Estimating Structural Parameters in Regression Models with Adaptive Learning

Norbert Christopeit[1]
Michael Massmann[2]

[1] *University of Bonn, Germany;*
[2] *Faculty of Economics and Business Administration, VU University Amsterdam, the Netherlands.*

Estimating structural parameters in regression models with adaptive learning[a]

Norbert Christopeit[b]
University of Bonn

Michael Massmann[b,c]
WHU – Otto Beisheim School of Management
and
VU University Amsterdam

5th August, 2015, 15:01 hrs

Abstract

This paper examines the ordinary least squares (OLS) estimator of the structural parameters in a class of stylised macroeconomic models in which agents are boundedly rational and use an adaptive learning rule to form expectations of the endogenous variable. The popularity of this type of model has recently increased amongst applied economists and policy makers who seek to estimate it empirically. Two prominent learning algorithms are considered, namely constant gain and decreasing gain learning. For each of the two learning rules, the 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, taken the first stage learning dynamics as given. In the case of constant gain learning, the structural model effectively amounts to a static, 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 non-standard.

keywords: adaptive learning, non-stationary regression, ordinary least squares, consistency, asymptotic distribution.

JEL codes: C22, C51, D83

Contents

1 Introduction

The purpose of this paper is to investigate the properties of the ordinary least squares estimator of the structural parameters in a prototypical macroeconomic model with adaptive learning. In particular, we consider a model class that has received substantial attention in the economic theory literature:

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

where $y_{t|t-1}^e$ 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 & Savin (1986) and Fourgeaud, Gourieroux & Pradel (1986), as does the Lucas (1973) aggregate supply model and the New Keynesian Phillips curve (NKPC), cf. Roberts (1995) and Clarida, Galí & 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 is, however, that the slope parameter δ generally turns out to be unrestricted while the autoregressive coefficient β lies within the unit interval, as in the NKPC, or to be negative, as in the cobweb model. For the purposes of this paper, we will refer 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 $y_{t|t-1}^e$ are modelled. The traditional approach is via rational expectations, cf. Muth (1961) or Sargent (2008), which assumes that agents, when forming expectations $y_{t|t-1}^e$, have complete knowledge of the model and the past $\mathcal{F}_{t-1} = \sigma\left(y_s, s \leq t-1; x_s, s \leq t\right)$ and make best use of it, i.e. set $y_{t|t-1}^e = \mathbf{E}\left(y_t | \mathcal{F}_{t-1}\right)$. Taking conditional expectations in (1.1) yields $\mathbf{E}\left(y_t | \mathcal{F}_{t-1}\right) = \alpha x_t$ 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 & Honkapohja (2001), Hommes (2002) or Gaspar, Smets & 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 $y_{t|t-1}^e$. One way to specify this auxiliary model is to assume that its functional form corresponds to that of the REE in (1.3). Generally, agents are presumed not to know the parameter α and therefore replace it by some estimate a_{t-1}, based on information $\mathcal{F}_{t-1}$. Typically, the parameter α will be estimated by some recursive procedure which, in general, has the form of a stochastic

approximation algorithm:

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

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

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. For recent surveys of stochastic approximation algorithms and their applications, see Lai (2003) and Kushner (2010). With the learning scheme in (1.4), agents' expectation will be given by

$$y_{t|t-1}^e = a_{t-1} x_t,$$

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 $y_{t|t-1}^e$ 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.

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 & Mavroeidis (2010), the European Central Bank's New Multi-Country Model by Dieppe, González Pandiella, Hall & Willman (2011), the inflation model by Malmedier & Nagel (2013) and the model of stock market volatility by Adam, Marcet & Nicolini (2015). 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 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 & 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."*

---

[1] The issue of *strong* consistency of the OLS estimator is investigated in a companion paper, see Christopeit & Massmann (2012).

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 *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; a good account of seminal results can be found in Benveniste, Métivier & 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. This result presumes that $\beta < 1$ since it can be shown that, if $\beta \geq 1$, a_t diverges, see e.g. Christopeit & 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 & Williams (2013) and Brock & 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 & Sargent (1995) and for agents employing a Bayesian estimation approach by Vives (1993). Since, however, the asymptotic behaviour of a_t turns out to be crucial for the properties of the OLS estimator in the EEP 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. (2015). In this case, the regressor a_t is an autoregressive 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 & Wei (1985) for the stationary ergodic case, Chan & Wei (1988) for the (negative) unit root case, and Phillips & Magdalinos (2008) as well as Wang & 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:

$$y_t \sim \beta\alpha x_t + \delta x_t + \varepsilon_t, \tag{1.6}$$

i.e. the asymptotic moment matrix

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

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 & Rosenblatt (1957). In the econometrics literature the singularity of M is generally referred to as absence of strong asymptotic identification, see e.g. Davidson & MacKinnon (1993) or Newey & 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 $\widehat{\delta}$ in (1.5) possess desirable asymptotic properties. One leading example in which the asymptotic singularity of M

does not preclude the consistency 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. Thus, seminal results such as the necessary condition for weak consistency of OLS in the simple linear regression model with a deterministic explanatory variable, as derived by Lai & Robbins (1977), are not applicable in our setting. Corresponding results for predetermined regressors can be found in, for instance, Christopeit & Helmes (1980), Lai & Wei (1982a, 1982b) and Christopeit (1986). However, as will be discussed in Section 4, they do not satisfy our needs either.

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 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 & Nicolini (2003) combine both types of learning in a switching model of hyperinflation, while Evans & 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 the very simplest scenario in which $x_t = x$ is a constant. In this case, $r_t = x^2$ 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 (1.4a) reduces to

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

while the structural equation in (1.5) becomes

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

7

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

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

where we have defined

$$c_t = (1 - \beta)\, \gamma_t. \qquad (1.12)$$

Note that under constant gain learning, c_t is equal to

$$c = (1 - \beta)\gamma, \qquad (1.13)$$

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

The DGP 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. While a detailed analysis of these issues is left to future research the likely ramifications of both extensions will be looked at in Section 4.

The first central conclusion of our analysis is that the asymptotics of a_t and of $\widehat{\beta}$ as well as $\widehat{\delta}$ critically depend on the value of c in (1.13). In particular, regarding the interplay of IFP and EEP, it is clear 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 central conclusion is that in all of the aforementioned settings the structural parameters β and δ can be consistently estimated although in all but one scenario the OLS estimators will have differing, highly non-standard asymptotic distributions. This result hence provides an answer to Marcet & Sargent's speculation quoted above, in that the learning systems considered in this paper do yield useful, albeit complex, econometric models.

The main contribution of the paper to the literature is hence that it derives the asymptotics of the OLS estimator $\widehat{\delta}$ and $\widehat{\beta}$ in the setting of model (1.1) under both constant and decreasing gain learning. This has not been attempted before. The paper thus provides 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 provides 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 and only comparable to the results derived in special cases by Marcet & Sargent (1995) and Vives (1993). It will be of potential interest in the economic theory literature. Finally, econometricians will likely find the third contribution instructive, namely that the paper effectively provides a synopsis of models that yield vastly different dynamics depending on the choice of their parameter spaces, covering unit root and explosive behaviour as much as illustrating issues like non-standard convergence rates and singular variance-covariance matrices.

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 model extensions are discussed in Section 4 before Section 5 concludes. Proofs of the constant gain results are relegated to Appendix A and those of the decreasing gain results to Appendix B. The proofs of some auxiliary results are contained in Appendices C-E.

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, for $\beta \neq 1$, the recursion in (1.11) is now written as

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

since $c\alpha = \gamma\delta$. As before,

$$c = (1 - \beta)\gamma. \tag{2.3}$$

In the special case of $\beta = 1$, (2.2a) is replaced by

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

Indeed, (2.2b) can be considered as resulting from taking the limit $\beta \to 1$ in (2.2a). Note that it is common in the economics literature to assume that $\gamma \in (0, 1)$ and that $\beta < 1$. Yet, 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 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 EEP 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 (2.2a) satisfies $|\rho| < 1$. Therefore we are in the classical scenario of stable autoregressive processes.

(ii) If $c = 0$ then $\rho = 1$ and (2.2a) is a random walk with drift. As opposed to that, if $c = 2$ then $\rho = -1$ and (2.2b) 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 & 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.

9

(iii) When $c < 0$ or $c > 2$, we have $|\rho| > 1$, and (2.2a) 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.*

(ii) *If $c = 0$ then a_t is a random walk with drift $\delta\gamma$ and*

$$a_t = \gamma\delta t + o(t) \quad a.s..$$

If, instead, $c = 2$ then a_t is an alternating random walk with drift 2α and

$$\frac{1}{\sigma\gamma\sqrt{t}}a_t \overset{d}{\to} \mathcal{N}(0,1). \tag{2.4}$$

(iii) *If $c < 0$ or $c > 2$ then $(1-c)^{-t}a_t$ converges with probability one and in L^2 to a nondegenerate limit with 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 for the time being. 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 discussed in Section 4. 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 derived 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

$$\widehat{\theta}_T - \theta = M_T^{-1}u_T, \tag{2.5}$$

where the second moment matrix M_T is given by

$$M_T = \begin{pmatrix} T & \sum_{t=1}^{T} a_{t-1} \\ \sum_{t=1}^{T} a_{t-1} & \sum_{t=1}^{T} a_{t-1}^2 \end{pmatrix} \tag{2.6}$$

10

and

$$u_T = \begin{pmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} a_{t-1}\varepsilon_t \end{pmatrix}, \tag{2.7}$$

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:

$$a_t^* = \delta^* + \beta^* a_{t-1}^* + \gamma\varepsilon_t, \tag{2.8}$$

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

$$\widehat{\theta}_T^* - \theta^* = M_T^{*-1} u_T^*, \tag{2.9}$$

where M_T^* is defined analogously to (2.6) and

$$u_T^* = \gamma \begin{pmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} a_{t-1}^*\varepsilon_t \end{pmatrix}.$$

If $\beta \neq 1$, suppose that $\delta^* = c\alpha$ 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 (2.5) and (2.9) that

$$\widehat{\theta}_T - \theta = \gamma^{-1} \left(\widehat{\theta}_T^* - \theta^* \right). \tag{2.10}$$

This identity will continue to hold for $\beta = 1$, in which case the two pairs of coefficients are given by $\delta^* = \gamma\delta$ and $\beta^* = 1$. In particular, if $\delta = 0$, the asymptotic behaviour of $\widehat{\theta}_T$ needs to be evaluated at $\delta^* = 0$ and $\beta^* = 1$, corresponding to the Dickey-Fuller model with drift.

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 (2.8). This equivalence will be exploited in the proof of Theorem 2, see Appendix A.3.

For the explosive case of constant gain learning, we need the following additional assumption.

Assumption CG
The limit

$$z = \sum_{t=1}^{\infty} (1 - c)^{-t} \varepsilon_t$$

possesses a continuous distribution function.

Note that the limit exists both in L^2 and with probability one since $|\rho| = |1 - c| > 1$. Sufficient conditions for Assumption CG to hold will be discussed at the end of this section.

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 $\sqrt{T}$:

$$\sqrt{T}(\widehat{\theta}_T - \theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \overline{M}^{-1}\right), \qquad (2.11)$$

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

$$\Omega = \begin{pmatrix} 4 & -\frac{6}{\delta\gamma} \\ -\frac{6}{\delta\gamma} & \frac{12}{\delta^2\gamma^2} \end{pmatrix}. \qquad (2.12)$$

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

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

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

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, under the additional Assumption CG,

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

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

where u and v are independent L^2-variables, and the two estimators are asymptotically independent.

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 & Yu (2015), see also Remark 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), or indeed (2.8), and its proof builds on the results derived by Chan & 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 crucial aspect of Theorem 2 is that the behaviour of the OLS estimator differs markedly depending 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 super-consistent. Moreover, inference on both parameters is feasible in all scenarios, provided one takes account of the non-standard 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.

A note on Assumption CG seems warranted. Though it seems impossible to give a complete characterisation of those distribution functions F of ε_t for which the assumption holds, useful sufficient conditions are available. They are based on a theorem by Lévy, cf. Kawata (1972, Theorem 13.1.2), which, in turn, relies on the observation that the point spectrum of the sum of two independent random variables is the vector sum of the individual point spectra. A trivial consequence of the theorem is that, if the distribution function F of ε_t is continuous, then so is that of z. Hence Assumption CG is satisfied. A consequence of the nontrivial part of the theorem is that, if F has compact support and discontinuities outside some neighborhood of 0, then Assumption CG is also satisfied. It is definitely not satisfied for discontinuous F with unbounded support and for distribution functions having a saltus at 0.

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:

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

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

where, again,

$$c = (1 - \beta)\gamma. \tag{3.3}$$

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 the maintained Assumptions MA1 and MA2 mentioned in Section 2. The following two additional assumptions about the error term ε_t are made at various stages in the subsequent analysis.

Assumption DG1
The ε_t possess finite fourth moments m_4.

Assumption DG2
For $c < 1/2$ and for every sequence of numbers $\theta_i \to 1$, the limit

$$v = \sum_{i=1}^{\infty} \theta_i \frac{\varepsilon_i}{i^{1-c}},$$

has a continuous distribution function.

The limit v exists both in L^2 and with probability one. Comments apply similar to those on Assumption CG made at the end of Section 2.

We will now first discuss the internal forecasting problem (IFP) in the context of model (3.1)-(3.3) 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 & 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. To see this, consider for simplicity a simple linear model with a deterministic regressor: it was shown by Lai & Robbins (1977) that the condition that the sum of squared mean-adjusted regressors diverges is necessary and sufficient for the (strong) consistency of the OLS estimator. Although the structural equation in (3.1) is a linear regression model with a predetermined stochastic regressor, it is intuitive that a comparable condition will

again have to hold. The convergence rates of a_t are instrumental in the analysis of the EEP. They are of course also of interest per se.

The IFP in the context of the model in (3.1)-(3.3) is strikingly non-standard. 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 that 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 & 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 thus tends to a constant. As a result, the process a_t is highly non-stationary, as reflected in, for instance, the fact that its variance decreases to zero at a rate which increases with c. Finally, for fixed t and $h \to \infty$, the covariance $\mathbf{Cov}\,(a_t, a_{t+h})$ behaves as $O\left(h^{-c}\right)$ and the correlation

$$\mathbf{Corr}\,(a_t, a_{t+h}) = \begin{cases} O\left(h^{1/2-c}\right) & \text{if } c > 1/2, \\ O\left((\ln h)^{-1/2}\right) & \text{if } c = 1/2 \end{cases}$$

while, if $c < 1/2$, the correlation tends to a non-zero constant. In the context of stationary stochastic processes, the behaviour for $c > 1/2$ corresponds to what is called long memory, see Brockwell & Davis (1991). It is interesting to note that memory for $c = 1/2$ is even longer while, for $c < 1/2$, it is infinite.

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 $\sqrt{t}$:

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

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

  in L^2, where $u = (a_0 - \alpha)\,B_0 + \gamma v$, B_0 is some positive constant depending on c, and v is as defined in Assumption DG2.

The convergence rates of a_t are illustrated in Figure 1. It is plain that, as c decreases, the convergence of a_t to α gets progressively slower, from $t^{1/2}$ to $(t/\ln(t))^{1/2}$ and to t^c with $c < 1/2$. Moreover, 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

---

[2]We are grateful to George Evans for drawing our attention to this interpretation.
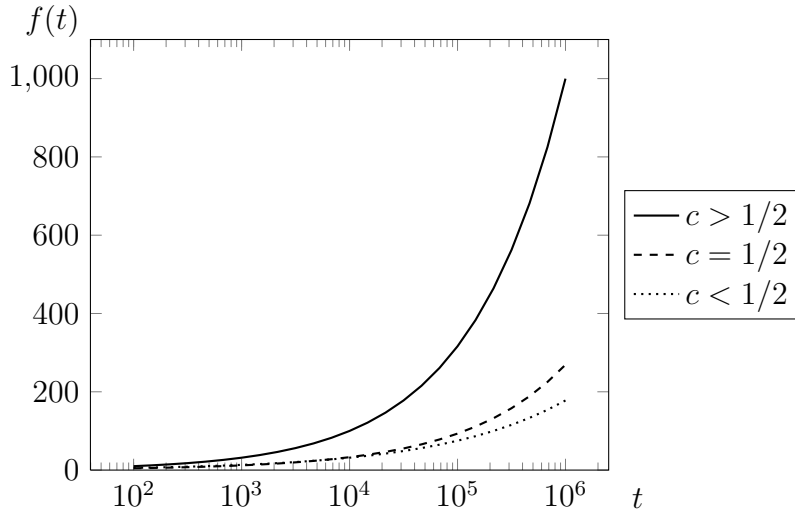
Figure 1: The convergence rates of a_t to normality. Solid line: $f(t) = \sqrt{t}$, dashed line: $f(t) = \sqrt{t/\ln t}$, dotted line: $f(t) = t^{3/8}$.

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 et al. (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 & Sargent (1995) and Evans & 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.

Theorem 3 states that, for $c < 1/2$, the asymptotic distribution of a_t is, in general, not normal. The reason for that is that v, as defined in Assumption DG2, is a random nuisance parameter. Yet, making the explicit assumption that the error terms are Gaussian, the ensuing corollary provides an extension of Theorem 3, simply by virtue of v being a weighted sum of the ε_t. Note that, in this case, Assumption DG2 is automatically satisfied.

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.

16

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

$$\widehat{\beta}_T - \beta = \frac{u_T}{A_T}, \tag{3.4}$$

with

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

$$A_T = \sum_{t=1}^{T} \left(a_{t-1} - \bar{a}_T^- \right)^2. \tag{3.6}$$

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. The reason for this is that neither necessary nor minimally sufficient conditions for the weak consistency of the OLS estimator in a linear regression model with predetermined stochastic regressors and i.i.d. errors are available in the literature. And the sufficient conditions that do exist turn out not to be met in our model for some parameter constellations. In particular, the *weak* consistency of OLS is not evident since, as argued in the introduction, one of the classical Grenander conditions is violated in model (3.1)-(3.3), namely the non-singularity of the limiting regressor sample second moment matrix. Incidentally, the best result for the *strong* consistency of OLS so far obtained is not satisfied in our model either if $c > 1/2$, as will be discussed further in Section 4.

It is hence plain that in examining the asymptotic properties of the OLS estimator in (3.4) no recourse can be taken to existing results. We hence 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 will deal with the cases $c > 1/2$ and $c < 1/2$. The boundary case $c = 1/2$ seems to require an approach entirely different from ours and is thus left to future research. A comment on the difficulties arising in that derivation will be made in Remark 7 as part of the proof of Theorem 4 in Appendix B.3.

Theorem 4
 (i) If $c > 1/2$ then, under the additional Assumption DG1,

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

The divergence rate of A_T is given by

$$\operatorname*{plim}_{T \to \infty} \frac{A_T}{\ln T} = \frac{\sigma^2 \gamma^2}{2c - 1}.$$
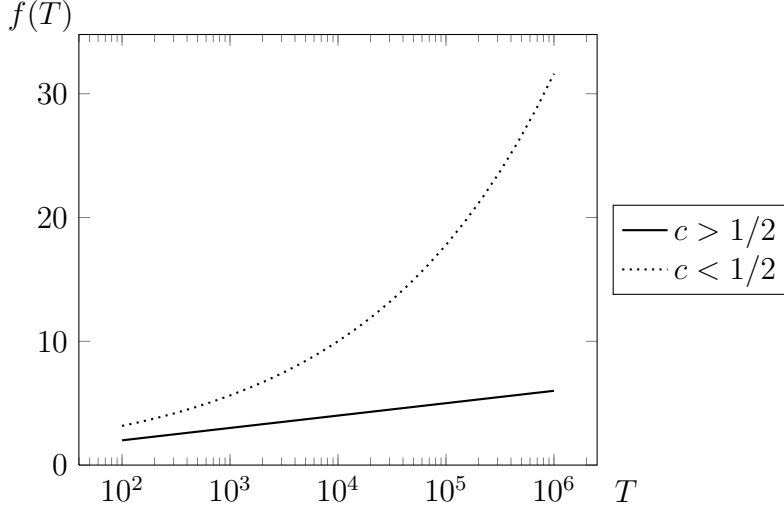
17

Figure 2: The divergence rates of A_T. Solid line: $f(T) = \ln T$, dotted line: $f(T) = T^{1-3/4}$.

(ii) If $c < 1/2$ then, under the additional Assumption DG2,

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

The divergence rate of A_T is given by

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

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

The divergence rates of A_T are illustrated in Figure 2. 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 $\widehat{\beta}_T$.

Note that A_T is a random normalising sequence whose presence can be interpreted as follows: Consider a textbook simple linear regression model with i.i.d.$(0, \sigma^2)$ error terms and exogenous regressors $x_1, \ldots, x_T$. It is well-known that the variance of the OLS estimator of the regression slope, conditional on $\mathcal{X}_T = \sigma(x_1, \ldots, x_T)$, can be written as

$$\mathbf{Var}(\widehat{\beta}_T \mid \mathcal{X}_T) = \frac{\sigma^2}{A_T}$$

with A_T denoting again the sum of the squared mean-adjusted regressors. The usual t-statistic is then

$$t = \frac{\sqrt{A_T}(\widehat{\beta}_T - \beta)}{\widehat{\sigma}}$$

where $\widehat{\sigma}^2$ is a consistent estimator of the unknown variance σ^2. Returning to the model of interest in (3.1)-(3.3), it is clear that the regressors are predetermined and not exogenous, yet part *(i)* of Theorem 4 effectively states that, when $c > 1/2$, it is the analogue of the conventional t-statistic that has an asymptotic standard normal distribution while

18

part *(ii)* shows that, when $c < 1/2$, the analogue of the usual Wald statistic is indeed asymptotically χ^2-distributed.

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, viz. v as defined in Assumption DG2. This is similar to the random nuisance parameter appearing in the asymptotic distribution theory of explosive AR(1)-models,

$$z_t = \phi z_{t-1} + \varepsilon_t, \qquad \text{with } |\phi| > 1, \tag{3.7}$$

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 & 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,

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

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

Indeed, this limit distribution is also obtained for the OLS estimator of ϕ in (3.7) when the normalisation is with the deterministic sequence $|a|^{-2T}$, cf. White (1958), and Shiryaev & Spokoiny (1997).

As mentioned earlier, the asymptotic properties of the estimators of the intercept δ and of the error variance σ^2 are easily derived from those of the slope estimator. In particular, consistency of the OLS estimator

$$\widehat{\delta}_T = \bar{y} - \widehat{\beta}_T \bar{a}^-$$

follows immediately from

$$\widehat{\delta}_T - \delta = (\beta - \widehat{\beta}_T)\bar{a}^- + \bar{\varepsilon}. \tag{3.8}$$

Since it is shown in the proof of Theorem 3 that $\operatorname{plim} a^- = \alpha$ for $c \gtrless 1/2$ and since $\bar{\varepsilon}$ converges to 0 by virtue of the LLN, $\widehat{\delta}_T$ is also weakly consistent. As to asymptotic normality, since $A_T = O_p(\ln T)$ for $c > 1/2$, it is clear that

$$\sqrt{A_T}\bar{\varepsilon} = o_p\left(1\right)$$

and

$$\sqrt{A_T}(\widehat{\delta}_T - \delta) = \sqrt{A_T}(\beta - \widehat{\beta}_T)\bar{a}^- + o_p\left(1\right). \tag{3.9}$$

As a consequence, joint normality of both OLS estimators follows for $c > 1/2$, i.e.

$$\sqrt{A_T}\left(\begin{array}{c}\widehat{\delta}_T - \delta \\ \widehat{\beta}_T - \beta\end{array}\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\left(\begin{array}{cc}\alpha^2 & -\alpha \\ -\alpha & 1\end{array}\right)\right).$$

Note that the asymptotic distribution is degenerate, namely concentrated on the line $y = -x/\alpha$. This implies that individual t-statistics can be constructed in the standard fashion. The linear combination $\widehat{\delta}_T + \alpha\widehat{\beta}_T$ also has a non-degenerate variance and may be subjected to a t-test, too. However, multivariate statistics involving the inverse of the asymptotic variance-covariance matrix, such as the Wald statistic, are infeasible. It would seem natural, however, to bootstrap a statistic like that or to use a generalised inverse to construct it.

For the case $c < 1/2$ with normally distributed errors, an analogue of equation (3.9) continues to hold with normalisation $\sqrt{T^{1-2c}}$ instead of $\sqrt{A_T}$. Consequently, the resulting bivariate distribution is again degenerate.

4 Discussion and extensions

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_T^2 is also consistent. The proof of the following corollary is provided in Appendix E.

Corollary 3
Define the OLS residual $\widehat{\varepsilon}_t = y_t - \widehat{\delta}_T - \widehat{\beta}_T a_{t-1}$. Then

$$\widehat{\sigma}_T^2 = \frac{1}{T} \sum_{t=1}^T \widehat{\varepsilon}_t^2 \xrightarrow{p} \sigma^2.$$

Regarding the gain parameter γ under constant or decreasing gain learning, see (2.2)-(2.3) and (3.2)-(3.3) 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

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

or

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

depending on which type of learning is stipulated. As a consequence, the OLS estimates $\widehat{\beta}_T$, $\widehat{\delta}_T$ and $\widehat{\sigma}_T^2$, 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 Theorem 2 and in Theorem 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 & Evans (2006, 2010) or Markiewicz & 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 non-linear techniques. For instance, Chevillon et al. (2010) use non-linear least squares while Adam et al. (2015) employ the method of simulated moments. The 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 $\mathcal{I}_t = \sigma\left(y_s, s \leq t; x_s, s \leq t\right)$, and assume that x_t follows an i.i.d. process with mean μ_x. This implies that $\mathbf{E}\left(y_{t+1} \mid \mathcal{I}_t\right) = \alpha \mu_x$ 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 & 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, instable 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 simply 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. However, Bullard & Mitra (2002), for instance, criticise this practice and argue that the use of $\mathcal{I}_{t-1}$ would be 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.

A road map for the econometric analysis of models with forward-looking expectations can now be sketched. First, it would seem natural in a first step to examine the properties of the parameter estimators in systems with adaptive learning whose equilibria are E-stable. Secondly, following the approach suggested by Adam (2003), models could be examined that have agents make good use of information $\mathcal{I}_t$ while their forecast $y^e_{t+1|t}$ is not endogenous. OLS estimation of the model parameters in that setting is still appropriate and the consistency and distributional results derived in the present paper should continue to hold. However, if the regressor $y^e_{t+1|t}$ in (4.3) *is* endogenous then OLS would no longer be appropriate and recourse would have to be taken to an instrumental variable technique. Suitable assumptions on instruments and their relationship to the endogenous variable and the error term would have to be made. The analysis would become more involved, yet preliminary derivations indicate that similar issues to those discussed in the present paper would arise.

4.3 Time-varying regressors

The assumption in this paper that the exogenous regressor x_t is constant serves the purpose of analytical tractability, especially in order to facilitate the examination of the asymptotic behaviour of a_t, at least to such an extent as is needed for the treatment of the EEP. Upon inspecting the proofs, however, it is apparent that time-varying regressors x_t may lead to a complication of the analysis, for instance if their behaviour leads to sequences a_t which can no longer be classified as stable, unit root or explosive for its entire trajectory. Of the many ways in which time-varying regressors could be introduced we now discuss a simple extension that has the beauty of resulting in essentially the same results as in Theorems 1-4. Although this generalisation is somewhat special it illustrates that the results in this paper have wider applicability.

Assumption E
The sequence x_t tends to an equilibrium value x: $\lim_{t \to \infty} x_t = x$.

Without loss of generality, we may again assume that $x = 1$. The x_t are taken to be deterministic for expositional simplicity. Identical calculations to those below would result for stochastic regressors if, for instance, *(i)* the regressors are strictly exogenous, i.e. the sequence x_t is independent of the error terms ε_t, and *(ii)* Assumption E holds with probability one.

4.3.1 Constant gain

Reconsider the recursion of r_t in (1.4b) with a constant gain $\gamma_t = \gamma$:

$$r_t = (1 - \gamma) \, r_{t-1} + \gamma x_t^2. \tag{4.4}$$

With the solution of (4.4) given by

$$r_t = \rho^t r_0 + \gamma \sum_{n=0}^{t-1} \rho^n x_{t-n}^2$$

it follows that r_t tends to the equilibrium value $r = x^2 = 1$, provided that $\gamma \in (0, 1)$. Substituting this into the dynamics of a_t in (1.4a) yields the recursion in (2.2), as indeed was obtained under the assumption of a constant $x_t = x$. As a consequence, the asymptotics of a_t are also the same.

Regarding the EEP, note that the structural equation is given by

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

or

$$y_t^x = \delta_t + \beta a_{t-1} + \varepsilon_t^x,$$

with $y_t^x = y_t / x_t$ and $\varepsilon_t^x = \varepsilon_t / x_t$. Remembering that $x_t \to 1$, it can be shown that passing from ε_t^x to ε_t does not affect the behaviour of the OLS estimator, so that all the results for the EEP in the case of constant gain remain valid.

4.3.2 Decreasing gain

Reconsider the recursion of r_t in (1.4b) with a decreasing gain sequence $\gamma_t = \gamma/t$, i.e.

$$r_t = \left(1 - \frac{\gamma}{t}\right) r_{t-1} + \frac{\gamma}{t} x_t^2.$$

This is of the same form as equation (B.3). Hence, performing the same analysis on r_t as is done on a_t in Appendices B.1.1 and B.1.2 shows that, for every $\gamma > 0$,

$$\lim_{t \to \infty} r_t = r\gamma x^2 = r\gamma$$

for some positive number r. Using this equilibrium value in the dynamics for a_t in (1.4a) we obtain

$$a_t = a_{t-1} + \frac{1/r}{t} \left(y_t - a_{t-1}\right),$$

which is just (1.9) with $\gamma_t = \widetilde{\gamma}/t$ and $\widetilde{\gamma} = 1/r$. Note, however, that in order to determine $\widetilde{\gamma}$ and, correspondingly, the value of $\widetilde{c} = (1 - \widetilde{\gamma})\beta$, one has to know r. This, however, is given by

$$r = \lim_{t \to \infty} \frac{1}{t^\gamma} \sum_{i=1}^{t} \frac{\theta_i}{i^{1-\gamma}},$$

cf. (B.17) and (B.20). Since $\theta_i \to 1$, it is clear that

$$r = \lim_{t \to \infty} \frac{1}{t^\gamma} \sum_{i=1}^{t} \frac{1}{i^{1-\gamma}} = \lim_{t \to \infty} \frac{1}{t^\gamma} \left[\int_1^t \frac{ds}{s^{1-\gamma}} + O(1)\right] = \frac{1}{\gamma}.$$

Hence, $\widetilde{\gamma} = \gamma$ and $\widetilde{c} = c$. As a consequence, up to a change in variance, we have the same asymptotics for a_t as for $x_t = 1$. The same is true for the EEP.

4.4 Consistency

As mentioned above, the weak consistency of the OLS estimator in Sections 2 and 3 is obtained as a byproduct of our results in Theorems 2 and 4. It is instructive, however, to look at our results in the light of the results available in the literature on consistency in models with predetermined regressors. The reason is that even the best of those conditions turn out not to be met by some of the constant and decreasing gain learning models we consider in this paper. This finding complements the failure of the Grenander condition for the decreasing gain model in Section 3, see also the discussion in the introduction.

To our knowledge, the best sufficient condition for the consistency of the OLS estimator in multivariate models with predetermined regressors is given in Lai & Wei (1982a). It requires that

$$\lambda_{\min}(T) \to \infty \quad \text{and} \quad \frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} \to 0 \quad \text{a.s.,} \tag{4.5}$$

where $\lambda_{\max}(T)$ and $\lambda_{\min}(T)$ are the maximal and minimal eigenvalue, respectively, of the regressors' moment matrix M_T. For the estimation of the slope parameter in a simple regression model, a slight improvement is given in Lai & Wei (1982b) with the condition

$$\frac{A_T}{\ln T} \to \infty \quad \text{a.s.,} \tag{4.6}$$

with A_T being again the sum of squares mean-adjusted regressors. To illustrate the strength of (4.5), Lai & Wei (1982a) discuss an example in which a marginal violation of the conditions leads to the inconsistency of the OLS estimator. They hence call the conditions in (4.5) "in some sense the weakest possible" (p. 155).

For the purpose of comparing (4.5) and (4.6) to our results on weak consistency, note that condition (4.6) may also be used in terms of convergence in probability, in the sense that

$$\frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} \xrightarrow{p} 0 \qquad (4.7)$$

implies the weak consistency of the OLS estimator, say $\widehat{\theta}_T$. This is because the basic result obtained by Lai & Wei (1982a) is that

$$\left\|\widehat{\theta}_T - \theta\right\|^2 = \frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} O(1) \quad \text{a.s.}$$

on the set $\{\lambda_{\min}(T) > 0\}$. Let us briefly discuss condition (4.7) for the various models considered in this paper.

4.4.1 Constant gain

Reconsider the model in (2.1)-(2.3). For the stable case, (4.7) is trivially satisfied since all entries of M_T in (2.6) satisfy a weak LLN. The same is true for the unit root case, as can be shown by some straightforward calculations on the eigenvalues, using the asymptotic behaviour of the properly normalised entries of M_T as obtained in Appendix B.3.2. For the explosive case, similar calculations making use of Theorem 1 *(iii)* show that

$$\frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} \to 4 \ln |1 - c| \quad \text{a.s..}$$

Hence (4.7) is violated, but weak consistency still holds.

4.4.2 Decreasing gain

Turn now to model (3.1)-(3.3). For $c < 1/2$, it can be verified that condition (4.7) is met. For $c > 1/2$, however, it is shown in Appendix B.3 that

$$\plim_{T \to \infty} \frac{A_T}{\ln T} = \frac{\sigma^2 \gamma^2}{2c - 1}.$$

Hence (4.6) is not satisfied. Also, Christopeit & Massmann (2013b) conclude that

$$\plim_{T \to \infty} \frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} = \left(\alpha^2 + 1\right) \frac{2c - 1}{\sigma^2 \gamma^2}$$

so that (4.7) is not satisfied either. Nevertheless, Theorem 4 implies that the slope estimator is weakly consistent. Under the more stringent conditions of Gaussian error terms, we even have strong consistency, as is shown in Christopeit & Massmann (2012).

5   Conclusion

This paper investigates the asymptotic properties of the OLS estimator of δ and β in a stereotypical macroeconomic model of the form

$$y_t = \beta y_{t|t-1}^e + \delta x_t + \varepsilon_t, \tag{5.1}$$

where agents form expectations $y_{t|t-1}^e = a_{t-1} x_t$ according to constant or decreasing gain adaptive learning. While empirical models of this type are becoming ever more popular in the literature, the asymptotics of OLS have so far not been investigated. The paper is the first to address this issue systematically.

To make the analysis tractable, the regressor sequence x_t is assumed to be constant. For both types of learning, we then provide a complete analysis of the asymptotics of agents' expectations a_t, 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. Subsequently, given the behaviour of a_t, the asymptotic properties of the OLS estimators $\widehat{\delta}$ and $\widehat{\beta}$ are derived for both constant and decreasing gain learning. This is both a novel and a challenging undertaking. Some of the models we consider are highly complex, e.g. self-referential models whose autoregressive parameter follows a non-stationary stochastic process. Yet the conclusions are wide-reaching for the theoretical and applied literature alike: It turns out that, contrary to what might be believed at first sight, $\widehat{\delta}$ and $\widehat{\beta}$ are consistent in all scenarios we consider. Their asymptotic distribution, however, is highly non-standard in all but one setting, so care must be taken when inference is to be conducted. For constant gain models, some results are related to the cointegration or co-explosiveness literature. With decreasing gain, the results are reminiscent of the literature on slowly varying regressions, yet not with deterministic regressors as in Phillips (2007) but with predetermined stochastic regressors. Several of the results we derive do not seem yet to have counterparts in the existing literature.

As extensions, forward-looking expectations and time-varying regressor sequences are contemplated, yet a complete analysis is beyond the scope of the present paper and is left to future research. Our results indicate, however, that an econometric analysis of such models as (5.1) can, at a fundamental level, be econometrically feasible and sound. This is encouraging, given the interest among economic theorists to develop further models of bounded rationality via adaptive learning and given the push by policy makers and applied economists to exploit these methods empirically.

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 (2.2a) 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 (2.2a):

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

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.

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 (2.2b),

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

Hence a_t is a random walk with drift:

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

where $S_t = \sum_{i=1}^{t} \varepsilon_i$. As a consequence, by the LLN,

$$t^{-1} a_t = \gamma\delta + o\left(1\right) \quad \text{a.s..}$$

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

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

Since the $\widetilde{\varepsilon}_j = (-1)^{-j}\varepsilon_j$ are i.i.d., (2.4) 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

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

Since its predictable quadratic variation is

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

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

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

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

$$(1 - c)^{-t} a_t^{\#} = a_0^{\#} + \gamma M_t \to a_0^{\#} + \gamma M$$

and hence

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

both a.s. and in L^2.

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 (2.8).

A.3.1 Stable case: $0 < c < 2$

In this case, $|\beta^*| < 1$. It is standard textbook knowledge that

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

with

$$\overline{M} = \plim_{T \to \infty} \frac{1}{T} M_T$$

and M_T being given in (2.6). Since

$$\plim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} a_t = \mathbf{E} a_t = \alpha$$

$$\plim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} a_t^2 = \mathbf{E} a_t^2 = \sigma_a^2 + \alpha^2,$$

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

A.3.2 Unit root case: $c = 0$ or $c = 2$

Case $c = 0$. Consider first the sub-case of $\delta \neq 0$. Making use of the OLS estimator in (2.5), 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 (2.6) and introducing the normalising matrix

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

we find that

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

Applying the CLT for bivariate martingales to (2.7), one obtains

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

Therefore,

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

Calculating the inverse yields (2.12).

Consider now the second sub-case of $\delta = 0$. 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

$$\widehat{\theta} - \theta = \gamma^{-1}\left(\widehat{\theta}^* - \theta^*\right), \tag{2.10}$$

where $\widehat{\theta}^* = (\widehat{\delta}^*, \widehat{\beta}^*)'$ is the OLS estimator for the parameters $\theta^* = (\delta^*, \beta^*)'$ in the AR(1) model (2.8) 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

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

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

Taking account of (2.9) yields the desired result.

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

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

with DGP

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

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

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

The asymptotics of z_t have, in a multivariate context, been analysed by Chan & Wei (1988). Introducing $\widetilde{\varepsilon}_j = (-1)^{-j}\varepsilon_j$ and

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

it is easily verified that

$$z_t = \gamma(-1)^t \widetilde{S}_t. \tag{A.7}$$

Putting $\widetilde{X}_T(u) = \frac{1}{\sigma\sqrt{T}}\widetilde{S}_{[uT]}$, we have the functional CLT

$$\widetilde{X}_T \Rightarrow \widetilde{B} \tag{A.8}$$

in $D[0,1]$, where $\widetilde{B}$ is a standard BM. As a first consequence,

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=1}^{T} z_{t-1}^2 &= \frac{\gamma^2}{T^2}\sum_{t=1}^{T} \widetilde{S}_{t-1}^2 = \gamma^2\sigma^2 \int_0^1 \widetilde{X}_T^2(s-)ds \\
&\xrightarrow{d} \gamma^2\sigma^2 \int_0^1 \widetilde{B}(s)^2 ds.
\end{aligned} \tag{A.9}$$

Returning to a_t, this means that

$$\frac{1}{T^2}\sum_{t=1}^{T} a_{t-1}^2 = \frac{1}{T^2}\sum_{t=1}^{T} z_{t-1}^2 + 2\alpha\frac{1}{T^2}\sum_{t=1}^{T} z_{t-1} + \frac{\alpha^2}{T}. \tag{A.10}$$

For the OLS estimator in (A.5), we need the asymptotics of the first two sample moments of a_t. For the mean, we must take account of the alternating factor $(-1)^t$ in (A.7) since the mean of $\widetilde{S}_t$ itself is weakly convergent at rate $T^{3/2}$. Since

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

consider

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

By partial summation,

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

Here we have put

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

Since $T^{-1/2} \sum_{t=1}^{T} \sigma_t \widetilde{\varepsilon}_t \xrightarrow{d} \mathcal{N}(0, \sigma^2/2)$ and $T^{-1/2} \widetilde{S}_T \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ (jointly), it follows that $W_T = O_p\left(T^{1/2}\right)$. Hence

$$\plim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} z_{t-1} = 0$$

and

$$\plim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} a_{t-1} = \alpha.$$

Revisiting (A.10) and taking account of (A.9), we find that

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

As a consequence, introducing the normalising matrix

$$D_T = \left(\begin{array}{cc} \sqrt{T} & 0 \\ 0 & T \end{array} \right),$$

we find that

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

Turning to

$$u_T = \left(\begin{array}{c} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} a_{t-1} \varepsilon_t \end{array} \right),$$

define $S_t = \sum_{j=1}^{t} \varepsilon_j$ and $X_T(u) = \frac{1}{\sigma \sqrt{T}} S_{[uT]}$. Then we have the usual functional CLT

$$X_T \Rightarrow B, \tag{A.12}$$

with a standard BM B. As a consequence,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \xrightarrow{d} \sigma B(1). \tag{A.13}$$

31

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

$$\sum_{t=1}^{T} z_{t-1}\varepsilon_t = \gamma \sum_{t=1}^{T} (-1)^{t-1} \widetilde{S}_{t-1}\varepsilon_t = -\gamma \sum_{t=1}^{T} \widetilde{S}_{t-1}\widetilde{\varepsilon}_t. \tag{A.14}$$

From

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

it follows that

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

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

Therefore, in view of (A.8),

$$\frac{1}{T}\sum_{t=1}^{T} \widetilde{S}_{t-1}\widetilde{\varepsilon}_t \xrightarrow{d} \frac{\sigma^2}{2}\left[\widetilde{B}^2(1) - 1 \right]. \tag{A.15}$$

Now look at

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

$$= -\gamma\frac{1}{T}\sum_{t=1}^{T} \widetilde{S}_{t-1}\widetilde{\varepsilon}_t + o(1).$$

In view of (A.14) and (A.15),

$$\frac{1}{T}\sum_{t=1}^{T} a_{t-1}\varepsilon_t \xrightarrow{d} -\gamma\frac{\sigma^2}{2}\left[\widetilde{B}^2(1) - 1 \right]. \tag{A.16a}$$

Or, if this version is preferred,

$$\frac{1}{T}\sum_{t=1}^{T} a_{t-1}\varepsilon_t \xrightarrow{d} -\gamma\sigma^2 \int_0^1 \widetilde{B}(s)d\widetilde{B}(s). \tag{A.16b}$$

(A.13) and (A.16) give the *separate* asymptotic behaviour of the entries of u_T. Actually, what we need in order to figure out the asymptotics of the OLS estimators

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

is the *joint* asymptotic behaviour of all entries of M_T and u_T. This is where the result of Chan & Wei is used, namely that weak convergence in (A.12) and (A.8) is *joint*, i.e.

$$(X_T, \widetilde{X}_T) \Rightarrow (B, \widetilde{B})$$

32

in $D\left[0,1\right]^{2}$, and $B, \widetilde{B}$ are *independent* BMs, cf. Chan & Wei (1988, Theorem 2.2). This allows us to conclude that

$$D_T^{-1} u_T = \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \\ \frac{1}{T} \sum_{t=1}^{T} a_{t-1} \varepsilon_t \end{pmatrix} \xrightarrow{d} \overline{u} = \begin{pmatrix} \sigma B(1) \\ -\gamma \sigma^2 \int_0^1 \widetilde{B}(s) d\widetilde{B}(s) \end{pmatrix}$$

jointly with the convergence in (A.11). Hence, finally,

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

with

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

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 & Yu (2015), the asymptotic distribution of the OLS estimator $\widehat{\theta}^*$ of $\delta^* = c\alpha$ and $\beta^* = 1 - c$ in (2.8) is given by

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

$$(1 - c)^T \left(\widehat{\beta}^* - \beta^* \right) \xrightarrow{d} \frac{\left[(1 - c)^2 - 1 \right] u}{v + (1 - c) c\alpha / \left[(1 - c) - 1 \right]}, \tag{A.17}$$

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

Remark 1
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. This is where Assumption CG enters the stage to guarantee that the right-hand side in (A.17) is well-defined.

B Proofs for decreasing gain

B.1 Preliminaries

We return now to the model (3.1)-(3.3):

$$
\begin{aligned}
y_t &= \delta + \beta a_{t-1} + \varepsilon_t & \text{(B.1a)} \\
a_t &= \left(1 - \frac{c}{t}\right) a_{t-1} + \frac{\gamma}{t}\left(\delta + \varepsilon_t\right), & \text{(B.1b)}
\end{aligned}
$$

where

$$
c = (1 - \beta)\gamma.
$$

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 center a_t about its asymptotic value

$$
\alpha = \frac{\delta}{1 - \beta} = \frac{\gamma\delta}{c}.
$$

To this end, introduce

$$
a_t^\# = a_t - \alpha. \tag{B.2}
$$

Then, by straightforward calculation,

$$
a_t^\# = \left(1 - \frac{c}{t}\right) a_{t-1}^\# + \frac{\gamma}{t}\varepsilon_t. \tag{B.3}
$$

Also,

$$
\begin{aligned}
y_t &= \delta + \beta\left(a_{t-1} - \alpha\right) + \alpha\beta + \varepsilon_t \\
&= \alpha + \beta a_{t-1}^\# + \varepsilon_t \tag{B.4}
\end{aligned}
$$

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.3) is given by

$$
a_t = a_0\phi_{t0} + \gamma\sum_{i=1}^{t}\phi_{ti}\frac{\varepsilon_i}{i}, \tag{B.5}
$$

where

$$
\begin{aligned}
\phi_{ti} &= \left(1 - \frac{c}{i+1}\right)\cdots\left(1 - \frac{c}{t}\right), \quad i = 0,\ldots,t-1, \tag{B.6} \\
\phi_{tt} &= 1.
\end{aligned}
$$

In particular, for $c = 1$,

$$
\phi_{ti} = \frac{i}{t}
$$

for all i. Define $i_0 = [c] =$ smallest integer $\geq c$. Then, for $i \geq i_0$, the factors in (B.6) are all positive. Note that $i_0 = 0$ for $c < 1$. Taking logarithms and using a first order Taylor expansion, we obtain for $i \geq i_0$

$$\ln \phi_{ti} = \sum_{j=i+1}^{t} \ln\left(1 - \frac{c}{j}\right) = -c \sum_{j=i+1}^{t} \frac{1}{j} - \sum_{j=i+1}^{t} \frac{R_j}{j^2}, \tag{B.7}$$

with

$$R_j = \frac{(1-\vartheta_j)\,c^2}{\left(1 - \vartheta_j\frac{c}{j}\right)^2}$$

for some $0 < \vartheta_j < 1$ (Cauchy form of the remainder). Note that $R_j \geq 0$ and

$$\sup_{j \geq i_0+1} R_j \leq \frac{c^2}{\left(1 - \frac{c}{i_0+1}\right)^2}. \tag{B.8}$$

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),

$$
\begin{aligned}
\ln \phi_{ti} &= -c\left(\ln t - \ln(i+1)\right) + O(1)\frac{1}{i+1} \\
&= -c\left(\ln t - \ln i - \ln\frac{i+1}{i}\right) + O(1)\frac{1}{i} \\
&= -c\left(\ln t - \ln i\right) + O(1)\frac{1}{i}.
\end{aligned}
$$

As a consequence, for $i \geq i_0 > 0$,

$$\phi_{ti} = \left(\frac{i}{t}\right)^c \left[1 + O\left(\frac{1}{i}\right)\right]. \tag{B.9a}$$

For $0 \leq i < i_0$, it follows from (B.6) and (B.9a) that

$$
\begin{aligned}
|\phi_{ti}| &= \prod_{j=i+1}^{i_0-1}\left|1 - \frac{c}{j}\right| \prod_{j=i_0}^{t}\left(1 - \frac{c}{j+1}\right) \\
&\leq C\phi_{ti_0} = O\left(t^{-c}\right).
\end{aligned} \tag{B.9b}
$$

In particular,

$$|\phi_{t0}| = O\left(t^{-c}\right). \tag{B.10}$$

Actually, for $c < 1$ (i.e. $i_0 = 0$), (B.9a) 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.7):

$$\ln \phi_{ti} = -c \sum_{j=i+1}^{t} \frac{1}{j} - \sum_{j=i+1}^{t} \frac{R_j}{j^2}. \tag{B.11}$$

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

Proposition (ICT) *Let $f(x)$ positive decreasing fct 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) $0 \leq d_{ni} - d_i \leq f(n)$.

(iv) $s_{ni} \leq f(i) + t_{ni}$.

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 (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

$$\sum_{j=i+1}^{t} \frac{1}{j} = \ln t - \ln (i+1) + E_i + O_i \left(\frac{1}{t} \right). \tag{B.12}$$

As to the second term on the right hand side of (B.11),

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

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

$$\sum_{j=i+1}^{t} \frac{R_j}{j^2} = C_i + O \left(\frac{1}{t} \right). \tag{B.13}$$

As a consequence, for $c < 1$, bringing together (B.11), (B.12) and (B.13) we may write

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

36

for all $i \geq 0$. Hence

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

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

$$
\begin{aligned}
t^c \phi_{ti} &= (i+1)^c\, e^{-(cE_i + C_i)}\exp\left[O\left(\frac{1}{t}\right)\right] \\
&= (i+1)^c\, B_i\left[1 + O\left(\frac{1}{t}\right)\right] \\
&= i^c\theta_i\left[1 + O\left(\frac{1}{t}\right)\right]
\end{aligned}
\tag{B.14}
$$

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.9a) will be needed for the case $c < 1/2$. Note that the $O\left(1/t\right)$-term also depend on i, but the dependence on $1/t$ is uniform in i: $\sup_{i \leq t}\left|O_i\left(1/t\right)\right| = O\left(1/t\right)$. Note also that

$$\lim_{t \to \infty} B_i = 1. \tag{B.15}$$

In particular, note that

$$B_0 = e^{-(cE_0 + C_0)},$$

where E_0 is the Euler constant and

$$C_0 = \sum_{j=1}^{\infty} \frac{R_j}{j^2}.$$

Making use of (B.7), we find that C_0 may be calculated from

$$C_0 = -\lim_{t \to \infty}\left[c\sum_{j=1}^{t}\frac{1}{j} + \ln\phi_{t0}\right] = -\lim_{t \to \infty}\sum_{j=1}^{t}\left[\frac{c}{j} + \ln\left(1 - \frac{c}{j}\right)\right].$$

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.5) and noting that $i_0 = 0$ for $c < 1/2$, we make use of (B.14) to write

$$
\begin{aligned}
t^c a_t &= t^c a_0 \phi_{t0} + \gamma\sum_{i=1}^{t} t^c \phi_{ti}\frac{\varepsilon_i}{i} \\
&= a_0 B_0 + \gamma\left(v_t + w_t\right) + O\left(t^{-1}\right),
\end{aligned}
\tag{B.16}
$$

where we have put

$$v_t = \sum_{i=1}^{t}\theta_i\frac{\varepsilon_i}{i^{1-c}}, \quad w_t = \frac{1}{t}\sum_{i=1}^{t}\frac{O_{ti}(1)}{i^{1-c}}\varepsilon_i. \tag{B.17}$$

37

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

$$a_t = t^{-c}a_0 B_0 + \gamma\left(\xi_t + \eta_t\right) + O\left(\frac{1}{t^{1+c}}\right), \tag{B.18}$$

with

$$\xi_t = \frac{1}{t^c}v_t, \quad \eta_t = \frac{1}{t^c}w_t. \tag{B.19}$$

B.1.2 Case $c \geq 1/2$

In this case, except for $c < 1$, we have $i_0 \geq 1$. Starting from (B.5), 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.9b), the first term on the right hand side is $O\left(t^{-c}\right)$. Making use of (B.9a), 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

$$v_t = \sum_{i=1}^{t}\frac{\varepsilon_i}{i^{1-c}}, \quad w_t = \sum_{i=1}^{t}\frac{O_{ti}(1)}{i^{2-c}}\varepsilon_i, \tag{B.20a}$$

$$\xi_t = \frac{1}{t^c}v_t, \quad \eta_t = \frac{1}{t^c}w_t \tag{B.20b}$$

(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, finally, taking account of (B.10), we may write (B.5) in the form

$$a_t = O\left(t^{-c}\right) + \gamma\left(\xi_t + \eta_t\right). \tag{B.21}$$

Remark 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 3
At first sight, (B.16) seems to be nothing but a minor improvement in precision in comparison with (B.21). Indeed, it is. But the exact identification of the $O\left(t^{-c}\right)$-term in (B.21) 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 $\bar{a}_T$ because it is needed in the treatment of the EEP in Appendix B.3. The four 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.21) above and examine first the behaviour of ξ_t (cf. (B.20b)). By the ICT, the predictable quadratic variation of v_t is given by

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

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, Shiryaev (1996, Chapter III, §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,

$$\frac{\sqrt{2c-1}}{\sigma} \sqrt{t} \xi_t \xrightarrow{d} \mathcal{N}(0,1). \tag{B.23}$$

Turning to η_t (cf. (B.20), it follows again by the ICT that

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

Hence

$$\mathbf{E} t \eta_t^2 = t^{1-2c} \mathbf{E} w_t^2 \to 0,$$

and, as a consequence,

$$\sqrt{t} \eta_t \xrightarrow{L^2} 0. \tag{B.24}$$

Taking account of (B.21), we find that

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

For later use, let us briefly consider the behavior of $\bar{a}_T$. 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

$$\begin{aligned} \mathbf{E} \left| \bar{a}_T \right| & \leq \frac{\text{const}}{T} \sum_{t=1}^{T} \left(t^{-c} + \mathbf{E} \left| \xi_t \right| + \mathbf{E} \left| \eta_t \right| \right) \\ & \leq \frac{\text{const}}{T} \sum_{t=1}^{T} \left(t^{-c} + t^{-1/2} \right) = \frac{\text{const}}{T} \sum_{t=1}^{T} t^{-1/2} \\ & = O\left(T^{-1/2}\right). \end{aligned}$$

Therefore,

$$\bar{a}_T = O_p\left(T^{-1/2}\right). \tag{B.25}$$

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

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

so that

$$\sum_{t=1}^{\infty} \eta_t^2 < \infty \quad a.s.. \tag{B.26}$$

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

B.2.2   Case $c = 1/2$

Reconsider (B.21). Regarding ξ_t, note that the predictable quadratic variation of v_t is

$$\langle v \rangle_t = \sigma^2 \sum_{i=1}^{t} i^{-1} = \sigma^2 \ln t + O(1). \tag{B.27}$$

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

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

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,

$$\mathbf{E}t\eta_t^2 = O\left(1\right),$$

so that

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

Therefore, by virtue of (B.21), we obtain that

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

Remark 5
The Lindeberg condition (LC) for the cases $c > 1/2$ and $c = 1/2$ will be verified in Appendix C.1

B.2.3   Case $c < 1/2$

In this case, the basic formula is (B.16):

$$t^c a_t = a_0 B_0 + \gamma \left(v_t + w_t\right) + O\left(t^{-1}\right), \tag{B.28}$$

with

$$v_t = \sum_{i=1}^{t} \theta_i \frac{\varepsilon_i}{i^{1-c}}, \quad w_t = \frac{1}{t} \sum_{i=1}^{t} \frac{O_{ti}\left(1\right)}{i^{1-c}} \varepsilon_i \tag{B.29}$$

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

$$v = \sum_{i=1}^{\infty} \theta_i \frac{\varepsilon_i}{i^{1-c}} \tag{B.30}$$

exists in L^2 and has nonzero variance. Hence it follows from (B.28) that

$$\lim_{t \to \infty} t^c a_t = a_0 B_0 + \gamma v = u \tag{B.31}$$

in L^2. Or, put differently,

$$t^c a_t = u + \rho_t, \tag{B.32}$$

with

$$\rho_t = \gamma \left(v_t - v\right) + \gamma w_t + O\left(t^{-1}\right). \tag{B.33}$$

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

$$\mathrm{plim}_{T \to \infty} T^c \overline{a}_T = \frac{1}{1-c} u. \tag{B.34}$$

Actually, it can be shown that

$$\mathbf{E}\left| T^c \overline{a}_T - \frac{1}{1-c} u \right|^2 = O\left(T^{2c-1}\right), \tag{B.35}$$

but we will not need this sharper result.

B.3   Proof of Theorem 4

B.3.1   The OLS estimator

Recall from Section 3 that the OLS estimator for the structural parameters in the DGP (3.1) in terms of the original non-centered variables a_{t-1} is given by

$$\widehat{\beta}_T - \beta = \frac{u_T}{A_T}, \tag{B.35}$$

where

$$u_T = \sum_{t=1}^{T} \left(a_{t-1} - \overline{a}_T^-\right) \varepsilon_t \tag{B.36}$$

and

$$A_T = \sum_{t=1}^{T} \left(a_{t-1} - \overline{a_T^-}\right)^2. \tag{B.37}$$

Here we have used the notation

$$\overline{a_T^-} = \frac{1}{T} \sum_{t=1}^{T} a_{t-1} \tag{B.38}$$

which will turn out convenient in the sequel.

A moment's thought shows that the OLS estimators obtained by using a_t or the centered forecasts $a_t^{\#} = a_t - \alpha$ are identical. To see this, let $\widehat{\beta}_T^{\#}$ denote the OLS estimator obtained using the latter, i.e.

$$\widehat{\beta}_T^{\#} - \beta = \frac{u_T^{\#}}{A_T^{\#}},$$

with obvious definitions for $u_T^{\#}$ and $A_T^{\#}$. Then

$$a_{t-1} - \overline{a_T^-} = a_{t-1}^{\#} - \overline{a_T^{\#-}},$$

and hence

$$u_T = u_T^{\#}, \quad A_T = A_T^{\#}.$$

As a consequence,

$$\widehat{\beta}_T - \beta = \widehat{\beta}_T^{\#} - \beta$$

and

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

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

$$U_T = \sqrt{A_T} \left(\widehat{\beta}_T - \beta\right) = \frac{u_T}{\sqrt{A_T}}. \tag{B.39}$$

Making use of the elementary algebraic identity

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

we may write

$$A_T = \sum_{t=2}^{T} \frac{t-1}{t} \left(a_{t-1} - \overline{a_{t-1}^-}\right)^2. \tag{B.40}$$

Remark 6
(B.40) 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

$$\text{plim}_{T \to \infty} \frac{A_T}{\alpha_T} = \lambda \tag{B.41}$$

42

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

$$\alpha_T = \ln T, \quad \lambda = \frac{\sigma^2 \gamma^2}{2c - 1} \tag{B.42a}$$

will do. In this case, U_T in (B.39) will be shown to be asymptotically normal.

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

$$\alpha_T = T^{1-2c}, \quad \lambda = \frac{u^2 c^2}{(1 - c)^2 (1 - 2c)}, \tag{B.42b}$$

with u as in (B.31). 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 $c > 1/2$

Our procedure will be as follows. First, we show asymptotic normality of U_T, assuming (B.41). Then we will prove the latter. From (B.25) we know that

$$\overline{a_T} = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{B.43}$$

Hence, starting with the second term on the right hand side of (B.36),

$$\begin{aligned}
\frac{\overline{a_T}}{\sqrt{A_T}} \sum_{t=1}^{T} \varepsilon_t &= \sqrt{\frac{\alpha_T}{A_T}} \frac{\overline{a_T}}{\sqrt{\alpha_T}} \sum_{t=1}^{T} \varepsilon_t \\
&= O_p(1) \frac{1}{\sqrt{\ln T}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \\
&= o_p(1)
\end{aligned} \tag{B.44}$$

by the CLT. Turning to the first term in (B.36), consider the martingale difference array (MDA)

$$M_T = \sum_{t=1}^{T} \frac{a_{t-1}}{\sqrt{\alpha_T}} \varepsilon_t. \tag{B.45}$$

Its predictable quadratic variation is

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

where

$$A_T' = \sum_{t=1}^{T} a_{t-1}^2. \tag{B.46}$$

From the definition of A_T (cf. (B.37)) together with (B.43) it follows readily that

$$A_T = A_T' - T\left(\overline{a_T}\right)^2 = A_T' + O_p(1). \tag{B.47}$$

Therefore, in view of (B.41),

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

and

$$\text{plim}_{T\to\infty}\frac{A_T'}{A_T} = 1. \tag{B.48}$$

In particular, (B.41) is equivalent to

$$\text{plim}_{T\to\infty}\frac{A_T'}{\alpha_T} = \lambda. \tag{B.49}$$

As a consequence,

$$\text{plim}_{T\to\infty}\langle M\rangle_T = \sigma^2\lambda.$$

Applying a CLT for MDAs, cf. Shiryaev (1996, Chapter VII, §8), we find that

$$M_T \xrightarrow{d} \mathcal{N}(0,\sigma^2\lambda), \tag{B.50}$$

or

$$\sqrt{\frac{\alpha_T}{A_T}}M_T \xrightarrow{d} \mathcal{N}(0,\sigma^2) \tag{B.51}$$

The proof of the LC for (B.50) is relegated to Appendix C.2. In view of (B.36), (B.39) and (B.45), bringing together (B.44) and (B.51), we then obtain as final consequence

$$U_T = \sqrt{\frac{\alpha_T}{A_T}}M_T + o_p(1)$$
$$\xrightarrow{d} \mathcal{N}(0,\sigma^2). \tag{B.52}$$

Or, in terms of the explicit rate of convergence,

$$\frac{\gamma}{\sqrt{2c-1}}\sqrt{\ln T}\left(\widehat{\beta}_T - \beta\right) \xrightarrow{d} \mathcal{N}(0,\sigma^2).$$

It remains to verify (B.49). For this crucial step, recall (B.21):

$$a_t = O\left(t^{-c}\right) + \gamma\left(\xi_t + \eta_t\right) = \gamma\xi_t + \zeta_t,$$

with $\zeta_t = O\left(t^{-c}\right) + \gamma\eta_t$. We may then write

$$\sum_{t=1}^{T} a_t^2 = \gamma^2 A_T^0 + R_T$$

with

$$A_T^0 = \sum_{t=1}^{T}\xi_t^2 \quad\text{and}\quad R_T = 2\sum_{t=1}^{T}\xi_t\zeta_t + \sum_{t=1}^{T}\zeta_t^2.$$

Introducing

$$S_T = \sum_{t=1}^{T}\zeta_t^2,$$

44

it follows from the Cauchy-Schwartz inequality that

$$|R_T| \leq 2\sqrt{A_T^0 S_T} + S_T = A_T^0 \left[2\sqrt{\frac{S_T}{A_T^0}} + \frac{S_T}{A_T^0} \right].$$

Since $\sum \eta_t^2 < \infty$ a.s., cf. Remark 4,

$$S_T = O(1) \quad \text{a.s..}$$

Hence

$$\sum_{t=1}^{T} a_t^2 = A_T^0 \left[\gamma^2 + O\left(\frac{1}{\sqrt{A_T^0}} \right) \right].$$

Therefore (B.49) is implied by

$$\text{plim}_{T \to \infty} \frac{A_T^0}{\ln T} = \frac{\sigma^2}{2c - 1}. \tag{B.53}$$

Note that, by monotonicity, this implies that $A_\infty^0 = \infty$ 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. The proof of (B.53) is relegated to Appendix D. Actually, there it is shown that convergence in (B.53) takes place in L^2. It is for this step that we need fourth moments of the ε_t.

Remark 7
If one tries to apply the rationale of this proof to the case of $c = 1/2$ one encounters the problem of A_T^0 / α_T not being a Cauchy sequence in L^2 for any choice of a deterministic sequence α_T. It is hence not clear whether an asymptotic distribution of $\widehat{\beta}_T$ can be derived using an approach similar to the ones used for $c \gtrless 1/2$ in this paper.

B.3.3 Case $c < 1/2$

We continue to work with the formulas (B.36), (B.37) and (B.39). The first term in u_T is the martingale $M_T = \sum_{t=1}^{T} a_{t-1} \varepsilon_t$, which has the predictable quadratic variation $\sigma^2 A_T'$. Consider

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

Then

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

Our procedure here is as follows.

<u>*Step (i)*</u> First we show that $A_\infty' = \infty$ a.s. and $\rho^2 = \lim_{T \to \infty} A_T' / A_T$ is finite.

<u>*Step (ii)*</u> Then we show that

$$U_T'^2 \xrightarrow{d} \sigma^2 \rho^{-2} \chi^2, \tag{B.56}$$

with $\chi^2 = \chi_1^2$.

Together *Step (i)* and *Step (ii)* will imply that

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

Regarding *Step (i)*, the essential tool will be the asymptotic behavior of A_T' and A_T. Note that (B.47) is no longer valid for $c < 1/2$. We therefore go back to the properties of a_t investigated in Theorem 3 *(iii)*. For the readers convenience, we repeat the fundamental facts here:

$$t^c a_t = u + \rho_t, \tag{B.58a}$$
$$\rho_t = \gamma (v_t - v) + \gamma w_t + O\left(t^{-1}\right), \tag{B.58b}$$
$$T^c \overline{a}_T = \frac{u}{1-c} + o_p(1) \tag{B.58c}$$

cf. (B.32) - (B.34). Performing some straightforward calculations it can be shown that

$$A_T' = T^{1-2c} \frac{u^2}{1-2c} \left[1 + o_p(1) \right], \tag{B.59a}$$
$$A_T = T^{1-2c} \nu^2 u^2 \left[1 + o_p(1) \right], \tag{B.59b}$$

with

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

This already settles the assertion in *Step (i)*, showing that

$$\rho^2 = \mathrm{plim}_{T \to \infty} \frac{A_T'}{A_T} = \left(\frac{1}{c} - 1 \right)^2 \tag{B.61}$$

on the set $\{u \neq 0\}$. Note that, by Assumption DG2, $\mathbf{P}\left(u \neq 0\right) = 1$.

Let us now turn to *Step (ii)*. Note that the following derivation presumes that Assumptions MA1 and DG2 are satisfied by the error terms. Note also that the ε_t are *not* assumed to be normally distributed. Gaussian error terms will be looked at in Appendix B.4.

We go back to (B.36):

$$u_T = \sum_{t=1}^{T} a_{t-1} \varepsilon_t - \overline{a}_T^- \sum_{t=1}^{T} \varepsilon_t. \tag{B.62}$$

The second term on the right hand side can be handled making use of (B.58c) to obtain

$$\overline{a}_T^- \sum_{t=1}^{T} \varepsilon_t = \frac{u}{1-c} T^{-c} \left[1 + o_p(1) \right] \sum_{t=1}^{T} \varepsilon_t.$$

By (B.59a),

$$\sqrt{A_T'} = \frac{|u|}{\sqrt{1-2c}} T^{1/2-c} \left[1 + o_p(1) \right]. \tag{B.63}$$

Therefore,

$$
\begin{aligned}
\frac{1}{\sqrt{A_T'}}\bar{a}_T \sum_{t=1}^{T}\varepsilon_t &= \frac{\sqrt{1-2c}}{1-c}\operatorname{sign}(u)\left[1+o_p(1)\right]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t \\
&= \frac{\sqrt{1-2c}}{1-c}\operatorname{sign}(u)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t + o_p(1) \\
&= \frac{\sqrt{1-2c}}{1-c}\operatorname{sign}(u)\,X_T'' + o_p(1).
\end{aligned} \tag{B.64}
$$

Here we have put

$$
X_T'' = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t.
$$

Coming to the first term of u_T, we go back to the decomposition (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

$$
\begin{aligned}
\frac{1}{\sqrt{A_T'}}\sum_{t=1}^{T}a_{t-1}\varepsilon_t &= \frac{\sqrt{1-2c}}{1+o_p(1)}\frac{1}{T^{1/2-c}}\left[\operatorname{sign}(u)\sum_{t=1}^{T}t^{-c}\varepsilon_t \right.\\
&\qquad\qquad \left. + \sum_{t=1}^{T}t^{-c}\rho_{t-1}\varepsilon_t + O_p(1)\right] \\
&= \frac{\sqrt{1-2c}}{1+o_p(1)}\left[\operatorname{sign}(u)\,X_T' + Z_T\right] + o_p(1).
\end{aligned} \tag{B.65}
$$

Here we have put

$$
\begin{aligned}
X_T' &= \frac{1}{T^{1/2-c}}\sum_{t=1}^{T}t^{-c}\varepsilon_t, \\
Z_T &= \frac{1}{T^{1/2-c}}\sum_{t=1}^{T}t^{-c}\rho_{t-1}\varepsilon_t.
\end{aligned} \tag{B.66}
$$

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

$$
\frac{1}{\sqrt{A_T'}}\sum_{t=1}^{T}a_{t-1}\varepsilon_t = \sqrt{1-2c}\left[\operatorname{sign}(u)\,X_T' + Z_T\right] + o_p(1). \tag{B.67}
$$

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 $Z_T \xrightarrow{P} 0$.

47

Introduce

$$X_T = X_T' - \frac{1}{1-c} X_T''$$

$$= \frac{1}{T^{1/2-c}} \sum_{t=1}^{T} \left(t^{-c} - \frac{T^{-c}}{1-c} \right) \varepsilon_t. \tag{B.68}$$

We now combine (B.64) and (B.65) to obtain the following decomposition for U':

$$U_T' = \frac{u_T}{\sqrt{A_T'}} = \sqrt{1-2c} \left[\text{sign}\,(u) \left(X_T' - \frac{1}{1-c} X_T'' \right) + Z_T \right] + o_p(1)$$

$$= \sqrt{1-2c} \left[\text{sign}\,(u)\, X_T + Z_T \right] + o_p(1). \tag{B.69}$$

As for X_T,

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

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

$$= \nu^2 + o(1),$$

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

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

The LC is verified in Appendix C.3. Hence, if it can be shown that

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

it will follow from (B.69) that

$$U_T'^2 = (1-2c)\, X_T^2 + o_p(1)$$

$$= (1-2c)\, \sigma^2 \nu^2 \left(\frac{X_T}{\sigma \nu} \right)^2 + o_p(1)$$

$$\xrightarrow{d} \sigma^2 \tau^2 \chi_1^2, \tag{B.72}$$

with

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

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

It remains to show that $Z_T \overset{p}{\to} 0$, see (B.71).

To that end, return to (B.66), with

$$\rho_{t-1} = \gamma\left(v_{t-1} - v\right) + \gamma w_{t-1} + O\left(t^{-1}\right)$$

(cf. (B.58b)). For the following, we need a somewhat more detailed representation of the term $v_{t-1} - v$. Remembering (B.29) and (B.30), we write

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

with

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

Then

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

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

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

Hence

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

with

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

$$S_T = \frac{1}{T^{1/2-c}} \sum_{t=1}^{T} t^{-c} w_{t-1} \varepsilon_t. \tag{B.77}$$

Ad R. We calculate

$$
\begin{aligned}
\mathbf{E}\left[\sum_{t=1}^{T} t^{-c} \zeta_t \varepsilon_t \right]^2 &= \mathbf{E} \sum_{s,t=1}^{T} t^{-c} \zeta_t \varepsilon_t s^{-c} \zeta_s \varepsilon_s \\
&= 2\mathbf{E} \sum_{t=1}^{T} t^{-c} \zeta_t \varepsilon_t \sum_{s=1}^{t-1} s^{-c} \zeta_s \varepsilon_s + \mathbf{E} \sum_{t=1}^{T} t^{-2c} \zeta_t^2 \varepsilon_t^2 \\
&= 2\sum_{t=1}^{T} \sum_{s=1}^{t-1} t^{-c} s^{-c} \mathbf{E}\zeta_t \varepsilon_t \zeta_s \varepsilon_s + \mathbf{E} \sum_{t=1}^{T} t^{-2c} \zeta_t^2 \varepsilon_t^2 \\
&= R_{1T} + R_{2T}. \tag{B.78}
\end{aligned}
$$

49

As to R_{1T}, making use of (B.74), we obtain for $s < t$ that

$$
\begin{aligned}
\mathbf{E}\zeta_t\varepsilon_t\zeta_s\varepsilon_s &= \mathbf{E}\left\{\left[\sum_{i=t}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_t\left[\sum_{i=s}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_s\right\} \\
&= \mathbf{E}\left\{\left[\sum_{i=t}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]^2\varepsilon_t\varepsilon_s\right\} \\
&\quad +\mathbf{E}\left\{\left[\sum_{i=t}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_t\left[\sum_{i=s}^{t-1}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_s\right\} \\
&= \mathbf{E}\left\{\left[\sum_{i=t}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_t\right\}\mathbf{E}\left\{\left[\sum_{i=s}^{t-1}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_s\right\} \\
&= \sigma^4\theta_t\theta_s\frac{1}{t^{1-c}}\frac{1}{s^{1-c}}.
\end{aligned}
$$

Hence, remembering that $\lim_{t\to\infty}\theta_t = 1$,

$$
\begin{aligned}
R_{1T} &= 2\sigma^4\sum_{t=1}^{T}t^{-c}\theta_t\frac{1}{t^{1-c}}\sum_{s=1}^{t-1}s^{-c}\theta_s\frac{1}{s^{1-c}} \\
&= 2\sigma^4\sum_{t=1}^{T}\theta_t\frac{1}{t}\sum_{s=1}^{t-1}\theta_s\frac{1}{s} \\
&= O(1)\sum_{t=1}^{T}\frac{1}{t}\left[\ln t + O(1)\right] \\
&= O(1)\ln^2 T.
\end{aligned}
\tag{B.79}
$$

As to R_{2T},

$$
\begin{aligned}
\mathbf{E}\zeta_t^2\varepsilon_t^2 &= \mathbf{E}\left[\zeta_{t+1}+\theta_t\frac{\varepsilon_t}{t^{1-c}}\right]^2\varepsilon_t^2 \\
&= \mathbf{E}\zeta_{t+1}^2\mathbf{E}\varepsilon_t^2 + \frac{\theta_t^2}{t^{2(1-c)}}\mathbf{E}\varepsilon_t^4 \\
&= \sigma^4\sum_{i=t+1}^{\infty}\frac{\theta_i^2}{t^{2(1-c)}} + \frac{\theta_t^2}{t^{2(1-c)}}m_4 \\
&= O\left(t^{2c-1}\right).
\end{aligned}
$$

Hence

$$
R_{2T} = \sum_{t=1}^{T}t^{-2c}\mathbf{E}\zeta_t^2\varepsilon_t^2 = O(1)\sum_{t=1}^{T}\frac{1}{t} = O(1)\ln T.
\tag{B.80}
$$

Therefore, taking account of (B.76) together with (B.78), (B.79) and (B.80), we find that

$$
\mathbf{E}\left(R_T\right)^2 = \frac{1}{T^{1-2c}}\left[R_{1T}+R_{2T}\right] = O\left(\frac{\ln^2 T}{T^{1-2c}}\right).
$$

In particular,

$$
\mathrm{plim}_{T\to\infty}R_T = 0.
\tag{B.81}
$$

Ad S. Recalling (B.77),

$$S_T = \frac{1}{T^{1/2-c}} \sum_{t=1}^{T} t^{-c} w_{t-1} \varepsilon_t.$$

Since

$$w_{t-1} = \frac{1}{t} \sum_{i=1}^{t-1} \frac{O_{ti}(1)}{i^{1-c}} \varepsilon_i, \tag{B.82}$$

$$\mathbf{E} w_{t-1}^2 = O(1) \frac{1}{t^2} \sum_{i=1}^{t} \frac{1}{i^{2(1-c)}} = O\left(t^{-2}\right), \tag{B.83}$$

(cf. (B.29)) is $\mathcal{F}_{t-1}$-measurable and

$$\mathbf{E} S_T^2 = \frac{\sigma^2}{T^{1-2c}} \sum_{t=1}^{T} t^{-2c} \mathbf{E} w_{t-1}^2$$

$$= O(1) \frac{1}{T^{1-2c}} \sum_{t=1}^{T} \frac{1}{t^{2(1+c)}} = O\left(\frac{1}{T^{1-2c}}\right).$$

In particular,

$$\text{plim}_{T\to\infty} S_T = 0. \tag{B.84}$$

This shows (B.71).

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.59b) and (B.61),

$$V_T = \sqrt{\frac{T^{1-2c}}{A_T}} \frac{u_T}{\sqrt{A_T}} = \sqrt{\frac{T^{1-2c}}{A_T}} \sqrt{\frac{A_T'}{A_T}} \frac{u_T}{\sqrt{A_T'}}$$

$$= \frac{\rho}{\nu |u|} U_T' + o_p(1).$$

In view of the representation (B.69),

$$V_T = \frac{\rho}{\nu |u|} \sqrt{1-2c} \left[\text{sign}(u) X_T + Z_T\right] + o_p(1)$$

$$= \frac{\rho \sqrt{1-2c}}{\nu} \frac{X_T}{u} + o_p(1). \tag{B.85}$$

The crucial fact is that, under normality,

$$\begin{pmatrix} u \\ X_T \end{pmatrix} \overset{d}{\to} \mathcal{N} \left\{ \begin{pmatrix} a_0 B_0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_X^2 \end{pmatrix} \right\}, \tag{B.86}$$

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.86) follows from the fact that

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

(cf. (B.31) and (B.30)), and since X_T and v are jointly normal with covariance

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

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

$$V_T' = \frac{\sigma_u}{\sigma_x} \frac{X_T}{u} = \frac{\kappa}{\nu} \frac{X_T}{u} \xrightarrow{d} \frac{\widetilde{X}}{\widetilde{u}}, \tag{B.87}$$

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

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

An alternative representation of the density in terms of an infinite series expansion may be found in Krishnan (1968). Going back to (B.87), we then find that

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

Finally, by asymptotic equivalence of V and V',

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

Remark 9

Reconsider the result in Theorem 2 (iii). It turns out that for Gaussian error terms, u and v 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 & Spokoiny (1997).

C The Lindeberg conditions

C.1 Theorem 3 for $c \geq 1/2$

We verify the Lindeberg condition for sums of independent random variables, cf. Shiryaev (1996, Chapter III, §4, Theorem 1). Put differently, for every $\delta > 0$,

$$V_t = \frac{1}{\langle v \rangle_t} \sum_{i=1}^{t} \mathbf{E} \frac{\varepsilon_i^2}{i^{2(1-c)}} 1_{\left\{|\varepsilon_i| > \delta i^{1-c} \langle v \rangle_t^{1/2}\right\}} \to 0.$$

For $c > 1/2$, taking account of (B.22),

$$
\begin{aligned}
\left\{|\varepsilon_i| > \delta i^{1-c} \langle v \rangle_t^{1/2}\right\} &= \left\{|\varepsilon_i| > \frac{\sigma}{\sqrt{2c-1}} \delta i^{1-c} \sqrt{t^{2c-1} + O(1)}\right\} \\
&= \left\{|\varepsilon_i| > \frac{\sigma}{\sqrt{2c-1}} (1 + o(1))\delta i^{1-c} t^{c-1/2}\right\} \\
&\subset \left\{|\varepsilon_i| > \kappa (1 + o(1)) t^p\right\}
\end{aligned}
$$

with $p = (c \wedge 1) - \frac{1}{2}$ and $\kappa > 0$. The last inclusion follows from the fact that $i^{1-c} \geq t^{1-c}$ for $c \geq 1$ and $i^{1-c} \geq 1$ for $c < 1$. Therefore, by square integrability of ε_i,

$$\mathbf{E}\varepsilon_i^2 1_{\left\{|\varepsilon_i| > \delta i^{1-c} \langle v \rangle_t^{1/2}\right\}} \leq \mathbf{E}\varepsilon_1^2 1_{\{|\varepsilon_1| > \kappa(1+o(1))t^p\}} = \pi_t \to 0$$

as $t \to \infty$. As a consequence,

$$V_t \leq \frac{\pi_t}{\langle v \rangle_t} \sum_{i=1}^{t} \frac{1}{i^{2(1-c)}} = \frac{\pi_t}{\sigma^2} \to 0.$$

For $c = 1/2$, the proof runs similarly, now making use of (B.27):

$$
\begin{aligned}
\left\{|\varepsilon_i| > \delta i^{1/2} \langle v \rangle_t^{1/2}\right\} &= \left\{|\varepsilon_i| > \sigma (1 + o(1))\delta i^{1/2} \sqrt{\ln t}\right\} \\
&\subset \left\{|\varepsilon_i| > \kappa (1 + o(1)) \sqrt{\ln t}\right\},
\end{aligned}
$$

so that

$$\mathbf{E}\varepsilon_i^2 1_{\left\{|\varepsilon_i| > \delta i^{1-c} \langle v \rangle_t^{1/2}\right\}} \leq \mathbf{E}\varepsilon_1^2 1_{\left\{|\varepsilon_1| > \kappa(1+o(1))\sqrt{\ln t}\right\}} = \pi_t \to 0$$

and hence

$$V_t = \frac{1}{\langle v \rangle_t} \sum_{i=1}^{t} \mathbf{E} \frac{\varepsilon_i^2}{i} 1_{\left\{|\varepsilon_i| > \delta i^{1/2} \langle v \rangle_t^{1/2}\right\}} \leq \frac{\pi_t}{\langle v \rangle_t} \sum_{i=1}^{t} \frac{1}{i} = \frac{\pi_t}{\sigma^2} \to 0.$$

C.2 Theorem 4 for $c > 1/2$

Reconsider the martingale in (B.45), reproduced here for convenience:

$$M_T = \sum_{t=1}^{T} \xi_{Tt} \varepsilon_t, \quad \xi_{Tt} = \frac{a_{t-1}}{\sqrt{\alpha_T}}.$$

We have to show that, for every $\delta > 0$,

$$R_T = \sum_{t=1}^{T} \mathbf{E} \left\{ \xi_{Tt}^2 \varepsilon_t^2 1_{\{|\xi_{Tt}\varepsilon_t| > \delta\}} | \mathcal{F}_{t-1} \right\} \xrightarrow{p} 0, \tag{C.1}$$

cf. Christopeit & Hoderlein (2006). To this end, we make use of the elementary implication $|ab| > \delta \Rightarrow a^2 > \delta$ or $b^2 > \delta$ to obtain the inclusion $\{|\xi_{Tt}\varepsilon_t| > \delta\} = \{|a_{t-1}\varepsilon_t| > \delta\sqrt{\alpha_T}\} \subset \{a_{t-1}^2 > \delta\sqrt{\alpha_T}\} \cup \{\varepsilon_t^2 > \delta\sqrt{\alpha_T}\}$. Therefore,

$$
\begin{aligned}
R_T &\leq \frac{1}{\alpha_T} \sum_{t=1}^{T} \mathbf{E} \left\{ a_{t-1}^2 \varepsilon_t^2 1_{\{a_{t-1}^2 > \delta\sqrt{\alpha_T}\}} | \mathcal{F}_{t-1} \right\} \\
&\quad + \frac{1}{\alpha_T} \sum_{t=1}^{T} \mathbf{E} \left\{ a_{t-1}^2 \varepsilon_t^2 1_{\{\varepsilon_t^2 > \delta\sqrt{\alpha_T}\}} | \mathcal{F}_{t-1} \right\} \\
&= \frac{\sigma^2}{\alpha_T} \sum_{t=1}^{T} a_{t-1}^2 1_{\{a_{t-1}^2 > \delta\sqrt{\alpha_T}\}} + \frac{1}{\alpha_T} \sum_{t=1}^{T} a_{t-1}^2 \mathbf{E} \left\{ \varepsilon_t^2 1_{\{\varepsilon_t^2 > \delta\sqrt{\alpha_T}\}} \right\} \\
&= R_T^0 + R_T^1.
\end{aligned}
$$

As to R_T^0, since $a_t \to \alpha$ a.s., there will be a T_0 (depending on ω) such that $a_{t-1}^2 \leq \delta\sqrt{\alpha_T}$ for all $t > T_0$. Hence the sum is finite and

$$R_T^0 \to 0 \quad \text{a.s.} \tag{C.2}$$

As to R_T^1,

$$\mathbf{E} \left\{ \varepsilon_t^2 1_{\{\varepsilon_t^2 > \lambda_T \delta\}} \right\} = \pi_T \to 0.$$

Hence, taking account of (B.49),

$$R_T^1 = \frac{\pi_T}{\alpha_T} \sum_{t=1}^{T} a_{t-1}^2 = \pi_T \frac{A_T'}{\alpha_T} \xrightarrow{p} 0. \tag{C.3}$$

(C.2) and (C.3) together show (C.1).

C.3 Theorem 4 for $c < 1/2$

By definition (cf. (B.68)),

$$X_T = \sum_{t=1}^{T} \xi_{Tt}\varepsilon_t$$

with

$$\xi_{Tt} = \frac{1}{T^{1/2-c}} \left(t^{-c} - \frac{T^{-c}}{1-c} \right)$$

To show:

$$R_T = \sum_{t=1}^{T} \mathbf{E} \left\{ \xi_{Tt}^2 \varepsilon_t^2 1_{\{|\xi_{Tt}\varepsilon_t| > \delta\}} | \mathcal{F}_{t-1} \right\} \xrightarrow{p} 0.$$

But
$$R_T = \sum_{t=1}^{T} \xi_{Tt}^2 \mathbf{E}\left\{\varepsilon_t^2 1_{\left\{\varepsilon_t^2 > \delta^2/\xi_{Tt}^2\right\}}\right\}.$$

Since
$$\max_{t \leq T} |\xi_{Tt}| \leq \frac{1}{T^{1/2-c}} + \frac{1}{1-c}\frac{1}{T^{1/2}} = m_T = o(1),$$

it follows that
$$\pi_T = \mathbf{E}\left\{\varepsilon_t^2 1_{\left\{\varepsilon_t^2 > \delta^2/\xi_{Tt}^2\right\}}\right\} \leq \mathbf{E}\left\{\varepsilon_t^2 1_{\left\{\varepsilon_t^2 > \delta^2/m_T^2\right\}}\right\} \to 0.$$

Therefore,
$$R_T \leq m_T \sum_{t=1}^{T} \xi_{Tt}^2 \to 0$$

since $\sum_{t=1}^{T} \xi_{Tt}^2 = O(1)$.

D Proof of equation (B.53)

D.1 Introduction

Here we give the proof of (B.53), as announced in Appendix B.3, namely that

$$\mathbf{E}\left[\frac{\sum_{t=1}^{T}\xi_t^2}{\sigma_T^2}-1\right]^2 \to 0 \tag{D.1}$$

for the sequence

$$\sigma_T^2 = \frac{\sigma^2}{2c-1}\ln T. \tag{D.2}$$

The sequence ξ_t was defined by

$$\xi_t = \frac{1}{t^c}\sum_{i=1}^{t}\frac{\varepsilon_i}{i^{1-c}}.$$

Remember that we are dealing with the case $c > 1/2$. Denote

$$X_T = \sum_{t=1}^{T}\xi_t^2.$$

Then, since

$$\begin{aligned}
\mathbf{E}\xi_t^2 &= \frac{1}{t^{2c}}\mathbf{E}\left[\sum_{i=1}^{t}\frac{1}{i^{1-c}}\varepsilon_i\right]^2 = \frac{\sigma^2}{t^{2c}}\sum_{i=1}^{t}\frac{1}{i^{2(1-c)}}\\
&= \frac{\sigma^2}{t^{2c}}\left[\frac{1}{2c-1}t^{2c-1}+O(1)\right]\\
&= \frac{\sigma^2}{2c-1}\frac{1}{t}+O\left(\frac{1}{t^{2c}}\right),
\end{aligned}$$

it follows from the integral comparison test (ICT) that

$$\mathbf{E}X_T = \frac{\sigma^2}{2c-1}\ln T + O(1). \tag{D.3}$$

Hence, if (D.1) holds,

$$\sigma_T^{-2}\mathbf{E}X_T \to 1,$$

so that the normalisation (D.2) comes up naturally.

In order to show (D.1), we will have to calculate 4th moments:

$$\mathbf{E}X_T^2 = \sum_{s,t=1}^{T}\mathbf{E}\xi_s^2\xi_t^2 = \sum_{t=1}^{T}\mathbf{E}\xi_t^4 + 2\sum_{s,t=1;s<t}^{T}\mathbf{E}\xi_s^2\xi_t^2. \tag{D.4}$$

D.2 Calculation of 4th moments

Henceforth, we will assume that $s \leq t$. The basic formula will be

$$
\mathbf{E}\varepsilon_i\varepsilon_{i'}\varepsilon_j\varepsilon_{j'} = \begin{cases} m_4, & i = i' = j = j', \\ \sigma^4, & i = i' \neq j = j' \ \text{ or } i = j \neq i' = j' \ \text{ or } i = j' \neq i' = j \\ 0, & \text{else} \end{cases} .
$$

Then

$$
\begin{aligned}
\mathbf{E}\xi_s^2\xi_t^2 &= \frac{1}{s^{2c}t^{2c}}\mathbf{E}\left[\sum_{i=1}^{s}\frac{1}{i^{1-c}}\varepsilon_i\right]^2\left[\sum_{j=1}^{t}\frac{1}{j^{1-c}}\varepsilon_j\right]^2 \\
&= \frac{1}{s^{2c}t^{2c}}\mathbf{E}\sum_{i,i'=1}^{s}\frac{1}{i^{1-c}i'^{1-c}}\varepsilon_i\varepsilon_{i'}\sum_{j,j'=1}^{t}\frac{1}{j^{1-c}j'^{1-c}}\varepsilon_j\varepsilon_{j'} \\
&= A_{st} + B_{st}.
\end{aligned} \tag{D.5}
$$

Here we have put

$$
A_{st} = \frac{1}{s^{2c}t^{2c}}\sum_{i,i',j,j'=1}^{s}\frac{1}{i^{1-c}i'^{1-c}}\frac{1}{j^{1-c}j'^{1-c}}\mathbf{E}\varepsilon_i\varepsilon_{i'}\varepsilon_j\varepsilon_{j'}, \tag{D.6a}
$$

$$
B_{st} = \frac{1}{s^{2c}t^{2c}}\mathbf{E}\sum_{i,i'=1}^{s}\frac{1}{i^{1-c}i'^{1-c}}\varepsilon_i\varepsilon_{i'}\sum_{j,j'=s+1}^{t}\frac{1}{j^{1-c}j'^{1-c}}\varepsilon_j\varepsilon_{j'}. \tag{D.6b}
$$

Remark 10
Note that the B-term vanishes for $s = t$.

D.2.1 Ad A_{st}

$$
\begin{aligned}
A_{st} &= \frac{1}{s^{2c}t^{2c}}\left[m_4\sum_{i=1}^{s}\frac{1}{i^{4(1-c)}} + 6\sigma^4\sum_{i=2}^{s}\frac{1}{i^{2(1-c)}}\sum_{j=1}^{i-1}\frac{1}{j^{2(1-c)}}\right] \\
&= \frac{1}{s^{4c}}\left[\sum_{i=1}^{s}\frac{1}{i^{4(1-c)}} + \left(\sum_{i=1}^{s}\frac{1}{i^{2(1-c)}}\right)^2\right]O(1) \\
&= [A'_s + A''_s]O(1),
\end{aligned}
$$

with

$$
A'_{st} = \frac{1}{s^{4c}}\sum_{i=1}^{s}\frac{1}{i^{4(1-c)}}, \quad A''_{st} = \frac{1}{s^{4c}}\left(\sum_{i=1}^{s}\frac{1}{i^{2(1-c)}}\right)^2.
$$

Ad A' Since

$$
\sum_{i=1}^{s}\frac{1}{i^{4(1-c)}} = \begin{cases} O(1), & c < 3/4, \\ O(\ln s), & c = 3/4, \\ O(s^{4c-3}), & c > 3/4, \end{cases}
$$

it follows that

$$A'_s = \begin{cases} O\left(\frac{1}{s^{4c}}\right), & c < 3/4, \\[2mm] O\left(\frac{\ln s}{s^3}\right), & c = 3/4, \\[2mm] O\left(\frac{1}{s^3}\right), & c > 3/4. \end{cases} \tag{D.7}$$

As a consequence,

$$\sum_{t=1}^{T}\sum_{s=1}^{t} A'_s = O(1). \tag{D.8}$$

Ad A'' Since

$$\sum_{i=1}^{s} \frac{1}{i^{2(1-c)}} = O\left(s^{2c-1}\right),$$

we have that

$$A''_s = \frac{1}{s^{4c}} O\left(s^{2(2c-1)}\right) = O\left(s^{-2}\right). \tag{D.9}$$

Hence

$$\sum_{t=1}^{T}\sum_{s=1}^{t} A''_s = O(\ln T). \tag{D.10}$$

As a consequence, from (D.6a),

$$\sum_{s \le t} A_{st} = \sum_{t=1}^{T}\sum_{s=1}^{t} A_{st} = O(1)\sum_{t=1}^{T}\sum_{s=1}^{t} [A'_s + A''_s] = O(\ln T). \tag{D.11}$$

D.2.2 Ad B_{st}

B will turn out the leading term in (D.5). Therefore we must be more explicit about $O(1)$-terms. We will make use of the formula

$$\sum_{j=s+1}^{t} j^p = \frac{t^{p+1}}{p+1}\left[1 - \left(\frac{s}{t}\right)^{p+1} + O_{st}\left(\frac{1}{t}\right)\right], \tag{D.12}$$

which is valid for all $p > -1$.

By (D.6b),

$$
\begin{aligned}
B_{st} &= \frac{1}{s^{2c}t^{2c}}\mathbf{E}\left[\sum_{i,i'=1}^{s}\frac{1}{i^{1-c}i'^{1-c}}\varepsilon_i\varepsilon_{i'}\sum_{j,j'=s+1}^{t}\frac{1}{j^{1-c}j'^{1-c}}\mathbf{E}\left\{\varepsilon_j\varepsilon_{j'}|\mathcal{F}_s\right\}\right] \\
&= \frac{1}{s^{2c}t^{2c}}\mathbf{E}\left[\sum_{i,i'=1}^{s}\frac{1}{i^{1-c}i'^{1-c}}\varepsilon_i\varepsilon_{i'}\sum_{j=s+1}^{t}\frac{\sigma^2}{j^{2(1-c)}}\right] \\
&= \frac{\sigma^4}{s^{2c}t^{2c}}\left[\sum_{i=1}^{s}\frac{1}{i^{2(1-c)}}\right]\left[\sum_{j=s+1}^{t}\frac{1}{j^{2(1-c)}}\right] \\
&= \frac{\sigma^4}{(2c-1)^2}\frac{s^{2c-1}t^{2c-1}}{s^{2c}t^{2c}}\left[1+o(1)\right]\left[1-\left(\frac{s}{t}\right)^{2c-1}+o(1)\right] \\
&= \frac{\sigma^4}{(2c-1)^2}\frac{1}{st}\left[1-\left(\frac{s}{t}\right)^{2c-1}+o(1)\right].
\end{aligned}
$$

As a consequence,

$$
\begin{aligned}
\sum_{s<t}^{T}B_{st} &= \sum_{t=2}^{T}\sum_{s=1}^{t-1}B_{st} \\
&= \frac{\sigma^4}{(2c-1)^2}\sum_{t=2}^{T}\left[\frac{1}{t}\sum_{s=1}^{t}\frac{1}{s}\left[1+o(1)\right]-\sum_{t=2}^{T}\frac{1}{t^{2c}}\sum_{s=1}^{t}\frac{1}{s^{2(1-c)}}\right] \\
&= \frac{\sigma^4}{(2c-1)^2}\left[\sum_{t=2}^{T}\frac{1}{t}\left[\ln t+O(1)\right]-\frac{1}{2c-1}\sum_{t=2}^{T}\frac{1}{t^{2c}}\left[t^{2c-1}+O(1)\right]\right].
\end{aligned}
$$

But

$$
\begin{aligned}
\sum_{t=2}^{T}\frac{1}{t}\left[\ln t+O(1)\right] &= \left[\int_{2}^{T}\frac{\ln t}{t}dt+O(1)\right]+O(1)\left[\int_{2}^{T}\frac{dt}{t}+1\right] \\
&= \left[\frac{1}{2}\ln^2 T+O(1)\right]+O(1)\left[\ln T+1\right] \\
&= \frac{1}{2}\ln^2 T+O(\ln T)
\end{aligned}
$$

and

$$
\sum_{t=2}^{T}\frac{1}{t^{2c}}\left[t^{2c-1}+O(1)\right]=\int_{2}^{T}\frac{dt}{t}+O(1)=\ln T+O(1).
$$

Hence

$$
\sum_{s<t}^{T}B_{st}=\frac{\sigma^4}{2\left(2c-1\right)^2}\ln^2 T+O(\ln T). \tag{D.13}
$$

D.3 Synthesis

From (D.5) together with (D.11) and (D.13) it follows that

$$
\begin{aligned}
\mathbf{E}X_T^2 &= \sum_{s\le t}^{T}\mathbf{E}\xi_s^2\xi_t^2 = 2\sum_{s<t}^{T}\mathbf{E}\xi_s^2\xi_t^2 + \sum_{t=1}^{T}\mathbf{E}\xi_t^4 \\
&= 2\sum_{s<t}^{T}B_{st} + O(\ln T) \\
&= \frac{\sigma^4}{(2c-1)^2}\ln\sigma^2 T + O(\ln T).
\end{aligned}
$$

Or, put differently, with σ_T as in (D.2),

$$
\begin{aligned}
\mathbf{E}\left[\sigma_T^{-2}X_T - 1\right]^2 &= \sigma_T^{-4}\left[\mathbf{E}X_T^2 - 2\sigma_T^2\mathbf{E}X_T + \sigma_T^4 + O(\ln T)\right] \\
&= \sigma_T^{-4}O(\ln T) \\
&= O\left(\ln^{-1}T\right).
\end{aligned}
$$

This proves the assertion.

E   Proof of Corollary 3

Consider the OLS residual $\widehat{\varepsilon}_t = y_t - \widehat{\delta} - \widehat{\beta} a_{t-1} = m_t + \varepsilon_t$, where

$$m_t = (\delta - \widehat{\delta}) + (\beta - \widehat{\beta}) a_{t-1}.$$

Then

$$\sum_{t=1}^{T} \widehat{\varepsilon}_t^2 = \sum_{t=1}^{T} m_t^2 + 2 \sum_{t=1}^{T} m_t \varepsilon_t + \sum_{t=1}^{T} \varepsilon_t^2.$$

Since

$$\frac{1}{T} \sum_{t=1}^{T} m_t^2 \leq \frac{2}{T} \left[T(\delta - \widehat{\delta})^2 + (\beta - \widehat{\beta})^2 \sum_{t=1}^{T} a_{t-1}^2 \right] = o(1),$$

$$\frac{1}{T} \left| \sum_{t=1}^{T} m_t \varepsilon_t \right| \leq \left[\frac{1}{T} \sum_{t=1}^{T} m_t^2 \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \right]^{1/2} = o(1),$$

it follows that

$$\frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_t^2 = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 + o(1) \to \sigma^2$$

with probability one or in probability according to whether both $\widehat{\delta}$ and $\widehat{\beta}$ are strongly or weakly consistent.

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., Marcet, A. & Nicolini, J. P. (2015), Stock market volatility and learning. Mimeo, Federal Reserve Bank of Minneapolis.

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

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

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

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

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

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

Brockwell, P. J. & Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd edn, Springer.

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

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

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

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

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

Christopeit, N. & Hoderlein, S. (2006), 'Local partitioned regression', *Econometrica* **74**, 787–817.

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

Christopeit, N. & Massmann, M. (2012), Strong consistency of the least-squares estimator in simple regression models with stochastic regressors. TI Discussion Paper 12-109/III.

63

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

Christopeit, N. & Massmann, M. (2013*b*), A note on an estimation problem in models with adaptive learning. TI Discussion Paper 13-151/III.

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

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

Dieppe, A., González Pandiella, A., Hall, S. & Willman, A. (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. & Honkapohja, S. (2001), *Learning and Expectations in Macroeconomics*, Princeton University Press, Princeton.

Evans, G. W., Honkapohja, S., Sargent, T. J. & Williams, N. (2013), Bayesian model averaging, learning and model selection, *in* T. J. Sargent & J. Vilmunen, eds, 'Macroeconomics at the Service of Public Policy', Oxford University Press, Oxford, UK, chapter 6, pp. 99–119.

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

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

Gaspar, V., Smets, F. & Vestin, D. (2010), Inflation expectations, adaptive learning and optimal monetary policy, *in* B. M. Friedman & M. Woodford, eds, 'Handbook of Monetary Economics', Vol. 3B, Elsevier.

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

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.

Kawata, T. (1972), *Fourier Analysis in Probability Theory*, Academic Press, New York.

Kottmann, T. (1990), Learning Procedures and Rational Expectations in Linear Models with Forecast Feedback, PhD 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. (2010), 'Stochastic approximation: a survey', *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(1), 87–96.

Lai, T. L. (2003), 'Stochastic approximation', *Annals of Statistics* **31**, 391–406.

Lai, T. L. & Robbins, H. (1977), 'Strong consistency of least squares estimates in regression models', *Proceedings of the National Academy of Sciences in the USA* **74**, 2667–2669.

Lai, T. L. & Wei, C. Z. (1982*a*), '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. & Wei, C. Z. (1982*b*), 'Asymptotic properties of projections with applications to stochastic regression problems', *Journal of Multivariate Analysis* **12**, 346–370.

Lai, T. L. & Wei, C. Z. (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, North-Holland, Amsterdam, pp. 375–393.

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

Malmedier, U. & Nagel, S. (2013), Learning from inflation experience. Mimeo, URL: `http://eml.berkeley.edu/~ulrike/Papers/InflExp_44.pdf`, accepted at *Quarterly Journal of Economics.*

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

Marcet, A. & Sargent, T. J. (1988), 'The fate of systems with "adaptive" learning', *American Economic Review* **78**, 168–172.

Marcet, A. & Sargent, T. J. (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', Blackwell, chapter 6, pp. 179–215.

Markiewicz, A. & Pick, A. (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. & McFadden, D. L. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle & D. L. McFadden, eds, 'Handbook of Econometrics', Vol. 4, Elsevier, Amsterdam, pp. 2111–2245.

Phillips, P. (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. & Magdalinos, T. (2007), 'Limit theory for moderate deviations from a unit root', *Journal of Econometrics* **136**, 115–130.

Phillips, P. C. B. & Magdalinos, T. (2008), 'Limit theory for explosively cointegrated systems', *Econometric Theory* **24**, 865–887.

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, Oxford.

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 edn, Palgrave Macmillan. 1st edition 1987.

Shiryaev, A. N. (1996), *Probability*, 2nd edn, Springer, New York. 1st edition 1984.

Shiryaev, A. & Spokoiny, V. (1997), 'On sequential estimation of an autoregressive parameter', *Stochastics and Stochastic Reports* **60**(3-4), 219–240.

Vives, X. (1993), 'How fast do rational agents learn?', *Review of Economic Studies* **60**, 329–347.

Wang, X. & Yu, J. (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.