Christopeit, Norbert; Massmann, Michael

Working Paper

Estimating Structural Parameters in Regression Models with Adaptive Learning

Tinbergen Institute Discussion Paper, No. 13-111/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Christopeit, Norbert; Massmann, Michael (2013) : Estimating Structural Parameters in Regression Models with Adaptive Learning, Tinbergen Institute Discussion Paper, No. 13-111/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/87537

TI 2013-111/III
Tinbergen Institute Discussion Paper

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, VU University Amsterdam, and Tinbergen Institute, The Netherlands.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Estimating structural parameters in regression models with adaptive learning

Norbert Christopeit
University of Bonn

Michael Massmann*
Vrije Universiteit Amsterdam
and Tinbergen Institute

July 17, 2013

Abstract

This paper investigates the asymptotic properties of the ordinary least squares (OLS) estimator of structural parameters in a stylised macroeconomic model in which agents are boundedly rational and use an adaptive learning rule to form expectations of the endogenous variable. In particular, when the learning recursion is subject to so-called decreasing gain sequences the model does not satisfy, in general, any of the sufficient conditions for consistent estimability available in the literature. The paper demonstrates that, for appropriate parameter sets, the OLS estimator nevertheless remains strongly consistent and asymptotically normally distributed.

keywords: non-stationary regression, strong consistency, asymptotic normality, bounded rationality, adaptive learning.

JEL codes: C22, C51, D83

*corresponding author: Michael Massmann, Department of Econometrics, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, phone: +31 (0)20 5986014, email: m.massmann@vu.nl

1 Introduction

In economic theory, there has been considerable interest in models of the form

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

where $y^e_{t|t-1}$ denotes agents' expectations about y_t based on the information available at time $t-1$ and the driving variable x_t is exogenous. 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), as does the Lucas (1973) aggregate supply model. Of central interest is the way in which the expectations $y^e_{t|t-1}$ are modelled. The traditional approach is via *rational expectations*, cf. Muth (1961) or Sargent (2008), which assumes that agents, when forming expectations $y^e_{t|t-1}$, have complete knowledge of the model and the past $\mathcal{F}_{t-1} = \sigma(y_s, s \leq t-1; x_s, s \leq t)$ and make best use of it, i.e. set $y^e_{t|t-1} = \mathbf{E}(y_t|\mathcal{F}_{t-1})$. Taking conditional expectations in (1.1) yields $\mathbf{E}(y_t|\mathcal{F}_{t-1}) = \alpha x_t$ with

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

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

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

Obviously, under the assumption of rational expecations, 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 $y^e_{t|t-1}$. One way to specify this auxiliary model is to assume that its functional form corresponds to that of the REE in (1.3). Generally, the agents will not 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 prodedure which, in general, has the form of a stochastic approximation algorithm:

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

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

where (γ_t) is some weighting, or gain, sequence. This updating algorithm can be viewed as generalising the recursive least squares estimator of α, which has $\gamma_t = 1/t$ and whose r_t is the sample second moment of x_t. For more details on stochastic approximation algorithms, see Lai

(2003). With the learning scheme in (1.4), agents' expectation will be given by $y^e_{t|t-1} = a_{t-1}x_t$, and the resulting so-called actual law of motion, or data generating process (DGP), is

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

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

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), and the inflation model by Malmedier & Nagel (2012). 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). We will henceforth refer to this issue as the *external estimation problem (EEP)*. In particular, the EEP concerns the question of (strong) consistency of the estimator and of its asymptotic distribution. Note that the model in (1.5) is a linear regression model with stochastic regressors. Yet even though there is a rich literature on the properties of OLS estimation in this setting, the particular model in (1.5) does not appear to have been analysed yet. More strikingly, when we apply the best general result available, namely the sufficient conditions for consistency established by Lai & Wei (1982*a*, 1982*b*), they turn out not to be satisfied. With the properties of the OLS estimator of β and δ in (1.5) thus being unsettled, it is as yet unclear whether empirical implementations of it are built on sound econometric principles.

With a view to examining the EEP we will have to address the question of whether a_t in (1.4a) converges to the REE α. This issue is usually called the *internal forecasting problem (IFP)* and, in contrast to the EEP in (1.5), has been conclusively answered in the literature. A good account of seminal results can be found in Benveniste, Métivier & Priouret (1990) and Kottmann (1990). 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 and provided that $\beta < 1$. Note, however, that if $\beta \geq 1$, it can be shown that a_t diverges. For details, see Christopeit & Massmann (2010).

As to the EEP, the few existing results on the asymptotic behaviour of the OLS estimator of

β and δ pertain mainly to the case of constant gain learning, see e.g. Chevillon et al. (2010). 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 with drift, or explosive, given suitable input sequences x_t. As a consequence, we may appeal to results on the strong consistency and the limiting distribution of the OLS estimator as obtained in Lai & Wei (1985) and Chan & Wei (1988). 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). Comparing the repercussions of constant gain versus decreasing gain learning, it is hence 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.

Given that the Grenander conditions are only sufficient but not necessary for consistent estimability, it is not clear a priori whether the OLS estimators β and δ in (1.5) possess desirable asymptotic properties. The focus of this paper will hence be on investigating the EEP with decreasing gain learning. Specifically, we consider gain sequences of the form $\gamma_t = \gamma/t$ for some constant $\gamma > 0$. In view of the discussion of the IFP above, we consider the case of $\beta < 1$. Moreover, for simplicity and in order not to obscure the main ideas, we will restrict ourselves to the case with constant x_t. Simple calculation shows that, without loss of generality, we can then set $x_t = 1$ since the value of x can be accounted for by a simple change of variance of ε_t from σ^2 to σ^2/x^2. Starting the recursion (1.4b) with the stationary value[1] $r_0 = 1$, we have $r_t = 1$ and

$$a_t = a_{t-1} + \frac{\gamma}{t}\left(y_t - a_{t-1}\right). \tag{1.7}$$

[1] Actually, for any starting value r_0 it will hold that $\lim_{t\to\infty} r_t = 1$.

The model in (1.4)-(1.5) then simplifies to

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t, \tag{1.8a}$$

$$a_t = \left(1 - \frac{c}{t}\right) a_{t-1} + \frac{\gamma}{t}\left(\delta + \varepsilon_t\right), \tag{1.8b}$$

where we have put

$$c = (1 - \beta)\,\gamma. \tag{1.9}$$

Note that the condition $\beta < 1$ corresponds to $c > 0$. The value c will turn out crucial for the behaviour of the OLS estimator. It is instructive to note at this early stage several characteristics of the process a_t in (1.8b): First, a_t is autoregressive of first order with time-varying coefficient which is intrinsically local-to-unity. Next, the impact of the intercept δ and of the disturbance ε_t on a_t tends to zero for large t. As a result, the process a_t is highly non-stationary. This is reflected, for instance, in the fact that its variance decreases to zero at a rate which increases with c. Moreover, for fixed t and $h \to \infty$, the covariance $Cov(a_t, a_{t+h})$ behaves as $O\left(h^{-c}\right)$ and the correlation

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

For $c < 1/2$, the correlation tends to a non-zero constant. This corresponds to what is called long memory behaviour of stationary processes.

The paper is structured as follows: Our results are presented in Section 2. Theorems 1 and 2 pertain to the IFP in that they present precise convergence rates for a_t to α that are not available in the literature but which are necessary for the ensuing derivations. Our main results, viz. Theorems 3 and 4, concern the EEP, showing the strong consistency and asymptotic normality of the OLS estimator. Proofs are relegated to Appendix A and B, respectively.

2 Main results

The model in (1.8a) is a simple linear regression model of the form

$$y_t = \delta + \beta z_t + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{2.1}$$

with predetermined stochastic regressors $z_t = a_{t-1}$. Our focus will be on the estimation of the slope coefficient β, from which the properties of the estimator of δ follow immediately. Since, with $a_0 = 0$,

$$\overline{z}_T = \frac{1}{T}\sum_{t=1}^{T} a_{t-1} = \frac{1}{T}\sum_{s=1}^{T-1} a_s = \frac{T-1}{T}\overline{a}_{T-1},$$

the OLS estimator is given by

$$\widehat{\beta}_T - \beta = \frac{u_T}{A_T}, \tag{2.2}$$

where we have put

$$u_T = \sum_{t=1}^{T} (z_t - \overline{z}_T)\, \varepsilon_t = \sum_{t=1}^{T} \left(a_{t-1} - \frac{T-1}{T} \overline{a}_{T-1} \right) \varepsilon_t \tag{2.3}$$

and

$$A_T = \sum_{t=1}^{T} (z_t - \overline{z}_T)^2 = \sum_{t=1}^{T} \left(a_{t-1} - \frac{T-1}{T} \overline{a}_{T-1} \right)^2 . \tag{2.4}$$

For linear regression models with predetermined stochastic regressors and i.i.d. errors ε_t, the best results for the consistent estimation of β available so far have been obtained by Lai & Wei (1982*a*, 1982*b*): For general multivariate models, the condition in Lai & Wei (1982*a*) is that

$$\lambda_{\min}(M_T) \to \infty \quad \text{and} \quad \log \lambda_{\max}(M_T) = o\left(\lambda_{\min}(M_T)\right) \text{ a.s.}, \tag{2.5}$$

where $\lambda_{\min}(M_T)$ and $\lambda_{\max}(M_T)$ denote the minimal and maximal eigenvalue of the regressors' second moment matrix M_T. Lai & Wei (1982*a*) present an example which shows that condition (2.5) may be considered minimal in the sense that even a marginal violation like $\log \lambda_{\max}(M_T)/\lambda_{\min}(M_T) \to \rho > 0$ may be destructive to consistency. For the slope parameter in the simple regression model (2.1), a slight improvement is given in Lai & Wei (1982*b*) with the condition

$$\frac{A_T}{\log T} \to \infty \text{ a.s..} \tag{2.6}$$

In our model, neither (2.5) nor (2.6) is satisfied for $c > 1/2$. Rather, it is shown in Christopeit & Massmann (2013*a*) that

$$\operatorname*{plim}_{T\to\infty} \frac{\log \lambda_{\max}(M_T)}{\lambda_{\min}(M_T)} = \frac{\sigma^2}{2c-1}$$

and that

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

where the moment matrix M_T in our setting is given by

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

Strikingly, it will turn out that the OLS estimator remains consistent. When $c < 1/2$, both conditions are satisfied. For (2.5), this is shown in Christopeit & Massmann (2013*a*), while for (2.6) it will follow from our analysis of A_T in Appendix B.

It is hence plain that in examining the asymptotic properties of the OLS estimator in (2.2) no recourse can be taken to existing results. We hence resort to analysing our model from first principles, starting with the behaviour of the predetermined regressor a_t, before proceeding to that of $\widehat{\beta}$ itself. The following provides a roadmap of the results to be shown below.

Theorem 1: almost sure behaviour of a_t,

Theorem 2: asymptotic distribution of a_t,

Theorem 3: strong consistency of OLS estimator,

Theorem 4: asymptotic normality of OLS estimator.

It will turn out that the asymptotic behaviour of a_t and of the OLS estimator differs markedly according to the three following cases: $c > 1/2$, $c = 1/2$ and $c < 1/2$. Theorems 1 and 2 cover all three of cases. Theorems 3 and 4 will deal with the cases $c > 1/2$ and $c < 1/2$ as the boundary case $c = 1/2$ seems to require an entirely different approach and is thus left to future research. Some comments on the difficulties arising in the derivation will be made in Appendix B.

We will make the following assumptions about the error term ε_t in (1.8).

Assumption 1 *The ε_t are i.i.d. with finite fourth moment and $\mathbf{E}\varepsilon_t = 0$ and $\mathbf{E}\varepsilon_t^2 = \sigma^2$.*

When dealing with the EEP for $c < 1/2$, we need to make a further assumption. Define

$$v_t = \sum_{i=1}^{t} \frac{\varepsilon_i}{i^{1-c}},$$

whose limit

$$v = \lim_{t\to\infty} v_t \tag{2.7}$$

exists with probability one by Kolmogorov's theorem.

Assumption 2 $\mathbf{P}(v \neq 0) = 1$.

Remark 1 *For Gaussian ε_i, v_t is itself normal with variance*

$$\sigma_v^2 = \sigma^2 \sum_{i=1}^{\infty} i^{2(c-1)}.$$

Hence Assumption 2 is trivially satisfied.

The following two theorems concern the IFP and describe the asymptotic behaviour of a_t. Actually, the mere convergence follows easily from well known results on recursive algorithms, cf. Christopeit & Massmann (2010, section 3.1). However, for our analyses further below, e.g. the asymptotic normality of the OLS estimator, we will need the exact rates of convergence of a_t which have not yet been derived in the literature. They are also of interest *per se*.

Theorem 1 *Under Assumption 1, the following is true with probability one.*

(i) For $c > 1/2$,

$$\overline{\lim_{t\to\infty}} \sqrt{\frac{t}{\log_2 t}}\,(a_t - \alpha) = \sigma\gamma\sqrt{\frac{2}{2c-1}}. \tag{2.8}$$

(ii) For $c = 1/2$,

$$\overline{\lim_{t\to\infty}} \sqrt{\frac{t}{\log t \log_3 t}}\,(a_t - \alpha) = \sigma\gamma\sqrt{2}. \tag{2.9}$$

(iii) For $c < 1/2$,

$$\lim_{t\to\infty} t^c\,(a_t - \alpha) = \gamma v. \tag{2.10}$$

The proof will be given in Appendix A.2. Hence, the value $1/2$ is a boundary separating 'good' asymptotic behaviour of a_t from 'poor' behaviour, in the sense of speed of convergence. 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 3.

Theorem 2 *Under Assumption 1, the following is true.*

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

$$\frac{\sqrt{2c-1}}{\sigma\gamma}\sqrt{t}\,(a_t - \alpha) \xrightarrow{d} \mathcal{N}(0,1). \tag{2.11}$$

(ii) For $c = 1/2$, $a_t - \alpha$ is asymptotically normal at rate $\sqrt{t/\log t}$:

$$\frac{1}{\sigma\gamma}\sqrt{\frac{t}{\log t}}\,(a_t - \alpha) \xrightarrow{d} \mathcal{N}(0,1). \tag{2.12}$$

(iii) For $c < 1/2$, the behaviour is as described in (2.10) of Theorem 1. Note that, for Gaussian ε_t, the limit is also normal.

The proof will be given in Appendix A.3. The next two theorem are devoted to the EEP and describe the asymptotic behaviour of the OLS estimator.

Theorem 3 *Under Assumption 1 and for Gaussian ε_t, the OLS estimator is strongly consistent for every $c > 1/2$. For $c < 1/2$, this remains true if, in addition, Assumption 2 is satisfied.*

For $c > 1/2$, this is proved in a companion paper, see Christopeit & Massmann (2013*b*). For $c < 1/2$, the proof will be given in Appendix B.2.

Theorem 4 *Under Assumption 1, the following holds.*

(i) For $c > 1/2$,

$$\sqrt{A_T}\left(\widehat{\beta}_T - \beta\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{2.13}$$

In particular, the rate of convergence is $\sqrt{\log T}$:

$$\frac{\gamma}{\sqrt{2c-1}}\sqrt{\log T}\left(\widehat{\beta}_T - \beta\right) \xrightarrow{d} \mathcal{N}(0, 1). \tag{2.14}$$

(ii) For $c < 1/2$, under the additional Assumption 2,

$$\sqrt{A_T}\left(\widehat{\beta}_T - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\left(1-2c\right)\right). \tag{2.15}$$

In particular, the rate of convergence is $T^{1/2-c}$:

$$v\sqrt{T^{1-2c}}\left(\widehat{\beta}_T - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\frac{(1-c)^2\left(1-2c\right)^2}{c^2\gamma^2}\right). \tag{2.16}$$

The proof will be given in Appendix B.3.

Though it seems hard to give a general description of those distributions of the ε_t that make Assumption 2 hold, the following result may rather easily be obtained from the observation that the point spectrum of the sum of two independent random variables is the vector sum of the individual point spectra, cf. Kawata (1972, Theorem 13.1.1).

Corollary 1 *For $c < 1/2$, strong consistency and asymptotic normality are valid whenever the distribution function F of ε_t is continuous.*

A necessary and sufficient condition for the case where F has discontinuities is given in Kawata (1972, Theorem 13.1.2).

3 Discussion and conclusion

In the case of $c > 1/2$ in Theorem 4, the normalizing sequence A_T is random and can be calculated from the data. In particular, no knowledge of the constants c and γ is required. Hence, if it comes to testing hypotheses on β, (2.13) is more useful than (2.14), the variance σ^2 allowing consistent estimation from the regression residuals, cf. Corollary 2 below.

A similar remark applies when $c < 1/2$ in Theorem 4 where, however, the constant c must be estimated. Note that A_T is observable for finite sample sizes, whereas v is not. Also, (2.16) shows that there is in general no deterministic normalizing sequence γ_T such that the asymptotic distribution of $\gamma_T(\widehat{\beta}_T - \beta)$ is normal.

The asymptotic properties of the slope estimator being settled, those of the remaining parameters δ and σ^2 are easily derived. In particular, consistency of $\widehat{\delta}$ follows immediately from

$$\widehat{\delta} - \delta = (\beta - \widehat{\beta})\overline{a} + \overline{\varepsilon} \tag{3.1}$$

where subscripts have been omitted for simplicity. The type of convergence is the same as that of $\widehat{\beta}$. As to asymptotic normality, it follows from (3.1) and from

$$\sqrt{A_T}\overline{\varepsilon} = o_P(1)$$

that

$$\sqrt{A_T}(\widehat{\delta} - \delta) = \sqrt{A_T}(\beta - \widehat{\beta})\overline{a} + o_P(1).$$

As a consequence,

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

with $\lambda = 1$ when $c > 1/2$ and with $\lambda = 1 - 2c$ when $c < 1/2$.

The usual residual-based estimator of σ^2 is also consistent. The proof of the following corollary is provided in Appendix C.

Corollary 2 *Define the OLS residual* $\widehat{\varepsilon}_t = y_t - \widehat{\delta} - \widehat{\beta}a_{t-1}$. *Then*

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2 \to \sigma^2$$

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

Finally, consider γ. When γ is known, c can be estimated consistently by

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

cf.(1.9). If, on the other hand, γ unknown, it may be consistently estimated from the observed a_t using the recursion (1.7):

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

This estimator will be needed when implementing (2.15), i.e.

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

In summary, this paper investigated the asymptotic properties of the OLS estimator in bivariate regression models in which the regressor is generated by a recursive algorithm and, as such, appears in the form of an autoregressive process with time varying, local-to-unity, coefficients as well as error terms whose magnitude decreases over time, see (1.4). The regressor thus exhibits a behavior that is completely different from that of processes generated by time-invariant dynamics. The setup was motivated by the problem of estimating structural parameters in a typical

macroeconomic model in which agents are boundedly rational and use an adaptive learning rule to form expectations of the endogenous variable. Our particular interest lay on analysing a model with a learning recursion that is subject to so-called decreasing gain sequences, implying, economically speaking, that agents learn to be rational asymptotically. From a statistical point of view we noticed that this model does not generally satisfy any of the sufficient conditions for consistent estimability available in the literature. The paper hence demonstrated that, for appropriate parameter sets, the OLS estimator of the structural parameters nevertheless remains strongly consistent and asymptotically normally distributed. Moreover, there turned out to be a trade-off between the speed of convergence of agents' forecasts to the rational expectations equilibrium on the one hand and that of the OLS estimator to the true parameter value on the other.

Several extensions of our setup seem appropriate yet were beyond the scope of the present paper and and are hence left to future research. Firstly, the question of whether or not the OLS estimator is srongly consistent and asymptotically normal in the boundary case of $c = 1/2$ appears to require a line of attack different from the one employed here. Secondly, our assumptions that the exogenous regressor x_t is constant was made for analytical tractability and should be generalised. Thirdly, the expectational term in the economic model (1.1) could be specified as $y^e_{t+1|t}$, i.e. as based on information on the past *and the present*: in this case the parameters would have to be estimated by means of an instrumental variable technique, which would require suitable assumptions on the (exogenous or endogenous) instruments. Nonetheless, by showing that strongly consistent and asymptotically normal estimation of the structural parameters in our sylised model is possible, we provide a theoretical justification of why the increasingly popular empirical estimation of such models is, in principle at least, econometrically sound.

A Proof of Theorems 1 and 2

A.1 Representation of a_t

We return now to the difference equation (1.8b) for a_t. It is intuitively clear and can, indeed, be easily shown that the choice of initial value a_0 has no influence on the asymptotic behaviour. For simplicity of exposition, we shall therefore suppose that $a_0 = 0$. Then

$$a_t = \sum_{i=1}^{t} \phi_{ti} \left(\delta + \varepsilon_i\right), \tag{A.1}$$

where

$$\begin{aligned} \phi_{ti} &= \frac{\gamma}{i}\left(1 - \frac{c}{i+1}\right)\cdots\left(1 - \frac{c}{t}\right), \quad i = 1, \ldots, t-1, \\ \phi_{tt} &= \frac{\gamma}{t}. \end{aligned} \tag{A.2}$$

In particular, for $c = 1$,

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

for all i. Define $i_0 = [c]$ for $c > 1$, and $i_0 = 1$ for $c \leq 1$. Taking logarithms and using a second order Taylor expansion, it is shown in Christopeit & Massmann (2010) that, for $i_0 \leq i$,

$$\begin{aligned} \phi_{ti} &= \gamma \frac{1}{t^c} \frac{1}{i^{1-c}} \exp\left[\frac{O_{ti}(1)}{i}\right] \\ &= \gamma \frac{1}{t^c} \frac{1}{i^{1-c}} \left[1 + \frac{O_{ti}(1)}{i}\right]. \end{aligned} \tag{A.3}$$

The $O_{ti}(1)$ are uniformly bounded in t and $i \leq t$. (A.3) is well defined for all $1 \leq i \leq t$. Henceforth, call it h_{ti}. For $i < i_0$,

$$\phi_{ti} = \frac{\gamma}{i} \prod_{j=i+1}^{i_0} \left(1 - \frac{c}{j}\right) \prod_{j=i_0+1}^{t} \left(1 - \frac{c}{j}\right) = \lambda_i \phi_{ti_0}$$

with

$$\lambda_i = \frac{i_0}{i} \prod_{j=i+1}^{i_0} \left(1 - \frac{c}{j}\right).$$

Hence, since $\max_{i<i_0} |\lambda_i| \leq K$, we have that $\phi_{ti} = O\left(t^{-c}\right)$ for $i < i_0$. The same is true for h_{ti}. Therefore

$$\phi_{ti} = \begin{cases} h_{ti}, & \text{for } i \geq i_0, \\ h_{ti} + O\left(t^{-c}\right), & \text{for } 1 \leq i < i_0. \end{cases}$$

Remark 2 *Note that, for $c \leq 1$, the $O\left(t^{-c}\right)$-term vanishes.*

Since

$$\begin{aligned} \sum_{i=1}^{i_0-1} \phi_{ti} &= O\left(t^{-c}\right), \\ \sum_{i=1}^{i_0-1} \phi_{ti}\varepsilon_i &= O\left(t^{-c}\right) \quad \text{a.s.}, \end{aligned}$$

with the same holding for h_{ti}, we may write

$$\begin{aligned} a_t &= \sum_{i=1}^{t} h_{ti}\left(\delta+\varepsilon_i\right)+O\left(t^{-c}\right) \\ &= \delta\zeta_t+\theta_t+O\left(t^{-c}\right), \end{aligned} \tag{A.5}$$

with

$$\begin{aligned} \zeta_t &= \sum\nolimits_{i=1}^{t} h_{ti}, \\ \theta_t &= \sum\nolimits_{i=1}^{t} h_{ti}\varepsilon_i. \end{aligned}$$

By the integral comparison theorem, hereafter referred to as ICT, for $c \neq 1$,

$$\zeta_t = \gamma\frac{1}{t^c}\sum\nolimits_{i=1}^{t}\frac{1}{i^{1-c}}\left[1+\frac{O_{ti}(1)}{i}\right] = \frac{\gamma}{c}+O\left(t^{-1}\right)+O(t^{-c}) \tag{A.6}$$

and, for $c = 1$,

$$\zeta_t = \gamma+O\left(\frac{\log t}{t}\right). \tag{A.7}$$

Next, write

$$\theta_t = \frac{\gamma}{t^c}\left[\sum\nolimits_{i=1}^{t}\frac{\varepsilon_i}{i^{1-c}}+\sum\nolimits_{i=1}^{t}\frac{O_{ti}(1)}{i^{2-c}}\varepsilon_i\right] = \gamma\frac{1}{t^c}\left(v_t+w_t\right), \tag{A.8}$$

with

$$\begin{aligned} v_t &= \sum\nolimits_{i=1}^{t}\frac{\varepsilon_i}{i^{1-c}}, \\ w_t &= \sum\nolimits_{i=1}^{t}\frac{O_{ti}(1)}{i^{2-c}}\varepsilon_i. \end{aligned}$$

Then, noting that $\delta\gamma/c = \alpha$ and introducing the processes

$$\begin{aligned} \xi_t &= \frac{1}{t^c}v_t \\ \eta_t &= \frac{1}{t^c}w_t \end{aligned}$$

we may write

$$a_t-\alpha = O\left(t^{-1}\right)+O(t^{-c})+\gamma\left(\xi_t+\eta_t\right), \tag{A.9}$$

with the $O(t^{-c})$ vanishing for $c \leq 1$.

A.2 Proof of Theorem 1

In the following, we will derive the asymptotic behaviour of $a_t-\alpha$. In addition, we shall also consider that of $\overline{a}_T-\alpha$ because it is needed in the treatment of the EEP in Appendix B.

A.2.1 Case $c > 1/2$

Reconsider the representation of $a_t - \alpha$ in (A.9) above and examine first the behaviour of ξ_t. By the ICT, the predictable quadratic variation of v_t is given by

$$\langle v \rangle_t = \sigma^2 \sum\nolimits_{i=1}^{t} i^{2(c-1)} = \frac{\sigma^2}{2c-1} t^{2c-1} + O(1). \tag{A.10}$$

Hence $\langle v \rangle_\infty = \lim_{t\to\infty} \langle v \rangle_t = \infty$ a.s.. By the law of iteration logarithms, henceforth denoted by LIL, cf. Chow & Teicher (1973),

$$\overline{\lim_{t\to\infty}} \frac{|v_t|}{\sqrt{2 \langle v \rangle_t \log_2 \langle v \rangle_t}} = 1 \text{ a.s..}$$

In view of (A.10), this means that

$$\overline{\lim_{t\to\infty}} \frac{|v_t|}{\sqrt{t^{2c-1} \log_2 t}} = \sigma \sqrt{\frac{2}{2c-1}} \text{ a.s.}$$

or

$$\overline{\lim_{t\to\infty}} \sqrt{\frac{t}{\log_2 t}} \, |\xi_t| = \overline{\lim_{t\to\infty}} \sqrt{\frac{t}{\log_2 t}} \frac{|v_t|}{t^c} = \sigma \sqrt{\frac{2}{2c-1}} \text{ a.s..} \tag{A.12}$$

Turning to η_t in (A.9), it follows again by the ICT that

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

Therefore

$$\mathbf{E} \sum_{t=1}^{\infty} t^{2\rho} \left(\frac{w_t}{t^c}\right)^2 < \infty$$

for every $0 \leq \rho < 1$. As a consequence,

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

In particular,

$$t^\rho \eta_t = o(1) \text{ a.s..} \tag{A.13}$$

(A.12) and (A.13) taken together then yield

$$\overline{\lim_{t\to\infty}} \sqrt{\frac{t}{\log_2 t}} \, (a_t - \alpha) = \sigma\gamma \sqrt{\frac{2}{2c-1}} \text{ a.s.,}$$

proving assertion (2.8) in Theorem 1.

Turning to $\overline{a}_T - \alpha$, by (2.8),

$$\begin{aligned} \overline{a}_T - \alpha &= \frac{1}{T} \sum_{t=1}^{T} (a_t - \alpha) = O(1) \frac{1}{T} \sum_{t=1}^{T} \sqrt{\frac{\log_2 t}{t}} \\ &= O\left(\sqrt{\frac{\log_2 T}{T}}\right) \text{ a.s..} \end{aligned} \tag{A.14}$$

A.2.2 Case $c = 1/2$

Reconsider $a_t - \alpha$ in (A.9). Regarding ξ_t, note that the predictable quadratic variation of v_t is

$$\langle v \rangle_t = \sigma^2 \sum\nolimits_{i=1}^{t} i^{-1} = \sigma^2 \log t + O(1). \tag{A.15}$$

Again by the LIL,

$$\overline{\lim_{t\to\infty}} \frac{|v_t|}{\sqrt{\log t \log_3 t}} = \sigma\sqrt{2} \text{ a.s..}$$

In particular,

$$\overline{\lim_{t\to\infty}} \sqrt{\frac{t}{\log t \log_3 t}}\, |\xi_t| = \sigma\sqrt{2} \text{ a.s..} \tag{A.16}$$

As for η_t,

$$\mathbf{E} w_t^2 = O\left(t^{-2}\right), \tag{A.17}$$

so that

$$\mathbf{E} \sum_{t=1}^{\infty} t^{2\rho} \left(\frac{w_t}{t^{1/2}}\right)^2 < \infty$$

for every $\rho < 1$. Therefore

$$t^{\rho} \eta_t = o(1) \text{ a.s..} \tag{A.18}$$

It then follows from (A.16) and (A.18) that

$$\overline{\lim_{t\to\infty}} \sqrt{\frac{t}{\log t \log_3 t}}\, (a_t - \alpha) = \sigma\gamma\sqrt{2} \text{ a.s.,}$$

as claimed in (2.9) of Theorem 1.

A.2.3 Case $c < 1/2$

Finally, we examine $a_t - \alpha$ in (A.9) in the case of $c < 1/2$. First, look again at ξ_t. By Kolomogorov's theorem, v_t converges with probability one to some finite random variable v, see also (2.7). Actually, it is easily verified that convergence of v_t to v also takes place in L^2, with the limit v having variance $\sigma_v^2 = \sigma^2 \sum_{i=1}^{\infty} i^{2(c-1)}$. Hence

$$t^c \xi_t = v_t = v + o(1) \text{ a.s..} \tag{A.19}$$

As far as η_t is concerned,

$$\mathbf{E} w_t^2 = O\left(t^{2c-3}\right),$$

so that

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

and

$$t^c \eta_t = w_t = o(1) \quad \text{a.s..} \tag{A.20}$$

Hence, remembering that the $O(t^{-c})$-term in (A.9) vanishes for $c < 1/2$,

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

This proves (2.10) in Theorem 1.

As to $\overline{a}_T - \alpha$, we have

$$\begin{aligned} \overline{a}_T - \alpha &= \frac{1}{T}\sum_{t=1}^{T}\left(a_t - \alpha\right) = \frac{1}{T}\sum_{t=1}^{T} t^{-c}\left(\gamma v + o(1)\right) \\ &= \frac{\gamma v}{1-c}\frac{1}{T^c} + o\left(T^{-c}\right) \end{aligned} \tag{A.21}$$

which follows from (2.10).

A.3 Proof of Theorem 2

A.3.1 Case $c > 1/2$

By (A.10) and the central limit theorem for martingales,

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

In terms of ξ_t, this is equivalent to

$$\frac{\sqrt{2c-1}}{\sigma}\sqrt{t}\xi_t \xrightarrow{d} \mathcal{N}(0,1). \tag{A.22}$$

Returning to (A.9), we obtain that

$$\sqrt{t}\left(a_t - \alpha\right) = \gamma\sqrt{t}\xi_t + \gamma\sqrt{t}\eta_t + O\left(t^{-1/2}\right) + O(t^{1/2-c}).$$

By virtue of (A.13),

$$\sqrt{t}\left(a_t - \alpha\right) = \gamma\sqrt{t}\xi_t + o(1),$$

which, due to (A.22), entails (2.11) in Theorem 2.

A.3.2 Case $c = 1/2$

In this case, by (A.15),

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

or, in terms of ξ_t,

$$\frac{1}{\sigma}\sqrt{\frac{t}{\log t}}\xi_t \xrightarrow{d} \mathcal{N}(0,1). \tag{A.23}$$

Taking account of (A.9) and (A.18),

$$\sqrt{\frac{t}{\log t}}\,(a_t-\alpha)=\gamma\sqrt{\frac{t}{\log t}}\xi_t+\gamma\sqrt{\frac{t}{\log t}}\eta_t+O\left(\sqrt{\frac{1}{t\log t}}\right).$$

By virtue of (A.18),

$$\sqrt{\frac{t}{\log t}}\,(a_t-\alpha)=\gamma\sqrt{\frac{t}{\log t}}\xi_t+o(1)\ \text{ a.s.},$$

which, due to (A.23), entails (2.12) in Theorem 2.

Remark 3 *For $c\geq 1/2$, the Liapunov condition for 4^{th} moments is satisfied. In particular, for $c>1/2$, this follows from*

$$\begin{aligned}\sum_{i=1}^{t}\mathbf{E}\left[\frac{\varepsilon_i}{i^{1-c}}\right]^4 &= m_4\sum_{i=1}^{t}i^{4(c-1)}=O\left(t^{4c-3}\right)+O(1),\\ \langle v\rangle_t^2 &= \frac{\sigma^4}{(2c-1)^2}t^{2(2c-1)}+O\left(t^{2c-1}\right),\end{aligned}$$

so that

$$\frac{1}{\langle v\rangle_t^2}\sum_{i=1}^{t}\mathbf{E}\left[\frac{\varepsilon_i}{i^{1-c}}\right]^4=O(t^{-1}).$$

For $c=1/2$,

$$\sum_{i=1}^{t}\mathbf{E}\left[\frac{\varepsilon_i}{i^{1/2}}\right]^4=O(1),\quad \langle v\rangle_t^2=\sigma^4\log^2 t\,(1+o(1)),$$

from which the Liapunov condition follows.

A.3.3 Case $c<1/2$

It follows from (2.10) that

$$t^c\,(a_t-\alpha)=\gamma v+o(1)\ \text{ a.s.},$$

where v is a random variable with variance

$$\sigma_v^2=\sigma^2\sum_{i=1}^{\infty}i^{2(c-1)}.$$

If the ε_t are normal, then so is v.

B Proof of Theorems 3 and 4

B.1 The OLS-estimator

Recall from Section 2 that the OLS estimator is given by

$$\widehat{\beta}_T - \beta = \frac{u_T}{A_T}, \tag{B.1}$$

where

$$u_T = \sum_{t=1}^{T} (x_t - \overline{x}_T)\,\varepsilon_t = \sum_{t=1}^{T} \left(a_{t-1} - \frac{T-1}{T}\overline{a}_{T-1} \right) \varepsilon_t$$

and

$$A_T = \sum_{t=1}^{T} (x_t - \overline{x}_T)^2 = \sum_{t=1}^{T} \left(a_{t-1} - \frac{T-1}{T}\overline{a}_{T-1} \right)^2.$$

Making use of the elementary algebraic identity

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

we may write

$$\begin{aligned} A_T &= \sum_{t=2}^{T} \frac{t-1}{t} (x_t - \overline{x}_{t-1})^2 \\ &= \sum_{t=2}^{T} \frac{t-1}{t} \left(a_{t-1} - \frac{t-2}{t-1}\overline{a}_{t-2} \right)^2. \end{aligned} \tag{B.2}$$

Remark 4 *(B.2) shows that A_T is nondecreasing.*

Making use of the fact that $\lim_{t\to\infty} a_t = \alpha$, our proofs will rely on the following decomposition:

$$\begin{aligned} u_T &= \sum_{t=1}^{T} \left(a_{t-1} - \frac{T-1}{T}\overline{a}_{T-1} \right) \varepsilon_t \\ &= \sum_{t=1}^{T} (a_{t-1} - \alpha)\,\varepsilon_t + \left(\alpha - \frac{T-1}{T}\overline{a}_{T-1} \right) \sum_{t=1}^{T} \varepsilon_t \\ &= \sum_{t=1}^{T} (a_{t-1} - \alpha)\,\varepsilon_t + (\alpha - \overline{a}_T) \sum_{t=1}^{T} \varepsilon_t + O\left(\frac{1}{T}\right) \sum_{t=1}^{T} \varepsilon_t. \end{aligned} \tag{B.3}$$

As to A_T, we simplify it to

$$\begin{aligned} A_T &= \sum_{t=1}^{T} \left(a_{t-1} - \overline{a}_{T-1} + \frac{1}{T}\overline{a}_{T-1} \right)^2 \\ &= \sum_{t=1}^{T} (a_{t-1} - \overline{a}_{T-1})^2 - \frac{2}{T}\overline{a}_{T-1} \sum_{t=1}^{T} (a_{t-1} - \overline{a}_{T-1}) + \frac{1}{T^2}\overline{a}_{T-1}^2 \\ &= A'_T + o(1), \end{aligned} \tag{B.4}$$

where we have put

$$A'_T = \sum_{t=1}^{T} (a_{t-1} - \overline{a}_T)^2 . \tag{B.5}$$

The expression A'_T will be further decomposed as

$$\begin{aligned} A'_T &= \sum_{t=1}^{T} [(a_{t-1} - \alpha) - (\overline{a}_T - \alpha)]^2 \\ &= \sum_{t=1}^{T} (a_{t-1} - \alpha)^2 + T (\overline{a}_T - \alpha)^2 - 2 (\overline{a}_T - \alpha) \sum_{t=1}^{T} (a_{t-1} - \alpha) . \end{aligned}$$

The last term may in turn be written as

$$\begin{aligned} (\overline{a}_T - \alpha) \sum_{t=1}^{T} (a_{t-1} - \alpha) &= (\overline{a}_T - \alpha) \left[\sum_{t=1}^{T} (a_t - \alpha) - a_T \right] \\ &= T (\overline{a}_T - \alpha)^2 - a_T (\overline{a}_T - \alpha) \\ &= T (\overline{a}_T - \alpha)^2 + o(1). \end{aligned}$$

Hence

$$A'_T = A^0_T - B_T + o(1), \tag{B.6}$$

where we have put

$$\begin{aligned} A^0_T &= \sum_{t=1}^{T} (a_{t-1} - \alpha)^2 \\ B_T &= T (\overline{a}_T - \alpha)^2 . \end{aligned}$$

The procedure will then be as follows. Using (B.3), write (B.1) in the form

$$\begin{aligned} \frac{u_T}{A_T} &= \frac{\sum_{t=1}^{T} (a_{t-1} - \alpha) \varepsilon_t}{A_T} + \frac{\alpha - \overline{a}_T}{A_T} \sum_{t=1}^{T} \varepsilon_t + O(1) \frac{1}{A_T} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \\ &= \left[\frac{\sum_{t=1}^{T} (a_{t-1} - \alpha) \varepsilon_t}{A^0_T} + \frac{\alpha - \overline{a}_T}{A^0_T} \sum_{t=1}^{T} \varepsilon_t + O(1) \frac{1}{A^0_T} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \right] \frac{A^0_T}{A_T}. \end{aligned} \tag{B.7}$$

(B.7) will be the basic decomposition of the OLS estimator on which the ensuing proofs are based. In particular, to show consistency we will proceed by verifying the following conditions, all holding with probability one:

$$\begin{aligned} \lim_{T \to \infty} A^0_T &= \infty, && \text{Condition (i)} \\ \overline{\lim}_{T \to \infty} \frac{B_T}{A^0_T} &< 1, && \text{Condition (ii)} \\ \sqrt{T \log_2 T} \frac{\overline{a}_T - \alpha}{A^0_T} &= o(1). && \text{Condition (iii)} \end{aligned}$$

Since, by virtue of (B.6),

$$\frac{A^0_T}{A'_T} = \frac{A^0_T}{A^0_T - B_T + o(1)} = \frac{1}{1 - \frac{B_T}{A^0_T} + o(1)}, \tag{B.8}$$

it will follow from Condition (ii) and (B.4) that, with probability one,

$$\overline{\lim}_{T\to\infty}\frac{A_T^0}{A_T}=\overline{\lim}_{T\to\infty}\frac{A_T^0}{A_T'}<\infty. \tag{B.9}$$

Hence, in view of Condition (i),

$$\lim_{T\to\infty}A_T'=\lim_{T\to\infty}A_T=\infty.$$

By the LIL, under Condition (iii) the middle term in brackets in (B.7) becomes

$$\begin{aligned}\frac{\alpha-\overline{a}_T}{A_T^0}\sum_{t=1}^{T}\varepsilon_t &= \sqrt{T\log_2 T}\frac{\alpha-\overline{a}_T}{A_T^0}\frac{1}{\sqrt{T\log_2 T}}\sum_{t=1}^{T}\varepsilon_t\\ &= \sqrt{T\log_2 T}\frac{\alpha-\overline{a}_T}{A_T^0}O(1)\\ &= o(1).\end{aligned}$$

As to the first term, it follows from the standard martingale convergence theorem applied to the martingale

$$M_T=\sum_{t=1}^{T}(a_{t-1}-\alpha)\,\varepsilon_t,$$

whose predictable quadratic variation is $\langle M\rangle_T=\sigma^2A_T^0$, that

$$\frac{M_T}{A_T^0}\to 0$$

with probability one. Hence, if Conditions (i)-(iii) are satisfied, it will follow from (B.7) that

$$\lim_{T\to\infty}\frac{u_T}{A_T}=0$$

with probability one. As a consequence, the OLS estimator in (2.2) is strongly consistent.

B.2 Proof of Theorem 3

B.2.1 Case $c>1/2$

The essence of the subsequent discussion is that, for $c>1/2$, Condition (i) turns out to be valid. As to Condition (ii) and Condition (iii), they require the specification of the almost sure rate of divergence of A_T^0 can be shown to be satisfied in probability yet it is not clear whether or not they also hold almost surely. Therefore, for the proof of *strong* consistency when $c>1/2$ we refer to a companion paper, see Christopeit & Massmann (2013*b*, Proposition 2), which adopts a decomposition of the OLS estimator that differs from (B.7) and thus does not require the specification of the almost sure rate of divergence of A_T^0. In that approach, we do make the additional assumption, however, that the error term ε_t is Gaussian.

The purpose of the following derivations is to examine the asymptotic behaviour of A_T^0 which, in turn, will be used in the proof of the asymptotic normality of the OLS estimator in Appendix B.3. It will be seen in the process that Condition (i) holds, as mentioned above. To that end, reconsider the expression of $a_t - \alpha$ given in (A.9) in Appendix A.1, namely

$$a_t - \alpha = O\left(t^{-1}\right) + O(t^{-c}) + \gamma\left(\xi_t + \eta_t\right),$$

with

$$\begin{aligned} \xi_t &= \frac{1}{t^c}\sum\nolimits_{i=1}^{t} \frac{\varepsilon_i}{i^{1-c}}, \\ \eta_t &= \frac{1}{t^c}\sum\nolimits_{i=1}^{t} \frac{O_{ti}(1)}{i^{2-c}}\varepsilon_i. \end{aligned}$$

Consider the asymptotic behaviour of $A_T^0 = \sum_{t=1}^{T}\left(a_{t-1} - \alpha\right)^2$. It hinges on the following two facts.

Firstly, it holds with probability one that

$$\sum_{t=1}^{\infty} \eta_t^2 < \infty. \tag{B.10}$$

Therefore,

$$A_T^0 = \gamma^2 \sum_{t=1}^{T} \xi_t^2 + O(1). \tag{B.11}$$

Actually, (B.10) follows immediately from

$$\begin{aligned} \mathbf{E}\eta_t^2 &= O(1)\frac{1}{t^{2c}}\sum\nolimits_{i=1}^{t} \frac{1}{i^{2(2-c)}} \\ &= \frac{1}{t^{2c}} \begin{cases} O(1), & 1/2 < c < 3/2, \\ \log t + O(1), & c = 3/2, \\ \frac{1}{2c-3}t^{2c-3} + O(1), & c > 3/2, \end{cases} \\ &= O(t^{-(1+\delta)}) \end{aligned}$$

for some $\delta > 0$ (depending on c). In any case, $\sum_{t=1}^{\infty} \mathbf{E}\eta_t^2 < \infty$ and hence (B.10).

Secondly, for the sequence

$$\alpha_T^2 = \frac{\sigma^2}{2c-1}\log T \tag{B.12}$$

it holds that

$$\alpha_T^{-2} \sum_{t=1}^{T} \xi_t^2 \xrightarrow{L^2} 1. \tag{B.13}$$

This can be shown by performing some tedious but rather straightforward calculations on 4^{th} moments. The proof is relegated to the Supplement I.

As a consequence of (B.11) and (B.13),

$$\frac{A_T^0}{\alpha_T^2} \xrightarrow{L^2} \gamma^2 \tag{B.14}$$

and therefore also

$$\text{plim}_{T\to\infty}\frac{A_T^0}{\alpha_T^2}=\gamma^2. \tag{B.15}$$

Since the sequence $\sum_{t=1}^{T}\xi_t^2$ is monotone increasing, (B.13) implies that

$$\sum_{t=1}^{\infty}\xi_t^2=\infty \quad \text{a.s..} \tag{B.16}$$

Hence, by (B.14),

$$\lim_{T\to\infty} A_T^0=\infty \quad \text{a.s.,} \tag{B.17}$$

so that Condition (i) is satisfied.

Recall now the order of magnitude of $\overline{a}_T-\alpha$ in (A.14), namely

$$\overline{a}_T-\alpha=O\left(\sqrt{\frac{\log_2 T}{T}}\right) \quad \text{a.s..}$$

It hence follows that, with probability one,

$$B_T=T\left(\overline{a}_T-\alpha\right)^2=O\left(\log_2 T\right). \tag{B.18}$$

Consequently, combining (B.12), (B.15) and (B.18) yields

$$\frac{B_T}{A_T^0}=\frac{B_T}{\log_2 T}\frac{\log_2 T}{\alpha_T^2}\frac{\alpha_T^2}{A_T^0}=O(1)o\left(1\right)\frac{\log_2 T}{A_T^0}=o_P(1). \tag{B.19}$$

It is hence *in probability* that Condition (ii) is satisfied, which will indeed be sufficient for weak consistency and asymptotic normality, see Theorem 4 and Appendix B.3. Specifically, we will need it in the form

$$\text{plim}_{T\to\infty}\frac{A_T^0}{A_T}=1, \tag{B.20}$$

which follows from (B.8) and $A_T=A_T'+o(1)$. Similarly, regarding Condition (iii), it follows from (A.14) in Appendix A.2 that

$$\sqrt{T\log_2 T}\frac{\overline{a}_T-\alpha}{A_T^0}=O\left(1\right)\frac{\log_2 T}{A_T^0}=o_P(1) \tag{B.21}$$

such that Condition (iii) is also satisfied *in probability.*

Remark 5 *It cannot be inferred from (B.15) that convergence of A_T^0/α_T^2 takes place with probability one. Otherwise convergence in (B.19) and (B.21), too, would hold with probability one, thus implying strong consistency.*

B.2.2 Case $c = 1/2$

Remark 6 *Note that the procedure outlined in Appendix B.2.1 does not work for $c = 1/2$. The reason is that there is no deterministic sequence α_T^2 such that $\alpha_T^{-2}\sum_{t=1}^T \xi_t^2$ converges in probability to some finite nonzero limit, see* (B.15)*. Actually, it is shown in Supplement II that $\alpha_T^{-2}\sum_{t=1}^T \xi_t^2$ is not a Cauchy sequence in L^2 and hence does not converge in L^2 for any deterministic sequence α_T^2. Together with the uniform integrability of the sequence $\alpha_T^{-2}\sum_{t=1}^T \xi_t^2$, which is also shown in Supplement II, this implies that it cannot converge in probability to a nonzero limit.*

B.2.3 Case $c < 1/2$

Remember that the $O(t^{-c})$ vanishes for $c \leq 1$, see Remark 2

Consider A_T^0 first. By definition and (2.10),

$$\begin{aligned} A_T^0 &= \sum_{t=1}^T (a_{t-1} - \alpha)^2 = \left[\gamma^2 v^2 + o(1)\right] \sum_{t=1}^T \frac{1}{t^{2c}} \\ &= \frac{\gamma^2 v^2}{1-2c} T^{1-2c} + o\left(T^{1-2c}\right). \end{aligned} \tag{B.22}$$

Hence Condition (i) in Appendix B.1 is met. Note that Assumption 2 that v is nonzero with probability one is crucial to obtain divergence with probability one.

The behaviour of B_T follows from (B.6):

$$B_T = T\left(\overline{a}_{T-1} - \alpha\right)^2 = \frac{\gamma^2 v^2}{(1-c)^2} T^{1-2c} + o\left(T^{1-2c}\right).$$

Consequently,

$$\frac{B_T}{A_T^0} = \frac{1-2c+o(1)}{(1-c)^2 + o(1)} = \frac{1-2c}{(1-c)^2} + o(1).$$

Since $(1-2c)/(1-c)^2 < 1$, this shows Condition (ii). Also,

$$\frac{\overline{a}_T - \alpha}{A_T^0} = \frac{\frac{\gamma v}{1-c} T^{-c}\left[1 + o(1)\right]}{\frac{\gamma^2 v^2}{1-2c} T^{1-2c}\left[1 + o(1)\right]} = \frac{1}{\gamma v}\frac{1-2c}{1-c}\frac{1}{T^{1-c}}\left[1 + o(1)\right].$$

Hence

$$\sqrt{T \log_2 T}\frac{\overline{a}_T - \alpha}{A_T^0} = O(1)\sqrt{\frac{\log_2 T}{T^{1-2c}}},$$

so that Condition (iii) is also satisfied. Consequently, the OLS estimator is strongly consistent.

Finally, for later reference, we note that

$$\lim_{T\to\infty} \frac{A_T^0}{A_T} = \frac{1}{1 - \frac{1-2c}{(1-c)^2}} = \frac{(1-c)^2}{c^2} \quad \text{a.s.} \tag{B.24}$$

cf. (B.8).

B.3 Proof of Theorem 4

B.3.1 Case $c > 1/2$

We return now to the decomposition (B.7), re-written as

$$\frac{u_T}{\sqrt{A_T^0}} = \frac{\sum_{t=1}^T (a_{t-1} - \alpha)\,\varepsilon_t}{\sqrt{A_T^0}} + \frac{\alpha - \overline{a}_T}{\sqrt{A_T^0}} \sum_{t=1}^T \varepsilon_t + O(1)\frac{1}{\sqrt{A_T^0}}\frac{1}{T}\sum_{t=1}^T \varepsilon_t. \tag{B.25}$$

The first term on the right-hand side is related to the martingale

$$M_T = \sum_{t=1}^T (a_{t-1} - \alpha)\,\varepsilon_t, \tag{B.26}$$

whose predictable quadratic variation is

$$\langle M \rangle_T = \sigma^2 A_T^0.$$

Therefore, by the central limit theorem (CLT) for martingales,

$$W_T = \frac{M_T}{\sqrt{A_T^0}} \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{B.27}$$

The corresponding Lindeberg condition is verified in Appendix B.3.3 below. Regarding the second term, (A.14) and (B.15) yield

$$\frac{\overline{a}_T - \alpha}{\sqrt{A_T^0}} = \frac{\alpha_T}{\sqrt{A_T^0}} \alpha_T^{-1} O\left(\sqrt{\frac{\log_2 T}{T}}\right) = \left(\gamma^{-1} + o_P(1)\right) O\left(\sqrt{\frac{\log_2 T}{T \log T}}\right),$$

so that by the CLT for i.i.d. sequences,

$$\frac{\alpha - \overline{a}_T}{\sqrt{A_T^0}} \sum_{t=1}^T \varepsilon_t = o_P(1). \tag{B.28}$$

The last term converges to zero by the law of large numbers. Synthesizing, we obtain

$$\frac{u_T}{\sqrt{A_T^0}} = W_T + o_P(1) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

In view of (2.2) and accounting for (B.20), this means that

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

More explicitly, using (B.15),

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

as claimed in (2.13) and (2.14) of Theorem 4.

B.3.2 Case $c < 1/2$

Reconsider again (B.25). Regarding the first term on the right-hand side, we find that, as in (B.27),

$$W_T = \frac{M_T}{\sqrt{A_T^0}} \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{B.29}$$

As for the second term, ìt follows from (A.21) and (B.22) that

$$\frac{\overline{a}_T - \alpha}{\sqrt{A_T^0}} = \frac{\sqrt{1-2c}}{1-c} \frac{1}{\sqrt{T}} (1 + o(1)).$$

Hence, by the CLT for i.i.d. sequences„

$$\frac{\overline{a}_T - \alpha}{\sqrt{A_T^0}} \sum_{t=1}^{T} \varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \frac{1-2c}{(1-c)^2}\right). \tag{B.30}$$

The last term in the brackets of (B.7) again tends to zero of course.

In view of (B.29) and (B.30), the first two terms on the right hand side of (B.7) must be treated together. It could be shown, e.g. using the Cramer-Rao device, that (B.29) and (B.30) converge jointly to a bivariate normal distribution. Yet since we are only interested in the sum

$$V_T = \frac{M_T}{\sqrt{A_T^0}} + \frac{\alpha - \overline{a}_T}{\sqrt{A_T^0}} \sum_{t=1}^{T} \varepsilon_t, \tag{B.31}$$

we proceed in a different way. By partial summation,

$$\begin{aligned} M_T &= \sum_{t=1}^{T} (a_{t-1} - \alpha) \varepsilon_t = (a_T - \alpha) \sum_{t=1}^{T} \varepsilon_t - \sum_{t=1}^{T} (a_t - a_{t-1}) \sum_{i=1}^{t} \varepsilon_i \\ &= (a_T - \alpha) \sum_{t=1}^{T} \varepsilon_t - \sum_{t=1}^{T} O\left(\frac{1}{t^{1+c}}\right) \sum_{i=1}^{t} \varepsilon_i. \end{aligned} \tag{B.32}$$

The last equality follows from (2.10), since

$$\begin{aligned} a_t - a_{t-1} &= \gamma v \left[\frac{1}{t^c} - \frac{1}{(t-1)^c}\right] (1 + o(1)) \\ &= -\gamma v c \frac{1}{t^{1+c}} (1 + o(1)). \end{aligned}$$

By the LIL, the second term in (B.32) becomes

$$\sum_{t=1}^{T} O\left(\frac{1}{t^{1+c}}\right) \sum_{i=1}^{t} \varepsilon_i = \sum_{t=1}^{T} O\left(\frac{\sqrt{\log_2 t}}{t^{1/2+c}}\right) \frac{1}{\sqrt{t \log_2 t}} \sum_{i=1}^{t} \varepsilon_i = O(1)$$

with probability one. Therefore

$$\frac{M_T}{\sqrt{A_T^0}} = \frac{a_T - \alpha}{\sqrt{A_T^0}} \sum_{t=1}^{T} \varepsilon_t + O\left(\frac{1}{T^{1/2-c}}\right),$$

and (B.31) simplifies to

$$V_T = \frac{a_T - \overline{a}_T}{\sqrt{A_T^0}} \sum_{t=1}^{T} \varepsilon_t + O\left(\frac{1}{T^{1/2-c}}\right).$$

By (B.22) together with (2.10) and (A.21),

$$\frac{a_T - \overline{a}_T}{\sqrt{A_T^0}} = -\frac{c\sqrt{1-2c}}{1-c} \frac{1}{\sqrt{T}} (1 + o(1)).$$

so that

$$V_T = -\frac{c\sqrt{1-2c}}{1-c} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t + o_P(1) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \frac{c^2(1-2c)}{(1-c)^2}\right).$$

Returning to (B.25), we thus find that

$$\frac{u_T}{\sqrt{A_T^0}} = V_T + o(1) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \frac{c^2(1-2c)}{(1-c)^2}\right)$$

or, taking account of (B.24),

$$\frac{u_T}{\sqrt{A_T}} = \frac{u_T}{\sqrt{A_T^0}} \sqrt{\frac{A_T^0}{A_T}} \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(1-2c)\right).$$

Hence, remembering that $\widehat{\beta}_T - \beta = u_T / A_T$, it follows that

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

or,

$$v\sqrt{T^{1-2c}}\left(\widehat{\beta}_T - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \frac{(1-c)^2(1-2c)^2}{c^2\gamma^2}\right). \tag{B.33}$$

This shows assertions (2.15) and (2.16) in Theorem 4, respectively.

B.3.3 On the Lindeberg condition

Reconsider the martingale in (B.26), reproduced here for convenience:

$$M_T = \sum_{t=1}^{T} (a_{t-1} - \alpha)\varepsilon_t.$$

Write M_T in the form

$$M_T = \lambda_T N_T \tag{B.34}$$

where N_T is the (square integrable) martingale difference array

$$N_T = \sum_{t=1}^{T} \xi_{Tt}\varepsilon_t, \qquad \xi_{Tt} = \frac{a_{t-1} - \alpha}{\lambda_T} \tag{B.35}$$

and λ_T is any deterministic sequence such that

$$\frac{A_T^0}{\lambda_T^2} \xrightarrow{P} v^2 \tag{B.36}$$

for some random variable v with $\mathbf{P}(v=0)=0$. Put differently, for $c>1/2$,

$$\lambda_T^2=\gamma^2\alpha_T^2=\frac{\gamma^2\sigma^2}{2c-1}\log T$$

(with $v=1$, cf. (B.15)) and, for $c<1/2$,

$$\lambda_T^2=\frac{\gamma^2}{2c-1}T^{1-2c},$$

cf. (B.22). We will show that

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

for every $\delta>0$. To this end, we make use of the elementary inequality $\{|\xi_{Tt}\varepsilon_t|>\delta\}=\{|(a_{t-1}-\alpha)\varepsilon_t|>\lambda_T\delta\}\subset\left\{|a_{t-1}-\alpha|^2>\lambda_T\delta\right\}\cup\left\{\varepsilon_t^2>\lambda_T\delta\right\}$ to obtain

$$\begin{aligned}R_T &= \sum_{t=1}^{T}\xi_{Tt}^2\mathbf{E}\left\{\varepsilon_t^2 1_{\{|\xi_{Tt}\varepsilon_t|>\delta\}}|\mathcal{F}_{t-1}\right\}\\ &\leq \frac{\sigma^2}{\lambda_T^2}\sum_{t=1}^{T}(a_{t-1}-\alpha)^2 1_{\left\{(a_{t-1}-\alpha)^2>\lambda_T\delta\right\}}+\frac{1}{\lambda_T^2}\sum_{t=1}^{T}(a_{t-1}-\alpha)^2\mathbf{E}\left\{\varepsilon_t^2 1_{\left\{\varepsilon_t^2>\lambda_T\delta\right\}}\right\}\\ &= R_T^0+R_T^1.\end{aligned}$$

Regarding R_T^0, since $a_{t-1}-\alpha\to 0$ a.s., there will be a T_0 (depending on ω) such that $(a_{t-1}-\alpha)^2\leq\lambda_T\delta$ for all $t>T_0$. Hence

$$R_T^0\leq\frac{\sigma^2}{\lambda_T^2}\sum_{t=1}^{T_0}(a_{t-1}-\alpha)^2\to 0$$

with probability one.

As to R_T^1, it follows from the Cauchy-Schwartz inequality that

$$\mathbf{E}\left\{\varepsilon_t^2 1_{\left\{\varepsilon_t^2>\lambda_T\delta\right\}}\right\}\leq\frac{1}{\lambda_T\delta}\mathbf{E}\varepsilon_t^4.$$

Hence

$$R_T^1\leq\frac{\mathbf{E}\varepsilon_t^4}{\lambda_T\delta}\frac{A_T^0}{\lambda_t^2}\xrightarrow{P}0.$$

As a consequence,

$$R_T\xrightarrow{P}0,$$

so that the Lindeberg condition is satisfied for the martingale difference array N_T. Therefore, since

$$V_T^2=\sum_{t=1}^{T}\mathbf{E}\left\{\xi_{Tt}^2\varepsilon_t^2|\mathcal{F}_{t-1}\right\}=\frac{\sigma^2}{\lambda_T^2}\sum_{t=1}^{T}(a_{t-1}-\alpha)^2=\frac{\sigma^2}{\lambda_T^2}A_T^0\xrightarrow{P}\sigma^2 v^2,$$

it follows from standard CLTs for martingale difference arrays, see e.g. Hall & Heyde (1980, Corollary 3.2), that

$$\frac{N_T}{V_T}=\frac{\lambda_T}{\sigma}\frac{N_T}{\sqrt{A_T^0}}\xrightarrow{d}\mathcal{N}(0,1)$$

or, in view of (B.34),

$$\frac{M_T}{\sigma\sqrt{A_T^0}} \xrightarrow{d} \mathcal{N}(0,1).$$

C 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

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

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

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

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

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.

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

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. (2013*a*), A note on an estimation problem in models with adaptive learning. Mimeo.

Christopeit, N. & Massmann, M. (2013*b*), Strong consistency of the least-squares estimator in simple regression models with stochastic regressors. TI Discussion Paper 12-109/III.

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

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.

Hall, P. & Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, Academic Press.

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.

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

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. (2012), Learning from inflation experience. Mimeo.

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.

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.

Supplements to "Estimating structural parameters in regression models with adaptive learning"

Norbert Christopeit
University of Bonn

Michael Massmann
Vrije Universiteit Amsterdam
and Tinbergen Institute

July 17, 2013

I Proof of equation (B.13)

I.1 Introduction

Here we give the proof of (B.13), as announced in Appendix B.2, namely that

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

for the sequence

$$\alpha_T^2=\frac{\sigma^2}{2c-1}\log T. \tag{I.2}$$

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

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

Hence, if (I.1) holds,

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

so that the normalization (I.2) comes up naturally.

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

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

I.2 Calculation of 4th moments

Henceforth, we will assumme that $s \le t$. 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

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

Here we have put

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

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

Remark 1 *Note that the B-term vanishes for $s = t$.*

I.2.1 Ad A_{st}

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

with

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

I.2.1.1 Ad A' Since

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

it follows that

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

As a consequence,

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

I.2.1.2 Ad A'' Since

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

we have that

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

Hence

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

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

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

I.2.2 Ad B_{st}

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

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

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

By (I.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,

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

But

$$
\begin{aligned}
\sum_{t=2}^{T}\frac{1}{t}\left[\log t+O(1)\right] &= \left[\int_2^T\frac{\log t}{t}dt+O(1)\right]+O(1)\left[\int_2^T\frac{dt}{t}+1\right] \\
&= \left[\frac{1}{2}\log^2 T+O(1)\right]+O(1)\left[\log T+1\right] \\
&= \frac{1}{2}\log^2 T+O(\log T)
\end{aligned}
$$

and

$$
\sum_{t=2}^{T}\frac{1}{t^{2c}}\left[t^{2c-1}+O(1)\right]=\int_2^T\frac{dt}{t}+O(1)=\log T+O(1).
$$

Hence

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

I.3 Synthesis

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

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

Or, put differently, with α_T as in (I.2),

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

This proves the assertion.

II Derivation of Remark 6

II.1 Introduction

Here we are concerned with the problem of plim-convergence of $\alpha_T^{-2} A_T^0$ in the case $c = 1/2$. In view of (B.11), which remains true for $c = 1/2$, it suffices to consider the behavior of

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

Remember that, for $c = 1/2$,

$$\xi_t = \frac{1}{\sqrt{t}} \sum_{i=1}^{t} \frac{1}{\sqrt{i}} \varepsilon_i = \frac{1}{\sqrt{t}} v_t. \tag{II.2}$$

Maintained assumption: The ε_t are i.i.d. with finite 4th moment $\mathbf{E}\varepsilon_t^4 = m_4, \mathbf{E}\varepsilon_t = 0$, $\mathrm{var}(\varepsilon_t) = \sigma^2$ and $\mathbf{E}\varepsilon_t^3 = 0$.

The last assumption (about 3rd moments) is introduced to simplify some of the calculations. It is not relevant for the results.

We will address the following two questions.

(i) Is there any deterministic sequence $\left(\alpha_T^2\right)$ *s.t.*

$$Z_T = \alpha_T^{-2} X_T \tag{II.3}$$

converges in L^2 *to some limit* Z *that is not identical to zero?*

(ii) Is there any such sequence $\left(\alpha_T^2\right)$ *and any such* Z *s.t.* $\mathrm{plim}_{T\to\infty} Z_T = Z$*?*

II.2 Ad question (i)

II.2.1 Filtering feasible sequences

By the ICT,

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

Hence

$$\mathbf{E}X_T = \frac{\sigma^2}{2}\ln^2 T + O(\ln T). \tag{II.4}$$

As a consequence, if the sequence (Z_T) should converge in L^2 to some nonvanishing square integrable random variable Z, it would follow that

$$\lim_{T\to\infty} \mathbf{E}Z_T = \lim_{T\to\infty} \frac{\mathbf{E}X_T}{\alpha_T^2} = \mathbf{E}Z = m > 0.$$

Hence any feasible choice of deterministic sequence $\left(\alpha_T^2\right)$ should satisfy

$$\frac{\alpha_T^2}{\ln^2 T} \to r \tag{II.5}$$

for some positive constant r. It therefore suffices to show that (Z_T) cannot be a Cauchy sequence in L^2 for any such sequence $\left(\alpha_T^2\right)$. Denoting

$$X_{T,N} = \left[\sum_{s=1}^{T} \xi_t^2 \sum_{t=T+1}^{T+N} \xi_t^2\right]^2,$$

we thus have to show that

$$\begin{aligned} & \sup_{N\geq 1} \mathbf{E}\left[Z_{T+N} - Z_T\right]^2 = \sup_{N\geq 1} \mathbf{E}\left[\frac{X_{T+N}}{\alpha_{T+N}^2} - \frac{X_T}{\alpha_T^2}\right]^2 \\ = & \sup_{N\geq 1} \underbrace{\left[\alpha_{T+N}^{-4}\mathbf{E}X_{T+N}^2 + \alpha_T^{-4}\mathbf{E}X_T^2 - 2\alpha_T^{-2}\alpha_{T+N}^{-2}\mathbf{E}X_T X_{T+N}\right]}_{D_{NT}} \end{aligned} \tag{II.6}$$

does not tend to zero as $T \to \infty$.

II.2.2 Fourth moments

We start as in Supplement 1 with formula (I.5) for $s \leq t$, which we repeat here for convenience:

$$C_{st} = \mathbf{E}\xi_s^2\xi_t^2 = A_{st} + B_{st}, \tag{II.7}$$

where now

$$A_{st} = \frac{1}{st}\sum_{i,j,k,l=1}^{s} \frac{1}{\sqrt{ijkl}}\mathbf{E}\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l, \tag{II.8a}$$

$$B_{st} = \frac{1}{st}\mathbf{E}\sum_{i,j=1}^{s} \frac{1}{\sqrt{ij}}\varepsilon_i\varepsilon_j \sum_{k,l=s+1}^{t} \frac{1}{\sqrt{kl}}\varepsilon_k\varepsilon_l. \tag{II.8b}$$

Remark 1 For $s = t$, the B-term vanishes.

$$\begin{aligned} A_{st} &= \frac{1}{st}\left[m_4\sum_{i=1}^{s}\frac{1}{i^2} + 6\sigma^4\sum_{i=2}^{s}\frac{1}{i}\sum_{j=1}^{i-1}\frac{1}{j}\right] \\ &= A'_{st} + A''_{st}, \end{aligned} \tag{II.9}$$

with

$$A'_{st} = \frac{m_4}{st}\sum_{i=1}^{s}\frac{1}{i^2}, \quad A''_{st} = \frac{6\sigma^4}{st}\sum_{i=1}^{s}\frac{1}{i}\sum_{j=1}^{i-1}\frac{1}{j}.$$

Ad A'

$$A'_{st} = O\left(\frac{1}{st}\right). \tag{II.10}$$

Ad A''

Since

$$\sum_{j=1}^{i-1} \frac{1}{j} = \ln i + O(1),$$

we have that

$$A''_{st} = \frac{6\sigma^4}{st} \sum_{i=1}^{s} \frac{1}{i} \left[\ln i + O(1)\right] = \frac{3\sigma^4}{st} \left[\ln^2 s + O\left(\ln s\right)\right]. \tag{II.11}$$

Putting together (II.10) and (II.11), it follows from (II.9) that

$$A_{st} = \frac{3\sigma^4}{st} \left[\ln^2 s + O\left(\ln s\right).\right] \tag{II.12}$$

In particular, by Remark 1,

$$\mathbf{E}\xi_t^4 = C_{tt} = O\left(\frac{\ln^2 t}{t^2}\right). \tag{II.13}$$

II.2.2.2 Ad B_{st} By definition,

$$\begin{aligned}
B_{st} &= \frac{1}{st}\mathbf{E} \sum_{i,j=1}^{s} \frac{1}{\sqrt{ij}} \varepsilon_i \varepsilon_j \sum_{k,l=s+1}^{t} \frac{1}{\sqrt{kl}} \varepsilon_k \varepsilon_l \\
&= \frac{1}{st}\mathbf{E} \left[\sum_{i,j=1}^{s} \frac{1}{\sqrt{ij}} \varepsilon_i \varepsilon_j \sum_{k,l=s+1}^{t} \frac{1}{\sqrt{kl}} \mathbf{E}\left\{\varepsilon_k \varepsilon_l | \mathcal{F}_s\right\} \right] \\
&= \frac{1}{st}\mathbf{E} \left[\sum_{i,j=1}^{s} \frac{1}{\sqrt{ij}} \varepsilon_i \varepsilon_j \sum_{k=s+1}^{t} \frac{\sigma^2}{k} \right] \\
&= \frac{\sigma^4}{st} \left[\sum_{i=1}^{s} \frac{1}{i}\right] \left[\sum_{k=s+1}^{t} \frac{1}{k}\right] \\
&= \frac{\sigma^4}{st} \left[\ln s + O(1)\right] \left[\ln t - \ln s + O(1)\right] \\
&= \frac{\sigma^4}{st} \left[\ln s \ln t - \ln^2 s + O(\ln t)\right].
\end{aligned} \tag{II.14}$$

Putting together (II.12) and (II.14), we obtain

$$\begin{aligned}
C_{st} &= A_{st} + B_{st} = \frac{3\sigma^4}{st} \left[\ln^2 s + O\left(\ln s\right)\right] + \sigma^4 \frac{1}{st} \left[\ln s \ln t - \ln^2 s + O(\ln t)\right] \\
&= \frac{\sigma^4}{st} \left[2\ln^2 s + \ln s \ln t + O(\ln t)\right]
\end{aligned} \tag{II.15}$$

As a consequence, since $\sum_{s=1}^{t} \left(\ln^m s\right)/s = \frac{1}{m+1} \ln^{m+1} t + O(1)$,

$$\begin{aligned}
\sum_{s=1}^{t} C_{st} &= \frac{\sigma^4}{t} \left[\frac{2}{3}\ln^3 t + \frac{1}{2}\ln^3 t + O(\ln^2 t)\right] \\
&= \frac{\sigma^4}{t} \left[\frac{7}{6}\ln^3 t + O(\ln^2 t)\right]
\end{aligned}$$

and

$$\sum_{t=2}^{T}\sum_{s=1}^{t} C_{st} = \frac{7}{24}\sigma^4 \ln^4 T + O\left(\ln^3 T\right). \tag{II.16}$$

From (II.7) it then follows that

$$\sum_{s<t}^{T} \mathbf{E}\xi_s^2\xi_t^2 = \frac{7}{24}\sigma^4 \ln^4 T + O(\ln^3 T). \tag{II.17}$$

II.2.3 Evaluating the Cauchy criterion

In view of (II.5), we evaluate (II.6) for the sequence

$$\alpha_T^2 \sim \sigma^2 \ln^2 T$$

(meaning that the quotient tends to 1 or $\alpha_T^2 = \sigma^2\left(1 + o(1)\right)\ln^2 T$).

II.2.3.1 Pure terms From (II.17) and (II.13) it follows immediately that

$$\mathbf{E}X_T^2 = \sum_{s,t=1}^{T} \mathbf{E}\xi_s^2\xi_t^2 = 2\sum_{s<t}^{T} \mathbf{E}\xi_s^2\xi_t^2 + \sum_{t=1}^{T} \mathbf{E}\xi_t^4 = \frac{7}{12}\sigma^4 \ln^4 T + O(\ln^3 T). \tag{II.18}$$

Then, as $T \to \infty$,

$$\frac{\mathbf{E}X_T^2}{\alpha_T^4} = \frac{7}{12}\frac{1}{1 + o_T(1)} + O\left(\frac{1}{\ln T}\right) = \frac{7}{12} + o_T(1). \tag{II.19a}$$

Also, as $N \to \infty$,

$$\begin{aligned} \frac{\mathbf{E}X_{T+N}^2}{\alpha_{T+N}^4} &= \frac{7}{12}\frac{1}{1 + o_{T+N}(1)} + O\left(\frac{1}{\ln\left(T+N\right)}\right) \\ &= \frac{7}{12} + o_{TN}(1). \end{aligned} \tag{II.19b}$$

II.2.3.2 Mixed term To calulate $\mathbf{E}X_T X_{T+N}$, write

$$\mathbf{E}X_T X_{T+N} = \mathbf{E}X_T^2 + \mathbf{E}X_T\left(X_{T+N} - X_T\right). \tag{II.20}$$

As to the first term,

$$\frac{\mathbf{E}X_T^2}{\alpha_T^2\alpha_{T+N}^2} = o_N(1). \tag{II.21}$$

As to the 2nd term on the right hand side of (II.20),

$$\mathbf{E}X_T\left(X_{T+N} - X_T\right) = \mathbf{E}\sum_{s=1}^{T}\xi_s^2 \sum_{t=T+1}^{T+N}\xi_t^2 = \sum_{s=1}^{T}\sum_{t=T+1}^{T+N} \mathbf{E}\xi_s^2\xi_t^2.$$

By virtue of (II.7) and (II.15),

$$\begin{aligned} \sum_{s=1}^{T}\sum_{t=T+1}^{T+N} \mathbf{E}\xi_s^2\xi_t^2 &= \sum_{s=1}^{T}\sum_{t=T+1}^{T+N} C_{st} \\ &= \sigma^4 \sum_{s=1}^{T}\frac{1}{s}\sum_{t=T+1}^{T+N}\frac{1}{t}\left[2\ln^2 s + \ln s \ln t + O_{st}(\ln t)\right]. \end{aligned}$$

But

$$\begin{aligned}\sum_{s=1}^{T}\frac{1}{s}\left[2\ln^2 s+\ln s\ln t+O_{st}(\ln t)\right] &= \frac{2}{3}\ln^3 T+\frac{1}{2}\ln^2 T\ln t+O_{Tt}\left(\ln T\ln t\right)\\ &= s_{Tt},\end{aligned}$$

and

$$\begin{aligned}\sum_{t=T+1}^{T+N}\frac{s_{Tt}}{t} &= \frac{2}{3}\ln^3 T\left[\ln\frac{T+N}{T}+O_{TN}(1)\right]+\frac{1}{4}\ln^2 T\left[\ln^2\frac{T+N}{T}+O_{TN}(1)\right]\\ &\qquad +O_{TN}(1)\ln T\left[\ln^2\frac{T+N}{T}+1\right]\\ &= \frac{1}{4}\ln^2\frac{T+N}{T}\left[\ln^2 T+O_{TN}\left(\ln T\right)\right]+\frac{2}{3}\ln\frac{T+N}{T}\ln^3 T+O_{TN}(\ln^3 T)\\ &= \frac{1}{4}\left(1+o_T(1)\right)\ln^2\frac{T+N}{T}\ln^2 T+\frac{2}{3}\ln\frac{T+N}{T}\ln^3 T+o_T\left(1\right)\\ &= :\lambda_{TN}.\end{aligned}$$

Therefore,

$$\mathbf{E}X_T\left(X_{T+N}-X_T\right)=\sigma^4\lambda_{TN}.$$

Simpler representation of λ

$$\begin{aligned}\ln\frac{T+N}{T} &= \ln\left(T+N\right)-\ln T,\\ \ln\frac{T+N}{T}\ln^3 T &= \ln\left(T+N\right)\ln^3 T-\ln^4 T,\\ \ln^2\frac{T+N}{T} &= \ln^2\left(T+N\right)-2\ln(T+N)\ln T+\ln^2 T,\\ \ln^2\frac{T+N}{T}\ln^2 T &= \ln^2\left(T+N\right)\ln^2 T-2\ln(T+N)\ln^3 T+\ln^4 T.\end{aligned}$$

Hence

$$\begin{aligned}\lambda &= \frac{1}{4}\ln^2\left(T+N\right)\ln^2 T\left[1-2\frac{\ln T}{\ln\left(T+N\right)}+\frac{\ln^2 T}{\ln^2\left(T+N\right)}\right]\left(1+o_T(1)\right)\\ &\qquad +\frac{2}{3}\ln^2\left(T+N\right)\ln^2 T\left[\frac{\ln T}{\ln\left(T+N\right)}-\frac{\ln^2 T}{\ln^2\left(T+N\right)}\right]+o_T\left(1\right)\\ &= \ln^2\left(T+N\right)\ln^2 T\left[\frac{1}{4}-\left(\frac{1}{2}-\frac{2}{3}\right)\frac{\ln T}{\ln\left(T+N\right)}\right.\\ &\qquad\qquad \left.+\left(\frac{1}{4}-\frac{2}{3}\right)\frac{\ln^2 T}{\ln^2\left(T+N\right)}\right]+o_T\left(1\right)\\ &= \ln^2\left(T+N\right)\ln^2 T\left[\frac{1}{4}+\frac{1}{6}\frac{\ln T}{\ln\left(T+N\right)}-\frac{5}{12}\frac{\ln^2 T}{\ln^2\left(T+N\right)}\right]+o_T\left(1\right)\end{aligned}$$

and

$$\begin{aligned}\frac{\sigma^4\lambda_{TN}}{\alpha_T^2\alpha_{T+N}^2} &= \left[\frac{1}{4}+\frac{1}{6}\frac{\ln T}{\ln\left(T+N\right)}-\frac{5}{12}\frac{\ln^2 T}{\ln^2\left(T+N\right)}\right]\left(1+o_{TN}(1)\right)\\ &= \frac{1}{4}+o_N(1).\end{aligned}$$

Hence

$$\frac{\mathbf{E}X_T\left(X_{T+N}-X_T\right)}{\alpha_T^2\alpha_{T+N}^2}=\frac{1}{4}+o_N(1). \tag{II.22}$$

II.2.3.3 Synthesis Going back to (II.6) and making use of (II.19) and (II.21),

$$\begin{aligned} D_{NT} &= \frac{\mathbf{E}X_{T+N}^2}{\alpha_{T+N}^4}+\frac{\mathbf{E}X_T^2}{\alpha_T^4}-2\frac{\mathbf{E}X_T^2}{\alpha_T^2\alpha_{T+N}^2}-2\frac{\mathbf{E}X_T\left(X_{T+N}-X_T\right)}{\alpha_T^2\alpha_{T+N}^2} \\ &= \frac{7}{6}-\frac{1}{2}+o_T(1)+o_N(1)+o_{TN}(1) \\ &= \frac{2}{3}+o_T(1)+o_N(1)+o_{TN}(1). \end{aligned}$$

As a consequence, letting $N\to\infty$ for fixed T,

$$\sup_{N\geq 1} D_{NT}\geq\frac{2}{3}+o_T(1)$$

and hence

$$\lim_{T\to\infty}\sup_{N\geq 1} D_{NT}\geq\frac{2}{3}.$$

Therefore with regard to question (i), we arrive at the following

Conclusion 2 *Z_T does not converge in L^2 to a nonzero limit for any choice of deterministic normalizing sequence α_T.*

II.3 Ad question (ii)

II.3.1 Line of argument

We show: $\left(Z_T^2\right)$ is uniformly integrable (ui).

To this end, we show that

$$\sup_T \mathbf{E}Z_T^4=\sup_T\frac{\mathbf{E}X_T^4}{\alpha_T^8}<\infty. \tag{II.23}$$

Once this is established, one may argue as follows.

Step 1. Assume that $\text{plim}_{T\to\infty}Z_T=Z$. Since$\left(Z_T^2\right)$ is ui, convergence also holds in L^2: $Z_T\overset{L^2}{\to}Z$.

Step 2. Taking account of Conclusion 1**,** we arrive at

Conclusion 3 *Z_T does not converge in probability to any nonzero limit.*

For Step 1, we refer to the following basic result of dominated convergence type, cf. Schürger (1998, Kapitel 4, Satz 6.5) and Shiryaev (1996, Chapter II, §6, Theorem 4).

Proposition *Let (x_T) be a sequence of random variables in L^r s.t. $x_T\overset{P}{\to}x$. Then $x_T\overset{L^r}{\to}x$ if and only if $(|x_T|^r)$ is ui.*

Proof. If $x_T \overset{L^r}{\to} x$, then also $x \in L^r$, and the assertion follows from Schürger (1998). On the other hand, $x_T \overset{P}{\to} x$ implies that $|x_T|^r \overset{P}{\to} |x|^r$. If $(|x_T|^r)$ is ui, then $|x_{T'}|^r \to |x|^r$ a.s. for some subsequence (T'). Hence, by Shiryaev (1996), $|x|^r$ is integrable, i.e. $x \in L^r$. Again by Schürger (1998), it then follows that $x_T \overset{L^r}{\to} x$. ■

In our context, this is applied to $x_T = Z_T$ and $r = 2$.

II.3.2 Bounding $\mathbf{E}X_T^4$

By (II.1),

$$\mathbf{E}X_T^4 = \sum_{q,r,s,t=1}^{T} \mathbf{E}\xi_q^2\xi_r^2\xi_s^2\xi_t^2. \tag{II.24}$$

By the (extended) Hölder inequality (cf. Schürger (1998)), with $p(i) = 4$,

$$\mathbf{E}\xi_q^2\xi_r^2\xi_s^2\xi_t^2 \leq \left[\mathbf{E}\xi_q^8\mathbf{E}\xi_r^8\mathbf{E}\xi_s^8\mathbf{E}\xi_t^8\right]^{1/4}.$$

Hence, by (II.24),

$$\mathbf{E}X_T^4 \leq \sum_{q,r,s,t=1}^{T} \left[\mathbf{E}\xi_q^8\right]^{1/4}\left[\mathbf{E}\xi_r^8\right]^{1/4}\left[\mathbf{E}\xi_r^8\right]^{1/4}\left[\mathbf{E}\xi_t^8\right]^{1/4}. \tag{II.25}$$

Note that

$$\mathbf{E}\xi_t^8 = \frac{1}{t^4}\sum_{i,j,k,l=1}^{t} \frac{1}{\sqrt{i_1\cdots i_8}}\mathbf{E}\left(\varepsilon_{i_1}\cdots\varepsilon_{i_8}\right) = \frac{1}{t^4}S_t. \tag{II.26}$$

II.3.2.1 Calculating 8-th moments

$$\mathbf{E}\left(\varepsilon_{i_1}\cdots\varepsilon_{i_8}\right) = \begin{cases} \text{moment} & \text{code} & \text{number} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ m_8 & 8 & t \\ \sigma^2 m_6 & 6{:}2 & \binom{8}{2}t\left(t-1\right) \\ m_4^2 & 4{:}4 & \binom{8}{4}t\left(t-1\right) \\ \sigma^4 m_4 & 4{:}2{:}2 & \binom{8}{4}\binom{4}{2}t\left(t-1\right)\left(t-2\right) \\ \sigma^8 & 2{:}2{:}2{:}2 & \binom{8}{2}\binom{6}{2}\binom{4}{2}t\left(t-1\right)\left(t-2\right)\left(t-3\right) \end{cases}.$$

Codes

$8 \equiv$ all i's equal

$6{:}2 \equiv 6$ i's equal, with the remaining 2 equal and different from the other 6

$4{:}4 \equiv 2$ groups à 4, equal within each group, different between groups

$4{:}2{:}2 \equiv 3$ groups, equal within each group, different between groups

$2{:}2{:}2{:}2 \equiv 4$ groups, equal within each group, different between groups

dominant term: code 2:2:2:2 with number $\sim t^4$

Hence

$$\begin{aligned} S_t &= m_8 \sum_{i=1}^{t} \frac{1}{i^4} + \sigma^2 m_6 \sum_{6:2} \frac{1}{i^3 j} + m_4^2 \sum_{4:4} \frac{1}{i^2 j^2} \\ &\quad + \sigma^4 m_4 \sum_{4:2:2} \frac{1}{i^2 jk} + \sigma^4 m_4 \sum_{\substack{i,j,k,l=1 \\ i \neq j \neq k \neq l}}^{t} \frac{1}{ijkl} \end{aligned}$$

The single terms

Code 8

$$\sum_{i=1}^{t} \frac{1}{i^4} = O(1)$$

Code 6:2

$$\sum_{6:2} \frac{1}{ij^3} = \binom{8}{2} \sum_{i=1}^{t} \frac{1}{i} \sum_{j \neq i} \frac{1}{j^3} = O\left(\ln t\right)$$

Code 4:4

$$\sum_{4:4} \frac{1}{i^2 j^2} = \binom{8}{4} \sum_{i=1}^{t} \frac{1}{i^2} \sum_{j \neq i} \frac{1}{j^2} = O\left(\ln^2 t\right)$$

Code 4:2:2

$$\sum_{4:2:2} \frac{1}{ijk^2} = \binom{8}{4} \binom{4}{2} \sum_{i=1}^{t} \frac{1}{i} \sum_{j \neq i} \frac{1}{j} \sum_{k \neq i \& j} \frac{1}{k^2} = O\left(\ln^2 t\right)$$

Code 2:2:2:2

$$\begin{aligned} \sum_{2:2:2:2} \frac{1}{ijk^2} &= \binom{8}{2} \binom{6}{2} \binom{4}{2} \sum_{i=1}^{t} \frac{1}{i} \sum_{j \neq i} \frac{1}{j} \sum_{k \neq i \& j} \frac{1}{k} \sum_{k \neq i \& j \& l} \frac{1}{l} \\ &= O\left(\ln^4 t\right) \end{aligned}$$

As a consequence, from (II.26),

$$\mathbf{E}\xi_t^8 = O\left(\frac{\ln^4 t}{t^4}\right)$$

and

$$\left[\mathbf{E}\xi_t^8\right]^{1/4} = O\left(\frac{\ln t}{t}\right).$$

Hence, by (II.25),

$$\begin{aligned} \mathbf{E}X_T^4 &\leq O(1) \sum_{q,r,s,t=1}^{T} \frac{\ln q}{q} \frac{\ln r}{r} \frac{\ln s}{s} \frac{\ln t}{t} \\ &= O(1) \left[\sum_{t=1}^{T} \frac{\ln t}{t}\right]^4. \end{aligned}$$

But

$$\sum_{t=1}^{T} \frac{\ln t}{t} = \frac{1}{2} \ln^2 T + O(1),$$

so that

$$\mathbf{E}X_T^4 = O\left(\ln^8 T\right). \tag{II.27}$$

II.3.3 Finale furioso

Putting together (II.27) and (II.5), the desired result (II.23) is obvious.

References

Schürger, K. (1998), *Wahrscheinlichkeitstheorie*, Oldenbourg.

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