, in such scenarios, an extra error term would be appended to (4.1) or (4.2) since it is unlikely to be an exact description of empirical data, see for instance Branch and Evans (2006, 2010) or Markiewicz and Pick (2014).

Finally, it is conceivable that both γ is unknown and a_t is not observed. In a situation like this, all structural parameters β, δ, and γ could in principle be estimated by some nonlinear techniques. For instance, Chevillon et al. (2010) use nonlinear least squares while Adam et al. (2016) employ the method of simulated

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

moments. The general properties of these estimation procedures in the context of adaptive learning models are not clear a priori, however. Further research on these estimators is imperative.

4.2. Forward-Looking Expectations

Some economic theories stipulate that agents' expectations are forward-looking, i.e., agents are thought of as forming expectations of a future value of the endogenous variable. For instance, agents might forecast y_{t+1} using information available at time t. This would amount to replacing the expectational term in model (1.1) by $y^e_{t+1|t}$, yielding

$$y_t = \beta y^e_{t+1|t} + \delta x_t + \varepsilon_t. \tag{4.3}$$

However, this type of modified model setup raises a number of issues, both economic and econometric.

The most immediate economic consequence of forward-looking models is that rational expectations equilibria are, in general, no longer unique. For instance, consider model (4.3), define the information set $\mathcal{I}_t = \sigma(y_s, s \leq t; x_s, s \leq t)$, and assume that x_t follows an i.i.d. process with mean μ_x. This implies that $\mathbf{E}(y_{t+1} \mid \mathcal{I}_t) = \alpha\mu_x$ and it is well-known that the resulting rational expectations solution $y_t = \alpha\mu_x + \varepsilon_t$ is only one of many, namely the minimum state variable solution. As an added complication, not all solutions need to be stable. As forcefully argued by Evans and Honkapohja (2001), however, one way to address the multiplicity issue is that adaptive learning can serve as a selection device, in that equilibria that are learnable appear more plausible than others. Moreover, unstable equilibria can be ruled out by focusing on those that satisfy the conditions for expectational stability, or so-called E-stability, a concept introduced by Evans (1989) to characterise those rational expectations solutions that are stable under learning.

A further economic issue with the model in (4.3) that has important repercussions for the analysis is that the use of $\mathcal{I}_t$ introduces simultaneity in the system, since agents' expectations are now endogenous, being partly based on the current value of y_t. However, it is common practice in parts of the economic theory literature, see for instance that on the indeterminacy of equilibria, to avoid this problem by, conversely, assuming that y_t is not observable by agents after all and that they base their prediction on $\mathcal{I}_{t-1}$, or indeed $\mathcal{F}_{t-1}$, instead. Bullard and Mitra (2002), for instance, criticise this practice and argue that the use of $\mathcal{I}_{t-1}$ would be informationally inconsistent within the model and that a switch between $\mathcal{I}_t$ and $\mathcal{I}_{t-1}$ in a system with adaptive learning might lead to a reversal of the stability properties of the equilibria. A solution to this issue is suggested by Adam (2003) who allows agents to make use of information $\mathcal{I}_t$ yet, through the introduction of sticky prices, avoids any simultaneity in the system.

The results of the present paper should therefore carry over to models with forward-looking expectations that are E-stable and that incorporate suitable informational or timing assumptions. OLS estimation would, however, no longer

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

be appropriate if forward-looking expectations create a simultaneity problem, and future work will have to address the properties of, for instance, instrumental variable techniques.

5. CONCLUSION

This paper investigates the asymptotic properties of the OLS estimators of the structural parameters in a stereotypical macroeconomic model in which agents form expectations by means of constant or decreasing gain adaptive learning. For both types of learning, we first provide a complete analysis of the asymptotics of agents' expectations, considering its convergence to the rational expectations equilibrium and, if appropriate, its convergence rate. The latter aspect has so far not been addressed comprehensively in the literature. In a second step, the asymptotic properties of the OLS estimators are derived for both constant and decreasing gain learning. This is both a novel and a challenging undertaking. While empirical models of this type are becoming ever more popular in the literature, the asymptotics of OLS are complex and this paper is the first to address them systematically.

The conclusions are potentially wide-reaching: It turns out that, contrary to what might be believed at first sight, the OLS estimators are consistent in all scenarios we consider. Their asymptotic distributions, however, are highly nonstandard in all but one setting, so care must be taken when inference is to be conducted. Yet our results indicate that an econometric analysis of such models with adaptive learning can, at a fundamental level, be feasible and sound. This is encouraging, given the push by policy makers and applied economists to exploit these methods empirically.

A complete analysis of forward-looking expectations and time-varying regressor sequences is beyond the scope of the present paper and is left to future research. Similarly, models that are nonlinear in the parameters and thus require alternative estimators such as nonlinear least squares will have to be looked at in future investigations.

NOTES

1. The issue of *strong* consistency of the OLS estimator is investigated in a companion paper, see Christopeit and Massmann (2016).
2. We are grateful to George Evans for drawing our attention to this interpretation.

REFERENCES

Abadir, K.M. (1993) The limiting distribution of the autocorrelation coefficient under a unit root. *Annals of Statistics* 21(2), 1058–1070.

Adam, K. (2003) Learning and equilibrium selection in a monetary overlapping generations model with sticky prices. *Review of Economic Studies* 70, 887–908.

Adam, K., A. Marcet, & J.P. Nicolini (2016) Stock market volatility and learning. *Journal of Finance* 71, 33–82.

Apostol, T.M. (1974) *Mathematical Analysis*, 2nd ed. Addison-Wesley.

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

Benveniste, A., M. Métivier, & P. Priouret (1990) *Adaptive Algorithms and Stochastic Approximation*. Springer. Orginally published in French in 1987.

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd ed. Wiley.

Branch, W.A. & G.W. Evans (2006) A simple recursive forecasting model. *Economics Letters* 91, 158–166.

Branch, W.A. & G.W. Evans (2010) Asset return dynamics and learning. *Review of Financial Studies* 23(4), 1651–1680.

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

Brock, W.A. & C.H. Hommes (1997) A rational route to randomness. *Econometrica* 65(5), 1059–1095.

Bullard, J.B. & K. Mitra (2002) Learning about monetary policy rules. *Journal of Monetary Economics* 49, 1105–1129.

Chan, N.H. & C.Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16(1), 367–401.

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

Chow, Y.S. & H. Teicher (1973) Iterated logarithm laws for weighted averages. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 26, 87–94.

Christopeit, N. (1986) Quasi-least-squares estimation in semimartingale regression models. *Stochastics* 16, 255–278.

Christopeit, N. & K. Helmes (1980) Strong consistency of least squares estimators in linear regression models. *Annals of Statistics* 8, 778–788.

Christopeit, N. & M. Massmann (2010) Consistent estimation of structural parameters in regression models with adaptive learning. Discussion Paper 10-077/4, Tinbergen Institute.

Christopeit, N. & M. Massmann (2013) Estimating structural parameters in regression models with adaptive learning. Discussion Paper 13-111/III, Tinbergen Institute.

Christopeit, N. & M. Massmann (2016) Strong Consistency of the Least-Squares Estimator in Simple Regression Models with Stochastic Regressors. Mimeo WHU – Otto Beisheim School of Management.

Clarida, R., J. Galí, & M. Gertler (2000) Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics* 115, 147–180.

Davidson, R. & J.G. MacKinnon (1993) *Estimation and Inference in Econometrics*. Oxford University Press.

Dieppe, A., A. González Pandiella, S. Hall, & A. Willman (2011) The ECB's New Multi-country Model for the Euro Area. Working paper 1316, European Central Bank.

Evans, G.W. (1989) The fragility of sunspots and bubbles. *Journal of Monetary Economics* 23, 297–317.

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

Evans, G.W., S. Honkapohja, T.J. Sargent, & N. Williams (2013) Bayesian model averaging, learning and model selection. In T.J. Sargent & J. Vilmunen (eds.), *Macroeconomics at the Service of Public Policy*, Chapter 6, pp. 99–119. Oxford University Press.

Evans, G.W. & G. Ramey (2006) Adaptive expectations, underparameterization and the Lucas critique. *Journal of Monetary Economics* 53, 249–264.

Fourgeaud, C., C. Gourieroux, & J. Pradel (1986) Learning procedures and convergence to rationality. *Econometrica* 54(4), 845–868.

Gaspar, V., F. Smets, & D. Vestin (2010) Inflation expectations, adaptive learning and optimal monetary policy. In B.M. Friedman & M. Woodford (eds.), *Handbook of Monetary Economics*, vol. 3B, pp. 1055–1095. Elsevier.

Grenander, U. & M. Rosenblatt (1957) *Statistical Analysis of Stationary Time Series*. Wiley.

Hommes, C.H. (2002) Modelling the stylized facts in finance through simple adaptive systems. *Proceedings of the National Academy of Science of the USA* 99(S3), 7221–7228.

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

Kottmann, T. (1990) Learning Procedures and Rational Expectations in Linear Models with Forecast Feedback. Ph.D. thesis, University of Bonn.

Krishnan, M. (1968) Series expansions of the doubly noncentral t-distribution. *Journal of the American Statistical Association* 63, 1104–1012.

Kushner, H. (1984) *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory*. MIT Press.

Kushner, H. (2010) Stochastic approximation: A survey. *Wiley Interdisciplinary Reviews: Computational Statistics* 2(1), 87–96.

Lai, T.L. (1989) Extended stochastic Lyapunov functions and recursive algorithms in linear stochastic systems. In N. Christopeit, K. Helmes, & M. Kohlmann (eds.), *Stochastic Differential Systems*, pp. 206–220. Springer.

Lai, T.L. (2003) Stochastic approximation. *Annals of Statistics* 31, 391–406.

Lai, T.L. & C.Z. Wei (1982a) Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics* 10, 154–166.

Lai, T.L. & C.Z. Wei (1982b) Asymptotic properties of projections with applications to stochastic regression problems. *Journal of Multivariate Analysis* 12, 346–370.

Lai, T.L. & C.Z. Wei (1983) A note on martingale difference sequences satisfying the local Marcinkiewicz-Zygmund condition. *Bulletin of the Institute of Mathematics, Academica Sinica* 11(1), 1–13.

Lai, T.L. & C.Z. Wei (1985) Asymptotic properties of multivariate weighted sums with applications to stochastic linear regression in linear dynamic stystems. In P.R. Krishnaiah (ed.), *Multivariate Analysis*, vol. VI, pp. 375–393. North-Holland.

Ljung, L. (1977) Analysis of recursive stochastic systems. *IEEE Transactions on Automatic Control* AC-22, 551–575.

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

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

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

Marcet, A. & T.J. Sargent (1988) The fate of systems with "adaptive" learning. *American Economic Review* 78, 168–172.

Marcet, A. & T.J. Sargent (1995) Speed of convergence of recursive least squares: learning with autoregressive moving-average perceptions. In A. Kirman & M. Salmon (eds.), *Learning and Rationality in Economics*, Chapter 6, pp. 179–215. Blackwell.

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

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

Muth, J.F. (1961) Rational expectations and the theory of price moevements. *Econometrica* 29, 315–335.

Newey, W.K. & D.L. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2111–2245. Elsevier.

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

Phillips, P.C.B. (2007) Regression with slowly varying regressors and nonlinear trends. *Econometric Theory* 23, 557–614.

Phillips, P.C.B. & T. Magdalinos (2007) Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.

Phillips, P.C.B. & T. Magdalinos (2008) Limit theory for explosively cointegrated systems. *Econometric Theory* 24, 865–887.

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

Preston, B. (2005) Learning about monetary policy rules when long-horizon expectations matter. - *International Journal of Central Banking* 1(2), 81–126.
Robbins, H. & S. Monro (1951) A stochastic approximation method. *Annals of Mathematical Statistics* 22(3), 400–407.
Robbins, H. & D. Siegmund (1971) A convergence theorem for non-negative almost-supermartingales and some applications. In J.S. Rustagi (ed.), *Optimizing Methods in Statistics*, pp. 233–257. Academic Press.
Roberts, J.M. (1995) New Keynesian economics and the Phillips curve. *Journal of Money, Credit and Banking* 27, 975–984.
Sargent, T.J. (1993) *Bounded Rationality in Macroeconomics*. Clarendon Press.
Sargent, T.J. (1999) *The Conquest of American Inflation*. Princeton University Press.
Sargent, T.J. (2008) Rational expectations. In S.N. Durlauf & L.E. Blume (eds.), *The New Palgrave Dictionary of Economics*, 2nd ed., pp. 877–882 Palgrave. Macmillan, 1st edition 1987.
Shiryaev, A.N. (1996) *Probability*, 2nd ed. Springer, 1st edition 1984.
Shiryaev, A. & V. Spokoiny (1997) On sequential estimation of an autoregressive parameter. *Stochastics and Stochastic Reports* 60(3–4), 219–240.
Solo, V. & X. Kong (1995) *Signal Processing Algorithms*. Prentice Hall.
Vives, X. (1993) How fast do rational agents learn? *Review of Economic Studies* 60, 329–347.
Walk, H. & L. Zsidó (1989) Convergence of the Robbins-Monro method for linear problems in a Banach space. *Journal of Mathematical Analysis and Applications* 139, 152–177.
Wang, X. & J. Yu (2015) Limit theory for an explosive autoregressive process. *Economics Letters* 126, 176–180.
West, K.D. (1988) Asymptotic normality, when regressors have a unit root. *Econometrica* 56, 1397–1417.
White, J.S. (1958) The limiting distribution of the serial correlation coefficients in the explosive case. *Annals of Mathematical Statistics* 29, 1188–1197.

APPENDIX A: Proofs for Constant Gain

A.1. Working Formula for a_t

If $\beta \neq 1$, it will sometimes turn out favourable to work with the transformed process $a_t^{\#} = a_t - \alpha$. For then (2.2) simplifies to

$$a_t^{\#} = (1-c)\, a_{t-1}^{\#} + \gamma\, \varepsilon_t . \tag{A.1}$$

We will also make repeated use of the solution of a_t in (2.2):

$$\begin{aligned} a_t &= (1-c)^t\, a_0 + \sum\nolimits_{i=0}^{t-1} (1-c)^i \left(c\alpha + \gamma\, \varepsilon_{t-i}\right) \\ &= \left[1-(1-c)^t\right]\alpha + (1-c)^t\, a_0 + \gamma \sum\nolimits_{i=0}^{t-1} (1-c)^i\, \varepsilon_{t-i} \end{aligned} \tag{A.2}$$

and of $a_t^{\#}$ in (A.1):

$$\begin{aligned} a_t^{\#} &= (1-c)^t\, a_0^{\#} + \gamma \sum\nolimits_{i=0}^{t-1} (1-c)^i\, \varepsilon_{t-i} \\ &= (1-c)^t\, a_0^{\#} + \gamma\, (1-c)^t \sum\nolimits_{j=1}^{t} (1-c)^{-j}\, \varepsilon_j , \end{aligned} \tag{A.3}$$

which are valid for all initial values a_0.

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

A.2. Proof of Theorem 1

A.2.1. Stable Case: $0 < c < 2$. In this case, $|1-c| < 1$, and the assertion is a classical standard result for stable AR-systems.

A.2.2. Unit Root Case: $c = 0$ *or* $c = 2$. If $c = 0$ then, by (2.2), $a_t = a_{t-1} + \gamma\,(\delta + \varepsilon_t)$. Hence a_t is a random walk with drift:

$$a_t = a_0 + \gamma\,(\delta t + S_t),$$

where $S_t = \sum_{i=1}^{t} \varepsilon_i$. As a consequence, by the LLN, $t^{-1}a_t = \gamma\,\delta + o(1)$ a.s.

If $c = 2$ then (A.2) becomes

$$\begin{aligned} a_t &= \left[1-(-1)^t\right]\alpha + (-1)^t a_0 + \gamma \sum_{i=0}^{t-1} (-1)^i \varepsilon_{t-i} \\ &= \left[1-(-1)^t\right]\alpha + (-1)^t a_0 + \gamma\,(-1)^t \sum_{j=1}^{t} (-1)^{-j} \varepsilon_j. \end{aligned}$$

Since the $\widetilde{\varepsilon}_j = (-1)^{-j}\varepsilon_j$ are i.i.d., (2.3) follows from the classical CLT applied to $\widetilde{S}_t = \sum_{j=1}^{t} \widetilde{\varepsilon}_j$ together with the symmetry of the limit distribution.

A.2.3. Explosive Case: $c < 0$ *or* $c > 2$. In this case, $|1-c| > 1$. Recall (A.3) and consider the martingale

$$L_t = \sum_{j=1}^{t} (1-c)^{-j} \varepsilon_j.$$

Since its predictable quadratic variation is

$$\langle L \rangle_t = \sigma^2 \sum_{j=1}^{t} (1-c)^{-2j},$$

and hence $\langle L \rangle_\infty < \infty$ a.s., it follows from the martingale convergence theorem that the limit

$$L = \lim_{t\to\infty} L_t = \sum_{j=1}^{\infty} (1-c)^{-j} \varepsilon_j$$

exists with probability one and also in L^2. Its variance is $1/[(1-c)^2 - 1]$. Therefore,

$$(1-c)^{-t}\, a_t^{\#} = a_0^{\#} + \gamma\, L_t \to a_0^{\#} + \gamma\, L$$

and hence

$$(1-c)^{-t}\, a_t \to a_0 - \alpha + \gamma\, L$$

both a.s. and in L^2.

Remark A.1. L has a continuous distribution function. This follows from results of Lai and Wei (1983, Corollary 3 and 4) and Lai and Wei (1985, Lemma 2). Actually, in these papers it is assumed that $\varepsilon_t \in L^p$ for some $p > 2$, in which case the results are true for martingale difference sequences ε_t. In the i.i.d. case considered here, finite second moments do suffice. This is elaborated upon in Section H of the Online Supplement.

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

A.3. Proof of Theorem 2

Throughout this section, we will make use of the observations made at the beginning of Section 2.2, namely that the asymptotic behaviour of the OLS estimator in (2.1) is, up to a factor of proportionality, equal to that of the OLS estimator in the AR(1)-model (2.7).

A.3.1. Stable Case: $0 < c < 2$. In this case, $|\beta^*| < 1$. It is standard textbook knowledge that

$$\sqrt{T}(\widehat{\theta}_T^* - \theta^*) \xrightarrow{d} \mathcal{N}\big(0, \gamma^2 \sigma^2 \overline{M}^{-1}\big)$$

with $\overline{M} = \operatorname{plim}_{T\to\infty} \frac{1}{T} M_T$ and M_T being given in (2.5). Since

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} a_t = \mathbf{E} a_t = \alpha, \quad \operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} a_t^2 = \mathbf{E} a_t^2 = \sigma_a^2 + \alpha^2,$$

(2.10) is an immediate consequence of (2.9).

A.3.2. Unit Root Case: $c = 0$ *or* $c = 2$.

Case $c = 0$**.** Consider first the sub-case $\delta \neq 0$. Making use of the OLS estimator in (2.4), the asymptotic behaviour of a_t as established in Theorem 1(*ii*) implies that, with probability one,

$$\lim_{T\to\infty} \frac{1}{T^2} \sum_{t=1}^{T} a_{t-1} = \frac{\delta\gamma}{2}, \quad \lim_{T\to\infty} \frac{1}{T^3} \sum_{t=1}^{T} a_{t-1}^2 = \frac{\delta^2 \gamma^2}{3}.$$

Therefore, remembering (2.5) and introducing the normalising matrix

$$D_T = \begin{pmatrix} \sqrt{T} & 0 \\ 0 & T^{3/2} \end{pmatrix},$$

we find that

$$\lim_{T\to\infty} D_T^{-1} M_T D_T^{-1} = \begin{pmatrix} 1 & \frac{\delta\gamma}{2} \\ \frac{\delta\gamma}{2} & \frac{\delta^2\gamma^2}{3} \end{pmatrix} = \overline{M} \text{ a.s.}$$

Applying the CLT for bivariate martingales to (2.6), one obtains

$$D_T^{-1} u_T = \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \\ \frac{1}{T^{3/2}} \sum_{t=1}^{T} a_{t-1} \varepsilon_t \end{pmatrix} \xrightarrow{d} \mathcal{N}\big(0, \sigma^2 \overline{M}\big).$$

Therefore,

$$\begin{pmatrix} \sqrt{T}(\widehat{\delta} - \delta) \\ T^{3/2}(\widehat{\beta} - 1) \end{pmatrix} = \Big[D_T^{-1} M_T D_T^{-1}\Big]^{-1} D_T^{-1} u_T \xrightarrow{d} \mathcal{N}\big(0, \sigma^2 \overline{M}^{-1}\big).$$

Calculating the inverse yields (2.11).

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

Consider now the second sub-case, i.e., $\delta = 0$. It is of course well-known that the asymptotic behaviour of the OLS estimator is very different according to whether $\delta \neq 0$ or $\delta = 0$. This is also obvious from the different behaviour of a_t as established in Theorem 1(*ii*), viz. a_t grows linearly in the former case and tends to zero in the latter. Recall from the discussion in Section 2.2 that the OLS estimator $\widehat{\theta} = (\widehat{\delta}, \widehat{\beta})'$ of the parameters $\theta = (\delta, \beta)'$ in the structural model (2.1) satisfies

$$\widehat{\theta} - \theta = \gamma^{-1}\left(\widehat{\theta}^* - \theta^*\right),$$

where $\widehat{\theta}^* = (\widehat{\delta}^*, \widehat{\beta}^*)'$ is the OLS estimator for the parameters $\theta^* = \left(\delta^*, \beta^*\right)'$ in the AR(1) model (2.7) provided that $\delta^* = \gamma\,\delta$ and $\beta^* = \beta$. As pointed out earlier, this AR(1)-model is the Dickey–Fuller model with drift, evaluated at the joint null $\delta^* = 0, \beta^* = 1$. As is then well-known,

$$\begin{pmatrix} \sqrt{T}\widehat{\delta}^* \\ T(\widehat{\beta}^* - 1) \end{pmatrix} \xrightarrow{d} Z^*,$$

where

$$Z^* = \upsilon^{-1}\begin{pmatrix} \sigma\gamma\, B(1)\int_0^1 B(s)^2 ds - \frac{\sigma\gamma}{2}\left(B(1)^2 - 1\right)\int_0^1 B(s)ds \\ \frac{1}{2}\left(B(1)^2 - 1\right) - B(1)\int_0^1 B(s)ds \end{pmatrix},$$

$$\upsilon = \int_0^1 B(s)^2 ds - \left(\int_0^1 B(s)ds\right)^2.$$

Taking account of (2.8) yields the desired result.

Case $c = 2$**.** In this case, according to (2.2), the dynamics of a_t are given by

$$a_t = 2\alpha - a_{t-1} + \gamma\,\varepsilon_t, \tag{A.4}$$

with DGP as in (2.1):

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t. \tag{A.5}$$

Introducing $z_t = a_t - \alpha$, (A.4) is equivalent to

$$z_t = -z_{t-1} + \gamma\,\varepsilon_t. \tag{A.6}$$

Define variables $\widetilde{\varepsilon}_j = (-1)^{-j}\varepsilon_j$ and

$$S_t = \sum_{j=1}^{t}\varepsilon_j, \quad \widetilde{S}_t = \sum_{j=1}^{t}\widetilde{\varepsilon}_j,$$

together with random functions

$$X_T(u) = \frac{1}{\sigma\sqrt{T}}S_{[uT]}, \quad \widetilde{X}_T(u) = \frac{1}{\sigma\sqrt{T}}\widetilde{S}_{[uT]}, \quad \text{for } 0 \leq u \leq 1.$$

It is shown in Chan and Wei (1988, Theorem 2.2) that

$$(X_T, \widetilde{X}_T) \Rightarrow (B, \widetilde{B}) \tag{A.7}$$

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

in $D[0,1]^2$, where B and $\widetilde{B}$ are *independent* standard Brownian motions (BM). The important fact about this is *joint* weak convergence.

For the OLS estimator

$$\begin{pmatrix} \widehat{\delta}-\delta \\ \widehat{\beta}-\beta \end{pmatrix} = M_T^{-1} u_T$$

of β and δ in (A.5), we need the joint asymptotics of the first two sample moments of a_t together with u_T.

As to the mean, it is easily verified,

$$z_t = \gamma\,(-1)^t\,\widetilde{S}_t. \tag{A.8}$$

Hence

$$\sum_{t=1}^{T} z_{t-1} = \gamma \sum_{t=1}^{T} (-1)^{t-1}\,\widetilde{S}_{t-1} = \gamma\, W_T.$$

By partial summation,

$$W_T = \sum_{t=1}^{T} (-1)^{t-1}\,\widetilde{S}_{t-1} = \sigma_T \widetilde{S}_T - \sum_{t=1}^{T} \sigma_t \widetilde{\varepsilon}_t,$$

where we have put

$$\sigma_t = \sum_{j=1}^{t} (-1)^{j-1} = \begin{cases} 1, & t \text{ odd}, \\ 0, & t \text{ even}. \end{cases}$$

Since $T^{-1/2}\widetilde{S}_T \xrightarrow{d} \mathcal{N}(0,\sigma^2)$ and $\mathbf{E}[\sum_{t=1}^{T} \sigma_t \widetilde{\varepsilon}_t]^2 \le T\sigma^2$, it follows that $W_T = O_p\big(T^{1/2}\big)$. Hence $\operatorname{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_{t-1} = 0$ and

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} a_{t-1} = \alpha. \tag{A.9}$$

Regarding the second moments, note that by virtue of (A.8) and the continuous mapping theorem,

$$\frac{1}{T^2} \sum_{t=1}^{T} z_{t-1}^2 = \frac{\gamma^2}{T^2} \sum_{t=1}^{T} \widetilde{S}_{t-1}^2 = \gamma^2\sigma^2 \int_0^1 \widetilde{X}_T^2(s-)ds \xrightarrow{d} \gamma^2\sigma^2 \int_0^1 \widetilde{B}(s)^2 ds.$$

Since

$$\frac{1}{T^2} \sum_{t=1}^{T} a_{t-1}^2 = \frac{1}{T^2} \sum_{t=1}^{T} z_{t-1}^2 + 2\alpha \frac{1}{T^2} \sum_{t=1}^{T} z_{t-1} + \frac{\alpha^2}{T},$$

we have

$$\frac{1}{T^2} \sum_{t=1}^{T} a_{t-1}^2 \xrightarrow{d} \gamma^2\sigma^2 \int_0^1 \widetilde{B}(s)^2 ds. \tag{A.10}$$

Introducing the normalising matrix

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$D_T = \begin{pmatrix} \sqrt{T} & 0 \\ 0 & T \end{pmatrix},$$

we then find that

$$D_T^{-1} M_T D_T^{-1} = \begin{pmatrix} 1 & \frac{1}{T^{3/2}} \sum_{t=1}^{T} a_{t-1} \\ \frac{1}{T^{3/2}} \sum_{t=1}^{T} a_{t-1} & \frac{1}{T^2} \sum_{t=1}^{T} a_{t-1}^2 \end{pmatrix}$$
$$\xrightarrow{d} \overline{M} = \begin{pmatrix} 1 & 0 \\ 0 & \gamma^2 \sigma^2 \int_0^1 \widetilde{B}(s)^2 ds \end{pmatrix}. \tag{A.11}$$

Now turn to u_T as defined in (2.6). The first component is just the standard CLT:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \xrightarrow{d} \sigma B(1). \tag{A.12}$$

As to the second component of u_T, we make use of the trivial identity $(-1)^{t-1} \varepsilon_t = -\widetilde{\varepsilon}_t$. Then

$$\sum_{t=1}^{T} z_{t-1} \varepsilon_t = \gamma \sum_{t=1}^{T} (-1)^{t-1} \widetilde{S}_{t-1} \varepsilon_t = -\gamma \sum_{t=1}^{T} \widetilde{S}_{t-1} \widetilde{\varepsilon}_t. \tag{A.13}$$

Writing

$$\widetilde{S}_t^2 = \left(\widetilde{S}_{t-1} + \widetilde{\varepsilon}_t\right)^2 = \widetilde{S}_{t-1}^2 + 2\widetilde{S}_{t-1}\widetilde{\varepsilon}_t + \widetilde{\varepsilon}_t^2,$$

it follows that

$$\sum_{t=1}^{T} \widetilde{S}_{t-1} \widetilde{\varepsilon}_t = \frac{1}{2}\left[\widetilde{S}_T^2 - \widetilde{S}_0^2\right] - \frac{1}{2} \sum_{t=1}^{T} \widetilde{\varepsilon}_t^2$$

or

$$\frac{1}{T} \sum_{t=1}^{T} \widetilde{S}_{t-1} \widetilde{\varepsilon}_t = \frac{\sigma^2}{2} \widetilde{X}_T^2(1) - \frac{1}{2T} \sum_{t=1}^{T} \widetilde{\varepsilon}_t^2.$$

Therefore, in view of (A.7),

$$\frac{1}{T} \sum_{t=1}^{T} z_{t-1} \varepsilon_t = -\frac{\gamma}{T} \sum_{t=1}^{T} \widetilde{S}_{t-1} \widetilde{\varepsilon}_t \xrightarrow{d} -\frac{\sigma^2 \gamma}{2}\left[\widetilde{B}^2(1) - 1\right].$$

In terms of a_t this means that

$$\frac{1}{T} \sum_{t=1}^{T} a_{t-1} \varepsilon_t = \frac{1}{T} \sum_{t=1}^{T} z_{t-1} \varepsilon_t + \alpha \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \xrightarrow{d} -\frac{\sigma^2 \gamma}{2}\left[\widetilde{B}^2(1) - 1\right]. \tag{A.14}$$

Since all entries of M_T, u_T may be expressed as functions of the jointly convergent random functions $(X_T, \widetilde{X}_T)$, the weak convergence of M_T, u_T is jointly in $D[0,1]^2$. This allows us to conclude that

$$D_T^{-1} u_T = \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \\ \frac{1}{T} \sum_{t=1}^{T} a_{t-1} \varepsilon_t \end{pmatrix} \xrightarrow{d} \overline{u} = \begin{pmatrix} \sigma B(1) \\ -\frac{\sigma^2 \gamma}{2}\left[\widetilde{B}^2(1) - 1\right] \end{pmatrix}$$

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

and

$$\begin{pmatrix} \sqrt{T}(\widehat{\delta}-\delta) \\ T(\widehat{\beta}-\beta) \end{pmatrix} = D_T \begin{pmatrix} \widehat{\delta}-\delta \\ \widehat{\beta}-\beta \end{pmatrix} = \left[D_T^{-1} M_T D_T^{-1}\right]^{-1} D_T^{-1} u_T \xrightarrow{d} \overline{M}^{-1}\overline{u},$$

with

$$\overline{M}^{-1}\overline{u} = \sigma \begin{pmatrix} 1 & 0 \\ 0 & \gamma^2\sigma^2 \int_0^1 \widetilde{B}(s)^2 ds \end{pmatrix}^{-1} \begin{pmatrix} \sigma B(1) \\ -\frac{\sigma^2\gamma}{2}\left[\widetilde{B}^2(1)-1\right] \end{pmatrix}$$

$$= \frac{1}{\int_0^1 \widetilde{B}(s)^2 ds} \begin{pmatrix} \sigma B(1) \int_0^1 \widetilde{B}(s)^2 ds \\ -\frac{1}{2\gamma}\left[\widetilde{B}^2(1)-1\right] \end{pmatrix}.$$

Remark A.2. Actually, Theorem 2.2 in Chan and Wei requires that $\mathbf{E}|\varepsilon_t|^{2+\alpha} < \infty$ for some $\alpha > 0$. This is because it is formulated for martingale difference sequences ε_t. In our case, where the errors are i.i.d. with finite second moments, it can easily be verified directly using the standard multivariate CLT for such variables. The clue is that, for every $h \geq 0$,

$$\mathbf{E} S_T \widetilde{S}_{T+h} = \mathbf{E} S_T \widetilde{S}_T = \begin{cases} 0, & T \text{ even}, \\ -1, & T \text{ odd}. \end{cases}$$

As a consequence, denoting $Z_T = (X_T, \widetilde{X}_T)$, it follows from the CLT that the distributions of $(Z_T(t_1), \ldots, Z_T(t_k)), 0 \leq t_1 \leq \cdots \leq t_k \leq 1$, converge to a $2k$-variate normal distribution with covariance matrix identical to that of $(Z(t_1), \ldots, Z(t_k))$, where $Z = (B, \widetilde{B})$ and $B, \widetilde{B}$ are independent standard Brownian motions. Moreover, the distribution measures of the Z_T are tight in $D[0,1]^2$ since the marginal distributions of X_T and $\widetilde{X}_T$ are tight in $D[0,1]$, cf. Billingsley (1999, Problem 5.9).

A.3.3. Explosive Case: $c < 0$ or $c > 2$. In this case, the corresponding AR(1)-model is

$$a_t = c\alpha + (1-c)\, a_{t-1} + \gamma\, \varepsilon_t.$$

According to Wang and Yu (2015), the asymptotic distribution of the OLS estimator $\widehat{\theta}^*$ of $\delta^* = c\alpha$ and $\beta^* = 1-c$ in (2.7) is given by

$$\sqrt{T}\left(\widehat{\delta}^* - \delta^*\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\sigma^2\right),$$

$$(1-c)^T\left(\widehat{\beta}^* - \beta^*\right) \xrightarrow{d} \frac{\left[(1-c)^2-1\right]u}{v + (1-c)\, c\alpha / [(1-c)-1]}, \tag{A.15}$$

where u and v are independent L^2-variables, and the two estimators are asymptotically independent. Premultiplying by γ^{-1} yields the assertion.

Remark A.3. Actually, the random variable v is the limit

$$\alpha\left[\sum_{t=1}^{\infty}(1-c)^{-t}\varepsilon_t + a_0\right] = \alpha(z + a_0)$$

a.s. and in L^2. Arguing as in Remark A.1 in A.2.3, z has a continuous distribution function. Therefore the right hand side of (A.15) is well defined.

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

APPENDIX B: Proofs for Decreasing Gain

B.1. Preliminaries

We return now to the model (3.1)–(3.2), with c again as in (1.13). In the following, we will derive a finite moving-average form for a_t in terms of its initial value a_0 and a weighted sum of the errors. It will simplify the calculations if we centre a_t about its asymptotic value $\alpha = \delta/(1-\beta) = \gamma\,\delta/c$. To this end, introduce the centred variable $a_t^{\#} = a_t - \alpha$. Then, by straightforward calculation,

$$a_t^{\#} = \left(1 - \frac{c}{t}\right) a_{t-1}^{\#} + \frac{\gamma}{t}\varepsilon_t. \tag{B.1}$$

Also,

$$y_t = \delta + \beta\left(a_{t-1} - \alpha\right) + \alpha\beta + \varepsilon_t = \alpha + \beta a_{t-1}^{\#} + \varepsilon_t \tag{B.2}$$

since $\delta + \alpha\beta = \alpha$. Henceforth, let us rename the $a_t^{\#}$ as a_t. The theorems in Section 3 are of course presented in terms of the original a_t.

The solution to (B.1) is given by

$$a_t = a_0\phi_{t0} + \gamma \sum_{i=1}^{t} \phi_{ti}\frac{\varepsilon_i}{i}, \tag{B.3}$$

where

$$\phi_{ti} = \left(1 - \frac{c}{i+1}\right)\cdots\left(1 - \frac{c}{t}\right), \quad i = 0, \ldots, t-1, \tag{B.4}$$

and $\phi_{tt} = 1$. In particular, for $c = 1, \phi_{ti} = i/t$ for all i. Define $i_0 = [c] =$ smallest integer $\geq c$. Then, for $i \geq i_0$, the factors in (B.4) are all positive. Note that $i_0 = 0$ for $c < 1$. Taking logarithms and using a first order Taylor expansion, we obtain for $i \geq i_0$

$$\ln\phi_{ti} = \sum_{j=i+1}^{t} \ln\left(1 - \frac{c}{j}\right) = -c\sum_{j=i+1}^{t}\frac{1}{j} - \sum_{j=i+1}^{t}\frac{R_j}{j^2}, \tag{B.5}$$

with $R_j = \left(1-\vartheta_j\right)c^2 \Big/ \left(1-\vartheta_j c/j\right)^2$ for some $0 < \vartheta_j < 1$ (Cauchy form of the remainder). Note that $R_j \geq 0$ and

$$\sup_{j \geq i_0+1} R_j \leq \frac{c^2}{\left(1 - \frac{c}{i_0+1}\right)^2}. \tag{B.6}$$

Hence

$$\ln\phi_{ti} = -c\sum_{j=i+1}^{t}\frac{1}{j} - O_{ti}(1)\sum_{j=i+1}^{t}\frac{1}{j^2},$$

with the $O_{ti}(1)$-term uniformly bounded in i, t. Henceforth, we will denote such terms simply by $O(1)$. Making use of the integral comparison test (ICT),

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$\ln \phi_{ti} = -c\,(\ln t - \ln(i+1)) + O(1)\frac{1}{i+1}$$
$$= -c\,(\ln t - \ln i) + O(1)\frac{1}{i}$$

since $\ln(i+1) = \ln i + \ln(i+1)/i = \ln i + O(1/i)$. As a consequence, for $i \geq i_0 > 0$,

$$\phi_{ti} = \left(\frac{i}{t}\right)^c \left[1 + O\left(\frac{1}{i}\right)\right]. \tag{B.7a}$$

For $0 \leq i < i_0$, it follows from (B.4) and (B.7a) that

$$|\phi_{ti}| = \prod_{j=i+1}^{i_0-1} \left|1 - \frac{c}{j}\right| \prod_{j=i_0}^{t} \left(1 - \frac{c}{j+1}\right)$$
$$\leq C\phi_{ti_0} = O\left(t^{-c}\right). \tag{B.7b}$$

In particular, $|\phi_{t0}| = O\left(t^{-c}\right)$. Actually, for $c < 1$ (i.e., $i_0 = 0$), (B.7a) remains valid also for $i = 0$ (in the form $\phi_{ti} = O\left(t^{-c}\right)$). But in this case we can do better for all $\phi_{ti}, i \geq 0$, (and indeed have to for $c < 1/2$). To see this, we go back to (B.5):

$$\ln \phi_{ti} = -c \sum_{j=i+1}^{t} \frac{1}{j} - \sum_{j=i+1}^{t} \frac{R_j}{j^2}. \tag{B.8}$$

In the sequel, we will make frequent use of the following version of the integral comparison theorem (ICT), which we cite here for convenience, cf. Apostol (1974, Theorem 8.33).

PROPOSITION B.1 (ICT). *Let $f(x)$ be a positive decreasing function s.t. $\lim_{x\to\infty} f(x) = 0$. For $n \geq i, i = 1, 2, \ldots$, define*

$$s_{ni} = \sum_{k=i}^{n} f(k), \quad t_{ni} = \int_i^n f(x)dx, \quad d_{ni} = s_{ni} - t_{ni}.$$

Then we have:

(*i*) $0 < f(n+1) \leq d_{n+1,i} \leq d_{ni} \leq f(i)$.
(*ii*) $d_i = \lim_{n\to\infty} d_{ni}$ *exists, and* $d_i \leq f(i)$.
(*iii*) $0 \leq d_{ni} - d_i \leq f(n)$.

By ICT (*ii*), the limit

$$E_i = \lim_{t\to\infty} \left[\sum_{j=i+1}^{t} \frac{1}{j} - \int_{i+1}^{t} \frac{dx}{x}\right]$$

exists, and $0 \leq E_i \leq 1/(i+1)$. For $i = 0$, E_0 is just the Euler constant. Making use of ICT (*iii*), we obtain

$$0 \leq \sum_{j=i+1}^{t} \frac{1}{j} - \int_{i+1}^{t} \frac{dx}{x} - E_i \leq \frac{1}{t},$$

so that we may write

$$\sum_{j=i+1}^{t} \frac{1}{j} = \ln t - \ln(i+1) + E_i + O_i\left(\frac{1}{t}\right). \tag{B.9}$$

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

As to the second term on the right hand side of (B.8),

$$\sum_{j=i+1}^{t} \frac{R_j}{j^2} \nearrow C_i = \sum_{j=i+1}^{\infty} \frac{R_j}{j^2},$$

with $0 \leq C_i \leq C/(i+1)$ for some constant C. Since $\sup_{j\geq 1} R_j \leq R = c^2/(1-c)^2$, using the ICT again,

$$C_i - \sum_{j=i+1}^{t} \frac{R_j}{j^2} = \sum_{j=t+1}^{\infty} \frac{R_j}{j^2} \leq R \sum_{j=t}^{\infty} \frac{1}{j^2} = O\left(\frac{1}{t}\right).$$

Hence

$$\sum_{j=i+1}^{t} \frac{R_j}{j^2} = C_i + O\left(\frac{1}{t}\right). \tag{B.10}$$

As a consequence, for $c < 1$, bringing together (B.8), (B.9), and (B.10) we may write

$$\ln \phi_{ti} = -c\left[\ln t - \ln(i+1) + E_i\right] - C_i + O\left(\frac{1}{t}\right)$$

for all $i \geq 0$. Hence

$$\ln\left(t^c \phi_{ti}\right) = c \ln(i+1) - cE_i - C_i + O\left(\frac{1}{t}\right)$$

or, denoting $B_i = e^{-(cE_i + C_i)}$,

$$\begin{aligned} t^c \phi_{ti} &= (i+1)^c e^{-(cE_i + C_i)} \exp\left[O\left(\frac{1}{t}\right)\right] \\ &= (i+1)^c B_i \left[1 + O\left(\frac{1}{t}\right)\right] \\ &= i^c \theta_i \left[1 + O\left(\frac{1}{t}\right)\right] \end{aligned} \tag{B.11}$$

with $\theta_i = \left(1 + i^{-1}\right)^c B_i$ for $i \geq 1$ and $\theta_0 = B_0$. This stronger form of (B.7a) will be needed for the case $c < 1/2$. Note that the $O(1/t)$-term also depend on i, but the dependence on $1/t$ is uniform in i: $\sup_{i \leq t} |O_i(1/t)| = O(1/t)$. Note also that

$$\lim_{t\to\infty} B_i = 1. \tag{B.12}$$

We are now ready to derive the basic working formulas for a_t used in the proofs below. They will be different for $c < 1/2$ and for $c \geq 1/2$.

B.1.1. Case $c < 1/2$. Starting from (B.3) and noting that $i_0 = 0$ for $c < 1/2$, we make use of (B.11) to write

$$t^c a_t = t^c a_0 \phi_{t0} + \gamma \sum_{i=1}^{t} t^c \phi_{ti} \frac{\varepsilon_i}{i} = a_0 B_0 + \gamma\,(v_t + w_t) + O\left(t^{-1}\right), \tag{B.13}$$

where we have put

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$v_t = \sum_{i=1}^{t} \theta_i \frac{\varepsilon_i}{i^{1-c}}, \; w_t = \frac{1}{t} \sum_{i=1}^{t} \frac{O_{ti}(1)}{i^{1-c}} \varepsilon_i. \tag{B.14}$$

Note that the $O\left(t^{-1}\right)$ and the $O_{ti}(1)$-terms are deterministic and uniformly bounded.

B.1.2. Case $c \geq 1/2$. In this case, except for $c < 1$, we have $i_0 \geq 1$. Starting from (B.3), we make the decomposition

$$\sum_{i=1}^{t} \phi_{ti} \frac{\varepsilon_i}{i} = \sum_{i=1}^{i_0-1} \phi_{ti} \frac{\varepsilon_i}{i} + \sum_{i=i_0}^{t} \phi_{ti} \frac{\varepsilon_i}{i}.$$

By (B.7b), the first term on the right hand side is $O\left(t^{-c}\right)$. Making use of (B.7a), we write the second term in the form

$$\begin{aligned} \sum_{i=1}^{t} \phi_{ti} \frac{\varepsilon_i}{i} &= \frac{1}{t^c} \sum_{i=i_0}^{t} \frac{\varepsilon_i}{i^{1-c}} + \frac{1}{t^c} \sum_{i=i_0}^{t} \frac{O_{ti}(1)}{i^{2-c}} \varepsilon_i \\ &= \xi_t^0 + \eta_t^0. \end{aligned}$$

Since it is more convenient to start the sums at $i = 1$ instead of $i = i_0$, we introduce

$$v_t = \sum_{i=1}^{t} \frac{\varepsilon_i}{i^{1-c}}, \quad w_t = \sum_{i=1}^{t} \frac{O_{ti}(1)}{i^{2-c}} \varepsilon_i, \tag{B.15a}$$

$$\xi_t = \frac{1}{t^c} v_t, \quad \eta_t = \frac{1}{t^c} w_t \tag{B.15b}$$

(with the $O_{ti}(1)$-term $= 1$ for $i < i_0$). Then $\xi_t^0 = \xi_t + O\left(t^{-c}\right)$, and same for η. Therefore, since $|\phi_{t0}| = O\left(t^{-c}\right)$ (cf. (B.7b)) we may write (B.3) in the form

$$a_t = O\left(t^{-c}\right) + \gamma \left(\xi_t + \eta_t\right). \tag{B.16}$$

Remark B.2. The definitions of v_t and w_t differ for the two cases. We desist, however, from using different symbols for the two processes since in the proofs it will always be clear which case is under investigation.

Remark B.3. At first sight, (B.13) seems to be nothing but a minor improvement in precision in comparison with (B.16). Indeed, it is. But the exact identification of the $O\left(t^{-c}\right)$-term in (B.16) will be crucial in the proof of Theorem 3, where it is indispensable to establish the existence of a limit for $t^c a_t$ instead of a mere $O_p(1)$ statement.

B.2. Proof of Theorem 3

In the following, we will derive the asymptotic behaviour of a_t. In addition, we shall also consider that of $\overline{a}_T$ because it is needed in the treatment of the EEP in Appendix B.3. The three subsections below will cover the cases of $c > 1/2$, $c = 1/2$, and $c < 1/2$ separately.

B.2.1. Case $c > 1/2$. Reconsider the representation of a_t in (B.16) above and examine first the behaviour of ξ_t (cf. (B.15b)). By the ICT, the predictable quadratic variation of v_t is given by

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$\langle v\rangle_t = \sigma^2 \sum_{i=1}^{t} i^{2(c-1)} = \frac{\sigma^2}{2c-1} t^{2c-1} + O(1). \tag{B.17}$$

Hence $\langle v\rangle_\infty = \lim_{t\to\infty}\langle v\rangle_t = \infty$ a.s. By the CLT for sums of independent random variables (cf. Shiryaev (1996, Chapter III, Section 4, Theorem 1)),

$$\frac{v_t}{\sqrt{\langle v\rangle_t}} \sim \frac{\sqrt{2c-1}}{\sigma}\frac{v_t}{t^{c-1/2}} \xrightarrow{d} \mathcal{N}(0,1)$$

or, in terms of ξ_t,

$$\frac{\sqrt{2c-1}}{\sigma}\sqrt{t}\xi_t \xrightarrow{d} \mathcal{N}(0,1). \tag{B.18}$$

Turning to η_t (cf. (B.15), it follows again by the ICT that

$$\mathbf{E}w_t^2 = \begin{cases} O\left(t^{2c-3}\right) & \text{for } c > 3/2, \\ O\left(\ln t\right) & \text{for } c = 3/2, \\ O(1) & \text{for } c < 3/2. \end{cases}$$

Hence $\mathbf{E}t\eta_t^2 = t^{1-2c}\mathbf{E}w_t^2 \to 0$ and, as a consequence,

$$\sqrt{t}\eta_t \xrightarrow{L^2} 0. \tag{B.19}$$

Taking account of (B.16), we find that

$$\frac{\sqrt{2c-1}}{\sigma\gamma}\sqrt{t}a_t \xrightarrow{d} \mathcal{N}(0,1),$$

which settles part (*i*) of Theorem 3. For later use, let us briefly consider the behaviour of $\overline{a}_T$. From the definition,

$$\mathbf{E}\xi_t^2 = \frac{\sigma^2}{t^{2c}}\sum_{i=1}^{t}\frac{1}{i^{2(1-c)}} = O\left(t^{-1}\right).$$

Hence, since $\mathbf{E}t\eta_t^2 = o(1)$, it follows that

$$\mathbf{E}|\overline{a}_T| \le \frac{\text{const}}{T}\sum_{t=1}^{T}\left(t^{-c} + \mathbf{E}|\xi_t| + \mathbf{E}|\eta_t|\right) = O\left(T^{-1/2}\right).$$

Therefore,

$$\overline{a}_T = O_p\left(T^{-1/2}\right). \tag{B.20}$$

Remark B.4. Actually, for $c > 1/2$, it holds that

$$\sum_{t=1}^{\infty}\mathbf{E}\eta_t^2 < \infty,$$

so that

$$\sum_{t=1}^{\infty}\eta_t^2 < \infty \text{ a.s.} \tag{B.21}$$

We will need this result later, cf. Appendix B.3.2.

B.2.2. Case $c = 1/2$. Reconsider (B.16). Regarding ξ_t, note that the predictable quadratic variation of v_t is

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$\langle v\rangle_t = \sigma^2 \sum_{i=1}^{t} i^{-1} = \sigma^2 \ln t + O(1). \tag{B.22}$$

Hence, arguing as above, it follows from the CLT that

$$\frac{v_t}{\sqrt{\langle v\rangle_t}} \sim \frac{v_t}{\sigma\sqrt{\ln t}} \xrightarrow{d} \mathcal{N}(0,1)$$

or, in terms of ξ_t,

$$\frac{1}{\sigma}\sqrt{\frac{t}{\ln t}}\xi_t \xrightarrow{d} \mathcal{N}(0,1).$$

As for η_t, $\mathbf{E}t\eta_t^2 = O(1)$, so that

$$\sqrt{\frac{t}{\ln t}}\eta_t \xrightarrow{p} 0.$$

Therefore, by virtue of (B.16), we obtain that

$$\frac{1}{\sigma\gamma}\sqrt{\frac{t}{\ln t}}a_t = O\left(\frac{1}{\ln t}\right) + \frac{1}{\sigma}\sqrt{\frac{t}{\ln t}}\xi_t + \frac{1}{\sigma}\sqrt{\frac{t}{\ln t}}\eta_t \xrightarrow{d} \mathcal{N}(0,1).$$

For later use, note that

$$\mathbf{E}a_t^2 = \sigma^2\frac{\ln t}{t}. \tag{B.23}$$

Remark B.5. The Lindeberg condition (LC) for the cases $c > 1/2$ and $c = 1/2$ will be verified in Section E.1 of the Online Supplement.

B.2.3. Case $c < 1/2$. In this case, the basic formula is (B.13):

$$t^c a_t = a_0 B_0 + \gamma\,(v_t + w_t) + O\left(t^{-1}\right), \tag{B.24}$$

with

$$v_t = \sum_{i=1}^{t}\theta_i\frac{\varepsilon_i}{i^{1-c}}, \quad w_t = \frac{1}{t}\sum_{i=1}^{t}\frac{O_{ti}(1)}{i^{1-c}}\varepsilon_i \tag{B.25}$$

and $\lim_{i\to\infty}\theta_i = 1$. The $O\left(t^{-1}\right)$ - and $O_{ti}(1)$-terms are deterministic and uniformly bounded. Apparently, w_t converges to zero in L^2. Also, the limit

$$v = \sum_{i=1}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}} \tag{B.26}$$

exists in L^2 and has a continuous distribution function (cf. Remark A.1). Hence it follows from (B.24) that

$$\lim_{t\to\infty} t^c a_t = a_0 B_0 + \gamma v = u \tag{B.27}$$

in L^2. Or, put differently,

$$t^c a_t = u + \rho_t, \tag{B.28}$$

with

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$\rho_t = \gamma\,(v_t - v) + \gamma\, w_t + O\left(t^{-1}\right). \tag{B.29}$$

Note that $\rho_t \to 0$ in L^2. Concerning the behaviour of $\overline{a}_T$, this means that

$$\operatorname*{plim}_{T\to\infty} T^c \overline{a}_T = \frac{1}{1-c} u. \tag{B.30}$$

Remark B.6. Actually, the convergences $w_t \to 0$ and $v_t \to v$ also occur with probability one. For w_t, this follows from the Borel-Cantelli lemma, while for v_t it is a consequence of the MCT for martingales with bounded quadratic variation. This is elaborated in more detail in Section I of the Online Supplement.

B.3. Proof of Theorem 4

B.3.1. The OLS Estimator. Recall from Section 3.2 formulas (3.3)–(3.5) for the OLSE of the structural parameters in the DGP (3.1), written in terms of the original noncentred variables a_t.

A moment's thought shows that the OLS estimators obtained by using a_t or the centred forecasts $a_t^{\#} = a_t - \alpha$ are identical. With obvious notations, $u_T = u_T^{\#}$, $A_T = A_T^{\#}$, $\widehat{\beta}_T = \widehat{\beta}_T^{\#}$. Hence

$$\sqrt{A_T}\left(\widehat{\beta}_T - \beta\right) = \frac{u_T}{\sqrt{A_T}} = \frac{u_T^{\#}}{\sqrt{A_T^{\#}}} = \sqrt{A_T^{\#}}\left(\widehat{\beta}_T^{\#} - \beta\right).$$

Henceforth, we will continue to work with the transformed $a_t^{\#}$ and denote them again by a_t. Our interest is in the asymptotic behaviour of

$$U_T = \sqrt{A_T}\left(\widehat{\beta}_T - \beta\right) = \frac{u_T}{\sqrt{A_T}}. \tag{B.31}$$

Making use of the elementary algebraic identity

$$\sum_{t=1}^{T} (a_t - \overline{a}_T)\left(b_t - \overline{b}_T\right) = \sum_{t=2}^{T} \frac{t-1}{t}\left(a_t - \overline{a}_{t-1}\right)\left(b_t - \overline{b}_{t-1}\right),$$

we may write

$$A_T = \sum_{t=2}^{T} \frac{t-1}{t}\left(a_{t-1} - \overline{a}_{t-1}^{-}\right)^2. \tag{B.32}$$

Remark B.7. (B.32) shows that A_T is nondecreasing.

The decisive step is to determine the rate of divergence of A_T, i.e., find a sequence of numbers $\alpha_T \nearrow \infty$ such that

$$\operatorname*{plim}_{T\to\infty} \frac{A_T}{\alpha_T} = \lambda \tag{B.33}$$

with λ positive and finite. Note that, since A_T is increasing, this implies that $A_\infty = \lim_{T\to\infty} A_T = \infty$ with probability one. For $c > 1/2$, it will turn out that

$$\alpha_T = \ln T, \quad \lambda = \frac{\sigma^2 \gamma^2}{2c-1} \tag{B.34a}$$

will do. In this case, U_T in (B.31) will be shown to be asymptotically normal.

For $c < 1/2$, we will show that

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$\alpha_T = T^{1-2c}, \quad \lambda = \frac{u^2 c^2}{(1-c)^2 (1-2c)}, \tag{B.34b}$$

with u as in (B.27). The fact that now the limit λ is random, incorporating the entire sequence of ε_t, poses some problems with asymptotic normality of U_T. Without further assumptions, it can, however, be shown that the square of U_T is asymptotically χ^2-distributed (up to a normalisation).

B.3.2. Case $c > 1/2$. We proceed in two steps. First, we show asymptotic normality of U_T, assuming (B.33). Then we will prove the latter.

Step 1. From (B.20) we know that

$$\overline{a}_T^- = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{B.35}$$

Hence, starting with the second term in the right-most expression of

$$u_T = \sum_{t=1}^{T} a_{t-1}\varepsilon_t - \overline{a}_T^- \sum_{t=1}^{T} \varepsilon_t = u'_T - u''_T, \tag{B.36}$$

we have

$$\frac{u''_T}{\sqrt{A_T}} = \sqrt{\frac{\alpha_T}{A_T}} \frac{\overline{a}_T^-}{\sqrt{\alpha_T}} \sum_{t=1}^{T} \varepsilon_t = O_p(1) \frac{1}{\sqrt{T \ln T}} \sum_{t=1}^{T} \varepsilon_t = o_p(1) \tag{B.37}$$

by the CLT. Turning to the first term, consider the martingale difference array (MDA)

$$M_T = \frac{u'_T}{\sqrt{\alpha_T}} = \sum_{t=1}^{T} \frac{a_{t-1}}{\sqrt{\alpha_T}} \varepsilon_t. \tag{B.38}$$

Its predictable quadratic variation is

$$\langle M \rangle_T = \sigma^2 \frac{A_T^0}{\alpha_T},$$

where

$$A_T^0 = \sum_{t=1}^{T} a_{t-1}^2. \tag{B.39}$$

From the definition of A_T (cf. (3.5)) together with (B.35) it follows readily that

$$A_T = A_T^0 - T\left(\overline{a}_T^-\right)^2 = A_T^0 + O_p(1). \tag{B.40}$$

Therefore, in view of (B.34a),

$$\frac{A_T}{\alpha_T} = \frac{A_T^0}{\alpha_T} + O_p\left(\frac{1}{\alpha_T}\right) = \frac{A_T^0}{\alpha_T} + o_p(1)$$

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

and

$$\operatorname*{plim}_{T\to\infty}\frac{A_T}{\alpha_T}=\operatorname*{plim}_{T\to\infty}\frac{A_T^0}{\alpha_T}. \tag{B.41}$$

In particular, (B.33) is equivalent to

$$\operatorname*{plim}_{T\to\infty}\frac{A_T^0}{\alpha_T}=\lambda. \tag{B.42}$$

As a consequence, $\operatorname{plim}_{T\to\infty}\langle M\rangle_T=\sigma^2\lambda$. Applying a CLT for MDAs, cf. Shiryaev (1996, Chapter VII, Section 8), we find that

$$M_T\xrightarrow{d}\mathcal{N}(0,\sigma^2\lambda), \tag{B.43}$$

or

$$\frac{u'_T}{\sqrt{A_T}}=\sqrt{\frac{\alpha_T}{A_T}}M_T\xrightarrow{d}\mathcal{N}(0,\sigma^2). \tag{B.44}$$

The proof of the LC for (B.43) is relegated to Section E.2 of the Online Supplement. Bringing together (B.31), (B.36), (B.37), and (B.44), we then obtain as final consequence

$$U_T=\frac{u'_T-u''_T,}{\sqrt{A_T}}=\sqrt{\frac{\alpha_T}{A_T}}M_T-\frac{u''_T}{\sqrt{A_T}}\xrightarrow{d}\mathcal{N}(0,\sigma^2). \tag{B.45}$$

Together with (B.33) and (B.34a) this settles part (*i*) of Theorem 4.

Step 2. It remains to verify (B.42):

$$\operatorname*{plim}_{T\to\infty}\frac{A_T^0}{\ln T}=\frac{\sigma^2\gamma^2}{2c-1}. \tag{B.46}$$

Note that, by monotonicity, this implies that $A_\infty^0=\infty$ a.s. We go back to equation (B.1):

$$a_t=\left(1-\frac{c}{t}\right)a_{t-1}+\frac{\gamma}{t}\varepsilon_t$$

(remember our renaming convention). Taking squares yields

$$a_t^2=\left(1-\frac{c}{t}\right)^2a_{t-1}^2+\frac{\gamma^2}{t^2}\varepsilon_t^2+2\left(1-\frac{c}{t}\right)\frac{\gamma}{t}a_{t-1}\varepsilon_t.$$

Multiplying by t and summing over t we obtain

$$\sum_{t=1}^{T}ta_t^2=\sum_{t=1}^{T}t\left(1-\frac{c}{t}\right)^2a_{t-1}^2+\gamma^2\sum_{t=1}^{T}\frac{1}{t}\varepsilon_t^2+2\gamma\sum_{t=1}^{T}\left(1-\frac{c}{t}\right)a_{t-1}\varepsilon_t. \tag{B.47}$$

Since

$$\left(1-\frac{c}{t}\right)^2=1-\frac{2c}{t}+\frac{c^2}{t^2},$$

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

some straightforward calculations show that we may write the first term on the right hand side of (B.47) as

$$\sum_{t=1}^{T} t\left(1-\frac{c}{t}\right)^2 a_{t-1}^2 = \sum_{t=1}^{T} t a_t^2 - T a_T^2 + (1-2c)\, A_T^0 + c^2 \sum_{t=1}^{T} \frac{1}{t} a_{t-1}^2 .$$

Returning to (B.47), this means that

$$\sum_{t=1}^{T} t a_t^2 = \sum_{t=1}^{T} t a_t^2 - T a_T^2 + (1-2c)\, A_T^0 + c^2 \sum_{t=1}^{T} \frac{1}{t} a_{t-1}^2 \qquad \textbf{(B.48)}$$
$$+ \gamma^2 \sum_{t=1}^{T} \frac{1}{t} \varepsilon_t^2 + 2\gamma \sum_{t=1}^{T} \left(1-\frac{c}{t}\right) a_{t-1}\varepsilon_t .$$

Solving for A_T^0 yields

$$(2c-1)\, A_T^0 = -T a_T^2 + c^2 \sum_{t=1}^{T} \frac{1}{t} a_{t-1}^2 + \gamma^2 \sum_{t=1}^{T} \frac{1}{t} \varepsilon_t^2 \qquad \textbf{(B.49)}$$
$$+ 2\gamma\, N_T - 2\gamma\, c \sum_{t=1}^{T} \frac{1}{t} a_{t-1}\varepsilon_t ,$$

where N_T is the martingale

$$N_T = \sum_{t=1}^{T} a_{t-1}\varepsilon_t .$$

To analyse the individual terms on the right hand side of (B.49), we bring in the asymptotic behaviour of a_t resulting from Theorem 3(i):

$$a_t = O_p\left(t^{-1/2}\right)$$

and its following consequences:

1. $T a_T^2 = O_p(1)$.
2. $\sum_{t=1}^{T} t^{-1} a_{t-1}^2 = O_p(1)$ since $\mathbf{E} a_t^2 = O\left(t^{-1}\right)$, as proved in Appendix B.2.1.
3. $\sum_{t=1}^{T} \frac{1}{t} a_{t-1}\varepsilon_t = O(1)$ a.s. by the martingale convergence theorem, since the predictable quadratic variation of the martingale on the left hand side is uniformly bounded in L^1 and thus converges a.s. to some finite limit.
4. Making the decomposition $\nu_t = \varepsilon_t^2 - \sigma^2$ and applying Kolmogorov's LLN to ν_t, it follows that
$$\sum_{t=1}^{T} \frac{1}{t} \varepsilon_t^2 = \sum_{t=1}^{T} \frac{1}{t} \sigma^2 + \sum_{t=1}^{T} \frac{\nu_t}{t} = \sigma^2 \ln T + O(1) \text{ a.s.}$$

Making use of these properties in (B.49), we obtain

$$(2c-1)\, A_T^0 = \gamma^2 \sigma^2 \ln T + O_p(1) + 2\gamma\, N_T . \qquad \textbf{(B.50)}$$

Noting that $\langle N \rangle_T = \sigma^2 A_T^0$, we then argue as follows. Suppose that $A_\infty^0 < \infty$ on some set Γ of positive measure. Then, by the martingale convergence theorem, N_T converges a.s.

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

on Γ to some finite limit. Dividing (B.50) by A_T^0, we obtain

$$2c-1=\gamma^2\sigma^2\frac{\ln T}{A_T^0}+\frac{O_p(1)}{A_T^0}+O(1)=\gamma^2\sigma^2\frac{\ln T}{A_T^0}\left[1+\frac{O_p(1)}{\ln T}\right]. \tag{B.51}$$

On Γ, the right hand side converges to ∞ in probability (in the sense that, for all M, $\mathbf{P}(\text{rhs} > M, \Gamma) \to \mathbf{P}(\Gamma)$ as $T \to \infty$), which is impossible since the left hand side is finite. As a consequence, $A_\infty^0 = \infty$ with probability one. Again from the martingale convergence theorem (now the version for martingales with unbounded bracket process) it then follows that

$$\frac{N_T}{A_T^0} \to 0 \text{ a.s.}$$

Dividing (B.50) by A_T^0 we again get (B.51), which now tells us that

$$\frac{A_T^0}{\ln T} \xrightarrow{P} \frac{\gamma^2\sigma^2}{2c-1}. \tag{B.52}$$

It is worth emphasising that, by monotonicity, $A_\infty^0 = \infty$ a.s. Note also that it is not sufficient to simply show that $A_T/\ln T = O_p(1)$ since the latter need not imply that A_T tends to infinity. This, however, is indispensable for asymptotic normality.

B.3.3. Case $c = 1/2$. For $c = 1/2$, the calculations for A_T^0 made in Appendix B.3.2 remain valid up to equation (B.49). Pursuing this line of argument any further, it can be shown that

$$\operatorname*{plim}_{T\to\infty}\frac{A_T^0}{\ln T}=\infty. \tag{B.53}$$

Actually, performing some tedious calculations involving fourth moments, it can be shown by a Cauchy sequence argument that there does not exist any *deterministic* sequence α_T s.t. A_T^0/α_T converges in L^2 to some finite positive number λ. Assuming moments of order high enough such that uniform integrability holds, this also precludes convergence in probability. The latter is, however, indispensable if one tries to mimic the proof for $c > 1/2$. It therefore seems that, for $c = 1/2$, the OLS is not asymptotically normal in the sense of Theorem 4*(i)*. Of course this does not exclude the existence of some nonstandard limit distribution. Needless to say that, nevertheless, (B.53) is sufficient for consistency.

B.3.4. Case $c < 1/2$. We continue to work with the formulas (3.4), (3.5), and (B.31). The first term in u_T is the martingale $N_T = \sum_{t=1}^T a_{t-1}\varepsilon_t$ already encountered in Appendix B.3.2. N_T has the predictable quadratic variation $\sigma^2 A_T^0$. Consider

$$U_T^0=\frac{u_T}{\sqrt{A_T^0}}. \tag{B.54}$$

Then

$$U_T=\sqrt{\frac{A_T^0}{A_T}}U_T^0. \tag{B.55}$$

Our procedure here is as follows. In a first step, we show that $A_\infty^0 = \infty$ a.s. and $\rho^2 = \lim_{T\to\infty} A_T^0/A_T$ is finite. In a second step, we show that

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$\left(U_T^0\right)^2 \xrightarrow{d} \sigma^2 \rho^{-2} \chi^2, \tag{B.56}$$

with $\chi^2 = \chi_1^2$. Together, both steps will imply that

$$U_T^2 = A_T \left(\widehat{\beta}_T - \beta\right)^2 \xrightarrow{d} \sigma^2 \chi^2. \tag{B.57}$$

Step 1. The essential tool will be the asymptotic behaviour of A_T^0 and A_T. Note that equation (B.40) is no longer valid for $c < 1/2$. We therefore go back to the properties of a_t investigated in Theorem 3 (*iii*). For convenience, we repeat the fundamental facts here:

$$t^c a_t = u + \rho_t, \tag{B.58a}$$

$$\rho_t = \gamma\,(v_t - v) + \gamma\, w_t + O\left(t^{-1}\right), \tag{B.58b}$$

$$T^c \overline{a}_T = \frac{u}{1-c} + o_p(1), \tag{B.58c}$$

and u has a continuous distribution function (cf. (B.28)–(B.30)). Performing some straightforward calculations it can be shown that

$$A_T^0 = T^{1-2c} \frac{u^2}{1-2c} \left[1 + o_p(1)\right], \quad A_T = T^{1-2c} \nu^2 u^2 \left[1 + o_p(1)\right], \tag{B.59}$$

with

$$\nu^2 = \frac{c^2}{(1-2c)\,(1-c)^2}. \tag{B.60}$$

This settles the assertion, showing that

$$\rho^2 = \operatorname*{plim}_{T\to\infty} \frac{A_T^0}{A_T} = \left(\frac{1}{c} - 1\right)^2 \tag{B.61}$$

on the set $\{u \neq 0\}$. Note that $\mathbf{P}(u \neq 0) = 1$.

Step 2. First we will treat the general case without assuming normally distributed errors. Gaussian error terms will be looked at in Appendix B.4.

We go back to (3.4):

$$u_T = \sum_{t=1}^{T} a_{t-1} \varepsilon_t - \overline{a}_T^- \sum_{t=1}^{T} \varepsilon_t = u_T' - u_T''. \tag{B.62}$$

Using (B.58c) we obtain

$$u_T'' = \overline{a}_T^- \sum_{t=1}^{T} \varepsilon_t = \frac{u}{1-c} T^{-c} \left[1 + o_p(1)\right] \sum_{t=1}^{T} \varepsilon_t.$$

By (B.59),

$$\sqrt{A_T^0} = \frac{|u|}{\sqrt{1-2c}} T^{1/2-c} \left[1 + o_p(1)\right]. \tag{B.63}$$

Therefore,

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$\frac{u''_T}{\sqrt{A^0_T}} = \frac{\sqrt{1-2c}}{1-c}\operatorname{sign}(u)\left[1+o_p(1)\right]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t$$
$$= \frac{\sqrt{1-2c}}{1-c}\operatorname{sign}(u)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t + o_p(1)$$
$$= \frac{\sqrt{1-2c}}{1-c}\operatorname{sign}(u)\,X^0_T + o_p(1). \tag{B.64}$$

Here we have put

$$X^0_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t.$$

Coming to u'_T, we make use of (B.58). Since $(t-1)^{-c} - t^{-c} = O\left(t^{-(1+c)}\right)$ and u as well as ρ_t are bounded in L^2 (since $\mathbf{E}\rho_t^2 \to 0$),

$$a_{t-1} = (t-1)^{-c}\left[u+\rho_{t-1}\right] = t^{-c}\left[u+\rho_{t-1}\right] + t^{-(1+c)}d_t,$$

with d_t bounded in L^2. Hence

$$\sum_{t=1}^{T}a_{t-1}\varepsilon_t = u\sum_{t=1}^{T}t^{-c}\varepsilon_t + \sum_{t=1}^{T}t^{-c}\rho_{t-1}\varepsilon_t + O_p(1)$$

and

$$\frac{u'_T}{\sqrt{A^0_T}} = \frac{\sqrt{1-2c}}{1+o_p(1)}\frac{1}{T^{1/2-c}}\left[\operatorname{sign}(u)\sum_{t=1}^{T}t^{-c}\varepsilon_t + \sum_{t=1}^{T}t^{-c}\rho_{t-1}\varepsilon_t + O_p(1)\right]$$
$$= \frac{\sqrt{1-2c}}{1+o_p(1)}\left[\operatorname{sign}(u)\,X'_T + Z_T\right] + o_p(1). \tag{B.65}$$

Here we have put

$$X'_T = \frac{1}{T^{1/2-c}}\sum_{t=1}^{T}t^{-c}\varepsilon_t, \quad Z_T = \frac{1}{T^{1/2-c}}\sum_{t=1}^{T}t^{-c}\rho_{t-1}\varepsilon_t. \tag{B.66}$$

If X'_T and Z_T are both $O_p(1)$, (B.65) can be written

$$\frac{u'_T}{\sqrt{A^0_T}} = \sqrt{1-2c}\left[\operatorname{sign}(u)\,X'_T + Z_T\right] + o_p(1). \tag{B.67}$$

For X'_T, this is clear since $\mathbf{E}\left(X'_T\right)^2 = O(1)$. For Z_T, it will be shown below that $Z_T \xrightarrow{P} 0$.

Introduce

$$X_T = X'_T - \frac{1}{1-c}X^0_T = \frac{1}{T^{1/2-c}}\sum_{t=1}^{T}\left(t^{-c} - \frac{T^{-c}}{1-c}\right)\varepsilon_t. \tag{B.68}$$

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

We now combine (B.54), (B.64), and (B.67) to obtain the following decomposition for U_T^0:

$$U_T^0 = \frac{u_T}{\sqrt{A_T^0}} = \sqrt{1-2c}\left[\operatorname{sign}(u)\left(X_T' - \frac{1}{1-c}X_T^0\right) + Z_T\right] + o_p(1)$$
$$= \sqrt{1-2c}\left[\operatorname{sign}(u)\, X_T + Z_T\right] + o_p(1). \tag{B.69}$$

As for X_T,

$$\frac{1}{T^{1-2c}}\sum_{t=1}^{T}\left(t^{-c} - \frac{T^{-c}}{1-c}\right)^2 = \frac{1}{T^{1-2c}}\left[\sum_{t=1}^{T} t^{-2c} - 2\frac{T^{-c}}{1-c}\sum_{t=1}^{T} t^{-c} + \frac{T^{1-2c}}{(1-c)^2}\right]$$
$$= \left[\frac{1}{1-2c} - \frac{1}{(1-c)^2}\right] + O\left(\frac{1}{T^{1-2c}}\right)$$
$$= \upsilon^2 + o(1),$$

with υ^2 as in (B.60). Hence it follows from the CLT for sums of independent random variables that

$$X_T \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\upsilon^2\right). \tag{B.70}$$

The LC is verified in Section E.3 of the Online Supplement. Hence, if it can be shown that

$$Z_T \xrightarrow{p} 0, \tag{B.71}$$

it will follow from (B.69) that

$$\left(U_T^0\right)^2 = (1-2c)\, X_T^2 + o_p(1) = (1-2c)\,\sigma^2\upsilon^2\left(\frac{X_T}{\sigma\upsilon}\right)^2 + o_p(1) \xrightarrow{d} \sigma^2\tau^2\chi_1^2, \tag{B.72}$$

with

$$\tau^2 = \upsilon^2(1-2c) = \left(\frac{c}{1-c}\right)^2 = \rho^{-2},$$

cf. (B.61). Or, returning to U_T, taking account of (B.55) and (B.61),

$$U_T^2 \xrightarrow{d} \sigma^2\chi^2.$$

Remark B.8. It should be noted that, without further knowledge, it does not follow from (B.70) that $\operatorname{sign}(u)\, X_T$ converges in distribution, let alone is asymptotically normal. The problem caused here by the *random* nuisance parameter $\operatorname{sign}(u)$ is similar to the one appearing in explosive AR(1)-models $y_t = a y_{t-1} + \varepsilon_t$, where the denominator $\sum_{t=1}^{T} y_{t-1}^2$ of the OLS-estimator tends geometrically fast (at rate $|a|^{2T}$) to some nondegenerate random limit u.

It remains to show that $Z_T \xrightarrow{p} 0$, see (B.71). To that end, return to (B.66), with

$$\rho_{t-1} = \gamma\left(\upsilon_{t-1} - \upsilon\right) + \gamma\, w_{t-1} + O\left(t^{-1}\right)$$

(cf. (B.58b)). For the following, we need a somewhat more detailed representation of the term $\upsilon_{t-1} - \upsilon$. Remembering (B.25) and (B.26), we write

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$\rho_{t-1} = \gamma\,\zeta_t + \gamma\,w_{t-1} + O\left(t^{-1}\right), \tag{B.73}$$

with

$$\zeta_t = \upsilon_{t-1} - \upsilon = \sum_{i=t}^{\infty} \theta_i \frac{\varepsilon_i}{i^{1-c}}. \tag{B.74}$$

Then

$$\begin{aligned} Z_T &= \frac{1}{T^{1/2-c}} \sum_{t=1}^{T} t^{-c} \rho_{t-1} \varepsilon_t \\ &= \frac{\gamma}{T^{1/2-c}} \left[\sum_{t=1}^{T} t^{-c} \zeta_t \varepsilon_t + \sum_{t=1}^{T} t^{-c} w_{t-1} \varepsilon_t + \sum_{t=1}^{T} \frac{O_t(1)}{t^{1+c}} \varepsilon_t \right]. \end{aligned}$$

Apparently (remember that the $O_t(1)$-terms are deterministic), the contribution of the last term on the right hand side is $o_p(1)$ since

$$\frac{1}{T^{1/2-c}} \sum_{t=1}^{T} \frac{\varepsilon_t}{t^{1+c}} \xrightarrow{L^2} 0.$$

Hence

$$Z_T = \gamma\,(R_T + S_T) + o_p(1) \tag{B.75}$$

with

$$R_T = \frac{1}{T^{1/2-c}} \sum_{t=1}^{T} t^{-c} \zeta_t \varepsilon_t, \quad S_T = \frac{1}{T^{1/2-c}} \sum_{t=1}^{T} t^{-c} w_{t-1} \varepsilon_t. \tag{B.76}$$

Some tedious, but rather straightforward calculations show that

$$\operatorname*{plim}_{T\to\infty} R_T = \operatorname*{plim}_{T\to\infty} S_T = 0. \tag{B.77}$$

The details are provided in Section G of the Online Supplement.

B.4. Proof of Corollary 2: Gaussian Errors

In this section, we consider the special case where the ε_t are i.i.d. normal. Then a more specific statement can be made about the asymptotics of the *deterministically normalised* OLS-estimator, i.e., about the asymptotic distribution of

$$V_T = \sqrt{T^{1-2c}}\left(\widehat{\beta}_T - \beta\right).$$

Using the notation introduced in Appendix B.3, taking account of (B.59) and (B.61),

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$V_T = \sqrt{\frac{T^{1-2c}}{A_T}}\frac{u_T}{\sqrt{A_T}} = \sqrt{\frac{T^{1-2c}}{A_T}}\sqrt{\frac{A_T^0}{A_T}}\frac{u_T}{\sqrt{A_T^0}}$$
$$= \frac{\rho}{\nu\,|u|}U_T^0 + o_p(1).$$

In view of the representation (B.69),

$$V_T = \frac{\rho}{\nu\,|u|}\sqrt{1-2c}\left[\operatorname{sign}(u)\,X_T + Z_T\right] + o_p(1)$$
$$= \frac{\rho\sqrt{1-2c}}{\nu}\frac{X_T}{u} + o_p(1). \tag{B.78}$$

The crucial fact is that, under normality,

$$\begin{pmatrix} u \\ X_T \end{pmatrix} \xrightarrow{d} \mathcal{N}\left\{\begin{pmatrix} a_0 B_0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_X^2 \end{pmatrix}\right\}, \tag{B.79}$$

with $\sigma_u^2 = \sigma^2\kappa^2, \sigma_X^2 = \sigma^2\nu^2$ and

$$\kappa^2 = \gamma^2\sum_{i=1}^{\infty}\frac{\theta_i^2}{i^{2(1-c)}}.$$

(B.79) follows from the fact that

$$u = a_0 B_0 + \gamma\,\upsilon, \quad \text{with } \upsilon = \sum_{i=1}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}$$

(cf. (B.27) and (B.26)), and since X_T and υ are jointly normal with covariance

$$\begin{aligned}\mathbf{Cov}(\upsilon, X_T) &= \frac{1}{T^{1/2-c}}\mathbf{E}\sum_{i=1}^{\infty}\frac{\theta_i}{i^{1-c}}\varepsilon_i\sum_{t=1}^{T}\left(t^{-c} - \frac{T^{-c}}{1-c}\right)\varepsilon_t \\ &= \frac{\sigma^2}{T^{1/2-c}}\sum_{t=1}^{T}\frac{\theta_t}{t^{1-c}}\left(t^{-c} - \frac{T^{-c}}{1-c}\right) \\ &= \frac{\sigma^2}{T^{1/2-c}}\left[\sum_{t=1}^{T}\frac{\theta_t}{t} - \frac{T^{-c}}{1-c}\sum_{t=1}^{T}\frac{\theta_t}{t^{1-c}}\right] \\ &= O(1)\frac{\sigma^2}{T^{1/2-c}}\left[\ln T + O(1)\right] \\ &= O\left(T^{c-1/2}\right).\end{aligned}$$

In particular, u and X_T are asymptotically independent. It then follows from the continuous mapping theorem that

$$\frac{X_T}{u} \xrightarrow{d} \frac{X}{u},$$

where X is an $\mathcal{N}\left(0, \sigma_X^2\right)$-distributed random variable independent of $u \sim \mathcal{N}\left(a_0 B_0, \sigma_u^2\right)$. Normalising, we find that

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press

$$V_T' = \frac{\sigma_u}{\sigma_x}\frac{X_T}{u} = \frac{\kappa}{\upsilon}\frac{X_T}{u} \xrightarrow{d} \frac{\widetilde{X}}{\widetilde{u}}, \tag{B.80}$$

where $\widetilde{X} \sim \mathcal{N}(0,1)$ is independent of $\widetilde{u} \sim \mathcal{N}(a_0 B_0, 1)$. As is well known, for $\mu = a_0 B_0 = 0$, the ratio $\widetilde{X}/\widetilde{u}$ is standard Cauchy-distributed. For nonvanishing μ, the limit distribution is a noncentral Cauchy distribution $\mathcal{C} = \mathcal{C}(\mu)$ which may also be regarded as double noncentral t-distribution with one degree of freedom whose noncentrality parameters are zero in the numerator and μ in the denominator. Its density is given by

$$f(\upsilon) = \frac{1}{\pi}\frac{1}{1+\upsilon^2}e^{-\mu^2/2}\left[1 + \frac{\mu}{2}\psi(\mu)\right],$$
$$\psi(\mu) = \frac{2}{1+\upsilon^2}e^{\mu^2/2(1+\upsilon^2)}\int_0^{\mu} e^{-x^2/2(1+\upsilon^2)}dx.$$

An alternative representation of the density in terms of an infinite series expansion may be found in Krishnan (1968). Going back to (B.80), we then find that

$$V_T' \xrightarrow{d} \mathcal{C}.$$

Finally, by asymptotic equivalence of V and V',

$$\kappa\upsilon V_T \xrightarrow{d} \mathcal{C}(\mu).$$

Remark B.9. Reconsider the result in Theorem 2(*iii*). It turns out that for Gaussian error terms, u and υ are independent normally distributed random variables with mean zero. Hence the limit distribution of $\widehat{\beta}$ is also a noncentral Cauchy distribution. For explosive AR(1)-model without intercept the usual Cauchy distribution $\mathcal{C}(0)$ was already identified as limit distribution in the Gaussian case by White (1958) and Shiryaev and Spokoiny (1997).

https://doi.org/10.1017/S0266466616000529 Published online by Cambridge University Press