

TI 2018-045/III  
Tinbergen Institute Discussion Paper



# Strong consistency of the least squares estimator in regression models with adaptive learning

Norbert Christopeit<sup>1</sup>  
Michael Massmann<sup>2</sup>

1: University of Bonn  
2: VU Amsterdam

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

Contact: [discussionpapers@tinbergen.nl](mailto:discussionpapers@tinbergen.nl)

More TI discussion papers can be downloaded at the [Tinbergen Site](#)

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 598 4580

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

# Strong consistency of the least squares estimator in regression models with adaptive learning

Norbert Christopeit\*  
University of Bonn

Michael Massmann\*,†  
WHU – Otto Beisheim School of Management  
and  
Vrije Universiteit Amsterdam

28<sup>th</sup> November, 2017

## Abstract

This paper looks at the strong consistency of the ordinary least squares (OLS) estimator in a stereotypical macroeconomic model with adaptive learning. It is a companion to Christopeit & Massmann (2017, *Econometric Theory*) which considers the estimator's convergence in distribution and its weak consistency in the same setting. Under constant gain learning, the model is closely related to stationary, (alternating) unit root or explosive autoregressive processes. Under decreasing gain learning, the regressors in the model are asymptotically collinear. The paper examines, first, the issue of strong convergence of the learning recursion: It is argued that, under constant gain learning, the recursion does not converge in any probabilistic sense, while for decreasing gain learning rates are derived at which the recursion converges almost surely to the rational expectations equilibrium. Secondly, the paper establishes the strong consistency of the OLS estimators, under both constant and decreasing gain learning, as well as rates at which the estimators converge almost surely. In the constant gain model, separate estimators for the intercept and slope parameters are juxtaposed to the joint estimator, drawing on the recent literature on explosive autoregressive models. Thirdly, it is emphasised that strong consistency is obtained in all models although the near-optimal condition for the strong consistency of OLS in linear regression models with stochastic regressors, established by Lai & Wei (1982), is not always met.

keywords: adaptive learning, non-stationary regression, ordinary least squares, almost sure convergence

JEL codes: C22, C51, D83

---

\*The authors would like to thank seminar participants at the University of Bonn and at the 6th RMSE Meeting at Erasmus University Rotterdam.

†corresponding author: Michael Massmann, WHU – Otto Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany; phone: +49 (0)261 6509370; email: [michael.massmann@whu.edu](mailto:michael.massmann@whu.edu)

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Asymptotic behaviour of <math>a_t</math></b>	<b>6</b>
2.1	Constant gain . . . . .	6
2.2	Decreasing gain . . . . .	7
<b>3</b>	<b>Strong consistency of the OLS estimator</b>	<b>8</b>
3.1	Constant gain . . . . .	8
3.1.1	Separate estimation of the parameters . . . . .	9
3.1.2	Joint estimation of the parameters . . . . .	11
3.1.3	Comparison . . . . .	13
3.2	Decreasing gain . . . . .	13
3.3	Comparison with Lai & Wei . . . . .	14
<b>4</b>	<b>Conclusion</b>	<b>15</b>
<b>A</b>	<b>Proof of Theorem 1</b>	<b>17</b>
A.1	$c > 1/2$ . . . . .	17
A.1.1	$c > 1$ . . . . .	17
A.1.2	$1/2 < c \leq 1$ . . . . .	18
A.2	$c = 1/2$ . . . . .	19
A.3	$c < 1/2$ . . . . .	20
<b>B</b>	<b>Proof of Theorems 2 and 3</b>	<b>21</b>
B.1	Prerequisites . . . . .	21
B.1.1	Path properties . . . . .	22
B.1.2	Eigenvalues of the moment matrix . . . . .	28
B.1.3	Generalities . . . . .	30
B.2	Proof of Theorem 2 . . . . .	31
B.2.1	Estimation of the slope . . . . .	31
B.2.2	Estimation of the intercept . . . . .	34
B.2.3	Stable case revisited . . . . .	35
B.3	Proof of Theorem 3 . . . . .	36
B.3.1	Stable case . . . . .	37
B.3.2	Explosive case . . . . .	37
B.3.3	Unit root case . . . . .	37
B.4	Comparison with Lai & Wei . . . . .	39
B.4.1	Stable case . . . . .	39
B.4.2	Explosive case . . . . .	39
B.4.3	Unit root case . . . . .	40
<b>C</b>	<b>Proof of Theorem 4</b>	<b>40</b>
C.1	Generalities . . . . .	40
C.2	Asymptotics of $A_T^0$ . . . . .	42
C.2.1	$c > 1/2$ . . . . .	42
C.2.2	$c < 1/2$ . . . . .	44
C.3	Consistency . . . . .	44
C.3.1	$c > 1/2$ . . . . .	44
C.3.2	$c < 1/2$ . . . . .	45
C.4	Comparison with Lai & Wei . . . . .	45
C.4.1	$c > 1/2$ . . . . .	46
C.4.2	$c < 1/2$ . . . . .	46

# 1 Introduction

This paper looks at the strong consistency of the ordinary least squares (OLS) estimator in a stereotypical macroeconomic model with adaptive learning. It is a companion to Christopheit & Massmann (2017), hereafter referred to as CM17, which considers the estimator's convergence in distribution and its *weak* consistency in the same setting. The following paragraphs serve as macroeconomic motivation of the econometric model considered in this paper. They draw in parts on the introductory remarks in CM17, which is recommended for more details.

There is a long tradition in macroeconomics to explain an observed variable of interest  $y_t$  by an expectational term as well as by so-called driving variables. To fix ideas, consider the following model class:

$$y_t = \beta y_{t|t-1}^e + \delta x_t + \varepsilon_t, \quad t = 1, 2, \dots \quad (1)$$

where  $y_{t|t-1}^e$  denotes agents' expectations about  $y_t$  based on the information available at time  $t - 1$ , the observed driving variable  $x_t$  is exogenous, and  $\varepsilon_t$  denotes an error term. Many seminal models fit into this class: the classical cobweb model, see e.g. Bray & Savin (1986), the aggregate supply model by Lucas (1973), and the New Keynesian Phillips curve (NKPC), cf. Clarida, Galí & Gertler (2000).

In much of the literature the expectational term  $y_{t|t-1}^e$  is considered unobserved. It is traditionally specified via rational expectations, an approach that is now, however, frequently seen as imputing too much knowledge to economic agents. The alternative assumption of agents being boundedly rational and forming their expectations via adaptive learning has recently gained popularity amongst macroeconomists, see Sargent (1993, 1999) and Evans & Honkapohja (2001). The basic idea underlying all adaptive learning procedures is that agents employ an auxiliary model, or so-called perceived law of motion, to form expectations. One way to specify this auxiliary model is to posit that its functional form corresponds to the rational expectations equilibrium  $y_t = \alpha x_t + \varepsilon_t$ , where  $\alpha = \delta/(1 - \beta)$ , and that agents replace the unknown parameter  $\alpha$  by some estimate  $a_{t-1}$  based on information  $\mathcal{F}_{t-1}$ . Typically, they are assumed to estimate the parameter  $\alpha$  by a recursive procedure of the form

$$a_t = a_{t-1} + \gamma_t \frac{x_t}{r_t} (y_t - a_{t-1} x_t) \quad (2a)$$

$$r_t = r_{t-1} + \gamma_t (x_t^2 - r_{t-1}), \quad (2b)$$

where  $\gamma_t$  is some weighting, or gain, sequence. This updating algorithm can be viewed as generalising the recursive least squares estimator of  $\alpha$ , which has  $\gamma_t = 1/t$  and whose  $r_t$  is the sample second moment of  $x_t$ . At the same time, (2) is a special case of a stochastic approximation algorithm, see Lai (2003) and Kushner (2010) for an overview of this large literature. With this learning scheme, 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 (2) coupled with

$$y_t = \beta a_{t-1} x_t + \delta x_t + \varepsilon_t. \quad (3)$$

The purpose of the present paper is to examine whether the OLS estimators of the parameters  $\beta$  and  $\delta$  in structural models such as (3), with  $a_t$  given in (2), are strongly consistent. This is of interest since the empirical estimation of learning models of this class has 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 (2013), the inflation model by Malmendier & Nagel (2016) and the model of stock market volatility by Adam, Marcet & Nicolini (2016). Yet the econometrics of adaptive learning models is still in its infancy and it is not in general clear on which econometric principles these empirical implementations are built. Our companion paper CM17 is one of the first

comprehensive attempts to examine the asymptotic behaviour of an econometric estimation procedure in models such as (2)-(3). In particular, for the same special case of the model that is also considered in the present paper, the authors show that the OLS estimator of the model parameters  $\delta$  and  $\beta$  remains weakly consistent for most parameter constellations, although the estimator's asymptotic distribution is highly non-standard in almost all settings. One of the main conclusions of that research is hence that care must be taken when inference in models with adaptive learning is conducted.

In the present paper, the issue of almost sure (a.s.) convergence of the OLS estimator of  $\beta$  and  $\delta$  is investigated. Apart from being of theoretical interest in its own right, a.s. convergence is also indispensable for the analysis of the long-run behaviour of trajectories.

Before the OLS estimators can be analysed the asymptotic behaviour of  $a_t$  needs to be examined. The latter, in turn, is contingent on the specification of the gain sequence  $\gamma_t$ . This paper considers two different types of gain: a *constant* gain, i.e.  $\gamma_t = \gamma > 0$ , and a *decreasing* gain sequence, such that  $\gamma_t \searrow 0$ . The choice of gain sequence has immediate repercussions on the dynamics of  $a_t$ , see e.g. Benveniste, Métivier & Priouret (1990). In particular, with constant gain learning, it is well-known that  $a_t$  does not in general converge to  $\alpha$ . As opposed to that, with decreasing gain learning, the convergence  $a_t \rightarrow \alpha$  does hold with probability one under suitable summability assumptions on  $\gamma_t$ , see below. These two gain specifications are popular in the macroeconomics literature due to their intuitive appeal, for a decreasing gain is tantamount to assuming that agents end up being rational in the limit while they learn *ad infinitum* under constant gain. Moreover, the estimation of  $\alpha$  based on a constant gain recursion is equivalent to a forecast formed by exponential smoothing. This procedure is frequently argued to be plausible for boundedly rational agents, in the same way as is a recursive least squares forecast constructed with the decreasing gain sequence  $\gamma_t = 1/t$ . Irrespective of the economic interpretation, a central ingredient to our analysis of the OLS estimator will be not the mere convergence of  $a_t$  but rather the *rate* at which it converges, if indeed it does. Section 2 therefore contains a comprehensive treatment of this issue.

Reconsider the structural equation in (1). Most models in the macroeconomics literature presume that  $\beta \in (0, 1)$ , yet there are some that also 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. The gain sequence  $\gamma_t$  in (2) is specified to be

$$\gamma_t = \begin{cases} \gamma \\ \gamma/t \end{cases} \quad (4)$$

for constant and decreasing gain learning, respectively, 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  $\alpha$  is  $\sum_t \gamma_t = \infty$  but  $\sum_t \gamma_t^2 \ln^2 t < \infty$ , provided that  $\beta < 1$ , cf. Kottmann (1990) and Christopheit & Massmann (2010) for details. These conditions would in fact suggest considering sequences  $\gamma_t = \gamma/t^\eta$ , with  $\eta \in (1/2, 1]$ . 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  $\eta$ .

With the  $a_t$  given by (2), the structural equation in (3) is a linear regression model with *pre-determined stochastic regressors*. In the 1950s and 60s research on this model was a byproduct of more general investigations in adaptive systems control theory. From the 1970s onwards, it has become a subject in its own right and an extensive literature has grown around many statistical issues pertaining to such models. An excellent account of the developments is given in Lai (2003). Besides asymptotic distribution theory for various estimators, their a.s. or *strong* convergence played a prominent part. One focus was on the performance of the OLS estimators of  $\beta$  and  $\delta$ , at least for models in which autocorrelation of the error terms was no issue. The results that, to our knowledge, still represent the current state of the art for the strong convergence of

the OLS estimator are those by Lai & Wei (1982a); but see also Lai & Wei (1982b), Wei (1985) and Christopheit (1986), the latter for the general semimartingale model. As to be expected, the sufficient conditions for models with stochastic regressors are more restrictive than those for deterministic regressors. A brief account of these results is given in Christopheit & Massmann (2012). Concerning our model, it will turn out that for some cases of both constant and decreasing gain learning the near optimal sufficient condition established in Lai & Wei (1982a) is not satisfied. Nevertheless strong consistency of the OLS estimators of  $\beta$  and  $\delta$  always obtains.

For constant gain learning, the model to be estimated is basically an autoregressive model of order one with a constant term. There is a vast literature on the strong consistency of the OLS estimator in general autoregressive models. Most of it, however, considers models without an intercept, cf., e.g., Lai & Wei (1983) and Lai & Wei (1985). As will be seen, the existence of an unknown intercept can make a considerable difference to the analysis. In the past, autoregressive models with intercept have frequently been treated within the framework of identification and control of dynamic systems, namely as input-output systems with a single constant input, cf. Lai & Wei (1982a, Section 3). This framework could be used for establishing a.s. convergence rates for the norm of the *joint*, i.e. bivariate, OLS estimator of  $\theta = (\beta, \delta)$ . Since, however, the theory for input-output systems allows for a variety of different input classes, such as feedback controls, one cannot expect to obtain “optimal” results for the model of interest. More recently, Nielsen (2005) presents a thorough investigation of a class of vector autoregressive models with deterministic terms in its own right. In addition, that class encompasses stationary, unit root as well as explosive models. It is therefore the approach by Nielsen that is used below for deriving the speed of convergence of the norm of the *joint* OLS estimator. Yet, in general, the convergence speed of the slope estimator can be expected to be different from that of the intercept. To capture that difference, the rate of convergence of the *separate* OLS estimators of  $\beta$  and  $\delta$  is also established in this paper. The convergence rates of both joint and separate estimators are presented and compared in Section 3.1.

For decreasing gain, the asymptotic second moment matrix is singular. This is due to the fact that the regressor  $a_t$  converges a.s. to the constant  $\alpha$ . This violation of the Grenander condition, cf. Grenander & Rosenblatt (1957), may affect the rates of weak convergence of the OLS estimator, see Phillips (2007) and CM17. Yet it does not pose any problem if one is concerned with a.s. convergence. The strong consistency of the OLS estimators in the decreasing gain model is derived in Section 3.2.

In order to keep the analysis tractable, a simple special case of (2)-(3) is considered in this paper: it is assumed that  $x_t = x$  is a constant. The resulting toy model will lay bare the difficulties arising from the feedback of the expectation formation process to the structural equation. When the regressors are constant,  $r_t = x^2$  is the stationary solution of (2b) such that, for any starting value  $r_0$ ,  $\lim_{t \rightarrow \infty} r_t = x^2$ . 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  $\varepsilon_t$  be changed from, say,  $\sigma^2$  to  $\sigma^2/x^2$ . Working with this stationary solution for  $r_t$ , the recursion in (2) reduces to  $a_t = a_{t-1} + \gamma_t (y_t - a_{t-1})$  while the structural equation in (3) becomes

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t. \quad (5)$$

Substituting  $y_t$  into  $a_t$ , the dynamics of  $a_t$  may also be written as

$$a_t = (1 - c_t) a_{t-1} + \gamma_t (\delta + \varepsilon_t), \quad (6)$$

where we have defined  $c_t = (1 - \beta) \gamma_t$ . With our choice of  $\gamma_t$  in (4),  $c_t$  becomes

$$c_t = \begin{cases} c \\ c/t \end{cases} \quad (7)$$

where

$$c = (1 - \beta)\gamma. \quad (8)$$

Recall that, for decreasing gain,  $\beta < 1$  and any  $\gamma > 0$  is admissible. As a consequence,  $c > 0$ . For constant gain, as opposed to that,  $\gamma \in (0, 1)$  while  $\beta$  and, consequently,  $c$  may take any value.

Putting the setup in a nutshell, the model we consider in this paper is (5)-(6), with  $\gamma_t$  and  $c_t$  given in (4) and (7), respectively. The parameters estimated by OLS are  $\beta$  and  $\delta$ , the value of  $\gamma$  is assumed known. We also make the following maintained assumption:

**Maintained assumption.** *The  $\varepsilon_t$  are independently and identically distributed (i.i.d.) with mean 0 and variance  $\sigma^2$ .*

To summarise our findings, this paper is the first to consider the question of strongly consistent OLS estimation in regression models with adaptive learning. In particular, it makes three main contributions to the literature: First, we derive the rates at which  $a_t$  converges almost surely to  $\alpha$  in the decreasing gain learning model, i.e. the almost sure rates at which agents learn to be rational. To the best of our knowledge, this has not been attempted before. Secondly, we establish the strong consistency of the OLS estimators of  $\delta$  and  $\beta$  in the constant and decreasing gain learning models as well as the rates at which they converge almost surely. Again, this is a novel undertaking and will be of use when interest lies on the long-run behaviour of trajectories. Thirdly, we relate our results to the available literature. The comparison of our convergence rates in the constant gain setting to those derived by Nielsen (2005) is of particular interest. Moreover, it is instructive to find in passing that the aforementioned near optimal sufficient condition by Lai & Wei (1982a) is not satisfied in some of our models.

It will be plain that the analysis of this model is far from trivial. The obvious extension to models with time-varying regressor sequences  $x_t$  is therefore beyond the scope of the present paper. Similarly, assuming agents' expectations in the structural equation (1) to be forward-looking, i.e. replacing  $y_{t|t-1}^e$  by  $y_{t+1|t}^e$ , introduces economic as well as econometric complications and is left to future research. Lastly, regarding  $\gamma$  to be unknown changes the structure of the setup fundamentally since it would make the model nonlinear in the parameters. The repercussions likely to arise from these extensions are discussed in CM17.

The outline of the paper is as follows: The asymptotics of  $a_t$  for the model in (5)-(6) are examined in Section 2, both for constant and decreasing gain learning. Subsequently, the strong consistency of the OLS estimators of  $\beta$  and  $\delta$  is proved in Section 3, again for both learning types. Our results based on the alternative approach inspired by the work by Nielsen (2005) is also contained in Section 3. Section 4 concludes. The proofs of all theorems and corollaries are relegated to Appendices A-C.

All convergence statements are of the almost sure (a.s.) type unless otherwise indicated.

## 2 Asymptotic behaviour of $a_t$

### 2.1 Constant gain

In this section, we consider the asymptotic behaviour of  $a_t$  under the assumption that agents employ a constant gain learning algorithm to produce their forecasts. Put differently, the model under examination is (5)-(6) while the constant gain sequence in (4) is  $\gamma_t = \gamma$  such that  $c_t = c$  in (7). As a result, the dynamics of  $a_t$  can be written as

$$a_t = (1 - c)a_{t-1} + \gamma(\delta + \varepsilon_t) \quad (9)$$

where  $c = (1 - \beta)\gamma$  as in (8).

It is well-known in the literature that constant gain recursions do not in general converge to the rational expectations equilibrium (REE). In particular, the precise limiting behaviour of  $a_t$  as given in (9) is derived in Theorem 1 of CM17 and depends crucially on parameter  $c$  since,



- (i) if  $0 < c < 2$ , the process  $a_t$  is a stable autoregression,
- (ii) if  $c = 0$ ,  $a_t$  follows a random walk with drift while, if  $c = 2$ , it follows an alternating random walk with drift,
- (iii) if  $c < 0$  or  $c > 2$ ,  $a_t$  in (9) is an explosive autoregressive process.

Seminal papers on autoregressive processes that CM17 appeal to and extend are 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.

The following reproduces Theorem 1 of CM17 for convenience.

**Result 1 (Christopeit & Massmann (2017, Theorem 1)).**

- (i) If  $0 < c < 2$  then  $a_t$  converges in distribution to the law of the stationary solution, i.e. to the invariant distribution. This is nondegenerate with mean  $\alpha$  and positive variance.
- (ii) If  $c = 0$  then  $a_t$  is a random walk with drift  $\delta\gamma$  and

$$a_t = \gamma\delta t + o(t) \text{ a.s.}$$

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

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

- (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  $\mathbf{E}a_0 - \alpha$ .

Clearly, for no value of  $c$ , and hence for no combination of  $\beta \in \mathbb{R}$  and  $\gamma \in (0, 1)$ , does  $a_t$  converge to the REE  $\alpha$  in any probabilistic sense. Agents will thus not be rational in the limit but learn ad infinitum.

## 2.2 Decreasing gain

Consider now the model under decreasing gain, i.e.  $\gamma_t = \gamma/t$  in (4) and  $c_t = c/t$  in (7). Consequently, the recursion of  $a_t$  in (6) becomes

$$a_t = \left(1 - \frac{c}{t}\right) a_{t-1} + \frac{\gamma}{t} (\delta + \varepsilon_t) \tag{10}$$

where, again,  $c = (1 - \beta)\gamma$ , see (8).

As mentioned in the introduction, for  $\beta < 1$  and  $\gamma > 0$ , the mere convergence of  $a_t$  to  $\alpha$  follows easily from well-known results on recursive algorithms. However, for our analysis of the strong consistency of the OLS estimator in Section 3, we will need the exact rates of convergence of  $a_t$ .

Note that the dynamics of  $a_t$  in (10) are highly nonstandard: First,  $a_t$  is autoregressive of first order with a time-varying coefficient that is intrinsically local-to-unity. The behaviour of models of this kind has been analysed by, for instance, Phillips (1987) and Phillips & Magdalinos (2007). Secondly, the impact of the intercept  $\delta$  and of the disturbance  $\varepsilon_t$  on  $a_t$  tends to zero for large  $t$ . In the limit,  $a_t$  is thus constant. Finally,  $a_t$  is generated by what Solo & Kong (1995) call a long memory algorithm.

It is shown in Theorem 1 below that  $a_t$  converges almost surely to the REE  $\alpha$  for all combinations of  $\beta$  and  $\gamma$ . The rates of convergence are, however, different for the three regimes  $c > 1/2$ ,  $c = 1/2$  and  $c < 1/2$ . The proof of the theorem is presented in Appendix A.

**Theorem 1.** For decreasing gain with gain sequence  $\gamma_t = \gamma/t$ , strong consistency of  $a_t$  holds at the following rates.

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

$$\limsup_{t \rightarrow \infty} \sqrt{\frac{t}{\ln_2 t}} |a_t - \alpha| = \sigma \gamma \sqrt{\frac{2}{2c - 1}}.$$

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

$$\limsup_{t \rightarrow \infty} \sqrt{\frac{t}{\ln t \ln_3 t}} |a_t| = \sigma \gamma \sqrt{2}.$$

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

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

$u$  has a continuous distribution function.

It is plain that, as  $c$  decreases, the convergence of  $a_t$  to  $\alpha$  gets progressively slower. The value  $c = 1/2$  can be interpreted as a boundary separating ‘good’ from ‘poor’ asymptotic behaviour of  $a_t$ , in the sense of speed of convergence. For an intuition of this boundary, the reader is referred to the exposition in CM17. The value of  $1/2$  indeed figures prominently in the context of weak convergence of stochastic approximation algorithms, see the results in Benveniste, Métivier & Priouret (1990, Theorem 3 on p. 11 and Theorem 13 on p. 332) which, in turn, is used by Marcet & Sargent (1995) and Evans & Honkapohja (2001). The threshold of  $1/2$  is also reminiscent of a similar boundary discussed in Evans, Honkapohja, Sargent & Williams (2013).

It is of interest to compare the convergence rates in Theorem 1 with those valid for weak convergence of  $a_t - \alpha$ , cf. Theorem 3 in CM17. For  $c > 1/2$ , the additional ‘path taming’ sequence  $(\ln_2 t)^{-1/2}$  corresponds to the passage from a central limit theorem (CLT) to a law of the iterated logarithm (LIL). As to be expected, this ‘path taming’ sequence is slower, namely  $(\ln_3 t)^{-1/2}$ , for  $c = 1/2$ . For  $c < 1/2$ , all convergence concepts coincide.

### 3 Strong consistency of the OLS estimator

#### 3.1 Constant gain

In this section we are concerned with the OLS estimation of  $\beta$  and  $\delta$  in

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t \tag{11}$$

under constant gain learning. As argued in Section 2.1,

$$a_t = (1 - c) a_{t-1} + \gamma (\delta + \varepsilon_t) \tag{12}$$

does not converge to the REE for any value of  $c = (1 - \beta)\gamma$ . There is hence no issue of asymptotic collinearity in (11)-(12).

It is shown in Section 2.2 of CM17 that the OLS estimator  $\hat{\theta}_T = (\hat{\delta}_T, \hat{\beta}_T)'$  of  $\theta = (\delta, \beta)'$  in (11) is, up to a constant of proportionality, equal to that of  $\theta^* = (\delta^*, \beta^*)'$  in the autoregressive model

$$a_t^* = \delta^* + \beta^* a_{t-1}^* + \gamma \varepsilon_t, \tag{13}$$

provided that  $\delta^* = \gamma\delta$  as well as  $\beta^* = 1 - c$  and the initial values of the two sequences  $a_t$  and  $a_t^*$  are the same. Put differently,

$$\hat{\theta}_T - \theta = \gamma^{-1} (\hat{\theta}_T^* - \theta^*). \tag{14}$$

By transforming the model in this fashion, we arrive at a first order autoregressive model *with drift*. Such models may be considered as special cases of input-output systems, for which there exists a vast literature, cf. e.g. Ljung (1977) for a seminal paper. However, when applied to our setting, this literature usually involves a restriction on the slope parameter. More importantly, it only provides a rate for the *bivariate* (henceforth called *joint*) estimator  $\hat{\theta}_T$ , i.e. a rate for its *norm*. For instance, by considering a single input  $u_t = 1$  and lag length  $k = 1$  the model in Lai & Wei (1982a, Section 3) is an AR(1)-model with intercept. Their Theorem 2 then provides a joint rate for the stable and the unit root case, except for intercept zero. We will pursue this *joint estimation* approach on the basis of a more recent paper by Nielsen (2005), which allows to cover also the explosive case.

On the other hand, some reflection shows that the speed of convergence of the estimator of the slope may be quite different from that of the intercept. In view of this observation, the joint approach will only produce rates valid for the *slower* of these estimators, which – not surprisingly – is that of the intercept. To take account of this difference and to obtain individual ‘optimal’ rates, we also pursue the *separate estimation* approach, treating the one-dimensional formula for each estimator on its own. This allows us to make use of the powerful martingale convergence theorems found in the literature, cf. Lai & Wei (1982a) and Wei (1985).

In the sequel, we will start with the separate approach in Section 3.1.1 and then consider the joint approach in Section 3.1.2.

### 3.1.1 Separate estimation of the parameters

Consider the separate OLS estimators of  $\beta^*$  and  $\delta^*$  in (13), namely

$$\begin{aligned}\hat{\beta}_T^* &= \frac{\sum_{t=1}^T (a_{t-1} - \bar{a}_T^-) (a_t - \bar{a}_T)}{A_T}, \\ \hat{\delta}_T^* &= \bar{a}_T - \hat{\beta}_T^* \bar{a}_T^-, \end{aligned}$$

where

$$\begin{aligned}\bar{a}_T &= \frac{1}{T} \sum_{t=1}^T a_t, & \bar{a}_T^- &= \frac{1}{T} \sum_{t=1}^T a_{t-1}, \\ A_T^0 &= \sum_{t=1}^T a_{t-1}^2, & A_T &= \sum_{t=1}^T (a_{t-1} - \bar{a}_T^-)^2 = A_T^0 - T (\bar{a}_T^-)^2.\end{aligned}$$

Theorem 2 and Corollary 1 below will summarise the properties of the OLS estimators of the original slope  $\beta$  and intercept  $\delta$ , see (11). The proofs, however, are conducted in terms of the starred model, see Appendix B. The main argument of the proofs consists in determining the rate of convergence of the slope estimator. In the process, the aforementioned martingale convergence theorems will be used as well as laws of large numbers (LLNs) and results of the LIL-type. The convergence rate of the intercept estimator will follow from that of the slope, taking account of the LIL. As will be plain, the behaviour of both estimators depends crucially on the relative magnitude of  $\beta$  and  $\gamma$ .

The case distinctions in Theorem 2 and Corollary 1 are phrased in terms of the parameter  $c$  and correspond to the original  $a_t$  in (12) being a stable, unit root or explosive process; see also the discussion in Section 2.1 above. They are equivalent to properties of the transformed  $a_t^*$  in (13), as indicated by the parameter  $\beta^*$ :

$$\begin{aligned}|\beta^*| < 1 &\Leftrightarrow 0 < c < 2, \\ \beta^* = 1 &\Leftrightarrow c = 0, \\ \beta^* = -1 &\Leftrightarrow c = 2, \\ |\beta^*| > 1 &\Leftrightarrow c < 0 \text{ or } c > 2.\end{aligned}$$

A special role is played by the scenario  $\beta^* = 1 \wedge \delta^* = 0$ , corresponding to  $\beta = 1 \wedge \delta = 0$  or indeed  $c = 0 \wedge \delta = 0$ , in which case no result is available. See also the comments on this combination of parameter values in Section B.1.1.3.

**Theorem 2.** *Strong consistency of the OLS estimator  $\widehat{\beta}_T$  of the slope parameter  $\beta$  holds at the following rates.*

(i) *Stable case:  $0 < c < 2$ . If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ ,*

$$\sqrt{\frac{T}{\ln_2 T}} (\widehat{\beta}_T - \beta) = O(1).$$

*If only second moments exist, then*

$$\sqrt{\frac{T}{(\ln T)^{1+\eta}}} (\widehat{\beta}_T - \beta) = o(1)$$

*for all  $\eta > 0$ .*

(ii) *Unit root case:  $c = 0 \wedge \delta \neq 0$ . If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ ,*

$$\sqrt{\frac{T^3}{\ln_2 T}} (\widehat{\beta}_T - \beta) = O(1).$$

*If only second moments exist, then*

$$\sqrt{\frac{T^3}{(\ln T)^{1+\eta}}} (\widehat{\beta}_T - \beta) = o(1)$$

*for all  $\eta > 0$ .*

(iii) *Unit root case:  $c = 2$ . If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ ,*

$$\frac{T}{(\ln_2 T)^3} (\widehat{\beta}_T - \beta) = O(1).$$

*If only second moments exist, then*

$$\frac{T}{(\ln T)^{1+\eta}} (\widehat{\beta}_T - \beta) = o(1)$$

*for all  $\eta > 0$ .*

(iii) *Explosive case:  $c < 0$  or  $c > 2$ . Assuming only second moments,*

$$\frac{|1-c|^T}{T^{1/2+\eta}} (\widehat{\beta}_T - \beta) = o(1) \tag{15}$$

*for all  $\eta > 0$ . If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ , (15) remains valid, with  $O(1)$  instead of  $o(1)$ , for  $\eta = 0$ .*

The following corollary summarises the behaviour of the intercept estimator.

**Corollary 1.** *Strong consistency of the OLS estimator  $\widehat{\delta}_T$  of the intercept  $\delta$  holds at the following rates.*

(i) *Stable case:*  $0 < c < 2$ . If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ ,

$$\sqrt{\frac{T}{\ln_2 T}} (\widehat{\delta}_T - \delta) = O(1).$$

If only second moments exist, then

$$\sqrt{\frac{T}{(\ln T)^{1+\eta}}} (\widehat{\delta}_T - \delta) = o(1)$$

for all  $\eta > 0$ .

(iia) *Unit root case:*  $c = 0 \wedge \delta \neq 0$ . If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ ,

$$\sqrt{\frac{T}{\ln_2 T}} (\widehat{\delta}_T - \delta) = O(1).$$

If only second moments exist, then

$$\sqrt{\frac{T}{(\ln T)^{1+\eta}}} (\widehat{\delta}_T - \delta) = o(1)$$

for all  $\eta > 0$ .

(iib) *Unit root case:*  $c = 2$ . Same as in case (iia).

(iii) *Explosive case:*  $c < 0$  or  $c > 2$ . Assuming only second moments,

$$T^{1/2-\eta} (\widehat{\delta}_T - \delta) = o(1) \tag{16}$$

for all  $\eta > 0$ . If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ , (16) remains valid, with  $O(1)$  instead of  $o(1)$ , for  $\eta = 0$ .

As is to be expected, the rate of the slope estimator is throughout at least as good as that of the intercept estimator, with equality holding in the stable case.

Let us compare the rates in Theorem 2 and Corollary 1 with the corresponding rates for weak convergence as obtained in Theorem 2 of CM17. In all cases, the strong consistency rates are basically those needed for weak convergence, corrected (i.e. divided) by some sequence  $r_T \nearrow \infty$ . For the nonexplosive cases, the sequences are of the form  $r_T = (\ln_2 T)^{m/2}$  or  $r_T = (\ln T)^p$ , depending on whether the innovations  $\varepsilon_t$  possess higher than second moments or not. In the stable case ( $m = 1$ ), this corresponds to the transition from the CLT to the LIL.

Since the proofs make use of theorems that are valid for general martingale difference sequences, it may be expected that for i.i.d. error terms some results may be weakened, in the sense that the  $(\ln_2 T)^{m/2}$ -corrections may also work if only second moments are assumed. This is true, for instance, in the stable case, in which a LIL for stationary ergodic processes by Stout (1970) can be applied, see Section B.2.3.

### 3.1.2 Joint estimation of the parameters

The second approach for estimating  $\theta = (\delta, \beta)'$  in (11) employs recent results derived in Nielsen (2005) on the rates of convergence of the *studentised version*

$$\tau_T = M_T^{1/2} (\widehat{\theta}_T^* - \theta^*) \tag{17}$$

of the OLS estimator of  $\theta^*$  in (13), where  $M_T$  is the sample second moment matrix of the regressor  $(1, a_{t-1}^*)$ :

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

Given rates for  $\|\tau_T\|$ , the idea is to find sequences of numbers  $\chi_T$  s.t.

$$\chi_T \|M_T^{-1/2}\| \|\tau_T\| = O(1) \quad (18)$$

where  $\|A\| = \lambda_{\max}^{1/2}(A'A)$  denotes the spectral norm of  $A$ . In view of (14) and (17) the sequence of numbers  $\chi_T$  will then satisfies

$$\chi_T (\hat{\theta}_T - \theta) = O(1). \quad (19)$$

Note that (18) involves calculating norms of the *inverse*  $M_T^{-1/2}$ . This amounts to estimating the *minimal* eigenvalue of  $M_T$  since

$$\|M_T^{-1/2}\|^2 = \lambda_{\max}(M_T^{-1}) = \frac{1}{\lambda_{\min}(M_T)}$$

so that (18) turns into

$$\frac{\chi_T}{\sqrt{\lambda_{\min}(M_T)}} \|\tau_T\| = O(1).$$

Hence this approach is tantamount to investigating the asymptotic behaviour of the minimal eigenvalues  $\lambda_T = \lambda_{\min}(M_T)$  and to finding sequences of numbers  $\chi_T$  s.t.

$$\chi_T \frac{\|\tau_T\|}{\sqrt{\lambda_T}} = O(1).$$

As a further complication, the rates of the two components of  $\hat{\theta}_T$  can (and will in the majority of cases) be different, so that (19) will only exhibit the behaviour of the *worse* of the two parameters.

Applying this approach to the starred model and then transforming back to the original one we obtain the following result. The proof is relegated to Section B.3.

**Theorem 3.** *Assume that  $\mathbf{E}|\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ . Then strong consistency of the joint OLS estimator  $\hat{\theta}_T$  holds at the following rates.*

(i) *Stable case:  $0 < c < 2$ .*

$$\sqrt{\frac{T}{\ln_2 T}} (\hat{\theta}_T - \theta) = O(1).$$

(ii) *Unit root case: For both  $c = 0$  and  $c = 2$ ,*

$$\sqrt{\frac{T}{\ln T}} (\hat{\theta}_T - \theta) = O(1).$$

(iii) *Explosive case:  $c < 0$  or  $c > 2$ .*

$$T^{1/2-\rho} (\hat{\theta}_T - \theta) = o(1)$$

for every  $\rho > 1/\alpha$ .

Note that, in contrast to the separate approach in Section 3.1.1, the case of  $c = 0 \wedge \delta = 0$  is covered in this theorem. It is, however, special in the sense that it is *not* covered by Nielsen (2005) such that the proof follows a different route, see the discussion in Section B.3.

### 3.1.3 Comparison

Before comparing Theorem 2 and Corollary 1 on the one hand and Theorem 3 on the other, recollect the following: First, the joint approach yields the rates for the bivariate OLS estimator  $\widehat{\theta}_T = (\widehat{\delta}_T, \widehat{\beta}_T)$ . Hence the rate of  $\widehat{\theta}_T$  cannot be better than that of the worse of its two constituents. Secondly, since when estimating the parameters separately the rate of the slope is always better than that of the intercept it appears reasonable to compare the rate of the latter, cf. Corollary 1, to the joint rate. Thirdly, Theorem 3 actually assumes the existence of somewhat higher than second moments, see Result 3 in Section B.3, so that only the corresponding results in the corollary need to be considered.

For the stable case, both approaches yield the same rates. For the unit root case with non-vanishing intercept, the rate of the intercept in Corollary 1 is better than the joint rate Theorem 3. For the explosive case, the separate rate is  $T^{1/2}$  and therefore better than the joint rate, which is  $T^{1/2-\eta}$ , for  $\eta > 1/\alpha$  and  $\mathbf{E}|\varepsilon_t|^\alpha < \infty$ . Note that if only second moments are assumed separate estimation yields the rate  $T^{1/2-\eta}$ , for all  $\eta > 0$ . In the joint approach, this would formally correspond to the case  $\alpha = \infty$ , i.e. bounded error terms.

Actually, Nielsen (2005) not only considers the asymptotics of the studentised version of the OLS estimator but also presents rates for the OLS estimator itself, see Theorem 2.5 *loc. cit.*. Applied to our model, they are as follows:

1.  $\widehat{\beta}_T - \beta = O\left(\sqrt{T/\ln T}\right),$
2.  $\widehat{\theta}_T - \theta = o\left(T^{1/2-\rho}\right)$  for all  $\rho > 1/\alpha$ .

Both results are valid only for a nonzero intercept. They constitute omnibus rates across all possible cases and might be of interest if no a priori knowledge about the size of the slope is available. For a fair comparison these rates ought to be compared to corresponding omnibus rates based on Theorem 2 and its corollary. For the slope, this is the rate for the stable case, namely  $\sqrt{T/\ln_2 T}$ . For both parameters simultaneously, this is the rate of the explosive case, which is also  $T^{1/2-\rho}$  for all  $\rho > 0$ . Both are faster than those in Nielsen (2005).

## 3.2 Decreasing gain

Consider now OLS estimation of  $\delta$  and  $\beta$  in

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t \tag{20}$$

under decreasing gain learning, i.e. with  $a_t$  is given by

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

see (10). Recall that the strong consistency of  $a_t$  is given by Theorem 1. That of the OLS estimator of  $\beta$  in (20) is presented in the following theorem, whose proof can be found in Appendix C. As in the context of weak consistency of  $\widehat{\beta}_T$  in CM17, only the cases  $c < 1/2$  and  $c > 1/2$  are considered. The boundary case of  $c = 1/2$  does not seem amenable to our methods and is left to future research.

**Theorem 4.** *For decreasing gain with gain sequence  $\gamma_t = \gamma/t$ , strong consistency of the OLS estimator  $\widehat{\beta}_T$  of the slope parameter  $\beta$  holds at the following rates.*

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

$$\lim_{T \rightarrow \infty} \sqrt{\frac{\ln T}{(\ln_2 T)^{1+\eta}}} \left(\widehat{\beta}_T - \beta\right) = 0$$

for every  $\eta > 0$ . If, in addition,  $\mathbf{E}|\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ , this may be sharpened to

$$\sqrt{\frac{\ln T}{\ln_3 T}} (\hat{\beta}_T - \beta) = O(1).$$

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

$$\sqrt{\frac{T^{1-2c}}{(\ln T)^{1+\eta}}} (\hat{\beta}_T - \beta) = O(1).$$

for every  $\eta > 0$ . If, in addition,  $\mathbf{E}|\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ , this may be sharpened to

$$\sqrt{\frac{T^{1-2c}}{\ln_2 T}} (\hat{\beta}_T - \beta) = O(1).$$

Let us compare the convergence rates in Theorem 4 to those established in the context of the weak consistency of  $\hat{\beta}_T$  in CM17. It was found in their Theorem 4 that,

(i) for  $c > 1/2$ ,  $A_T = O_p(\ln T)$  and

(ii) for  $c < 1/2$ ,  $A_T = O_p(T^{1-2c})$ .

It is hence plain from Theorem 4 above that the ‘path taming’ sequences are given by  $(\ln_2 T)^{-(1+\eta)}$  and  $(\ln T)^{-(1+\eta)}$ , respectively.

A comparison of Theorems 1 and 4 reveals that there is a trade-off between the convergence rates of  $a_t$  and the  $\hat{\beta}_T$ . For a further discussion of this issue, see CM17.

As a byproduct, rates of consistency for the OLS estimator of the intercept  $\delta$  are easily obtained from the formula

$$\hat{\delta}_T - \delta = (\hat{\beta}_T - \beta)\bar{a}_T + \bar{\varepsilon}_T.$$

In view of the LIL, any normalising sequence  $\chi_T$  should satisfy

$$\chi_T \sqrt{\frac{\ln_2 T}{T}} = O(1). \quad (21)$$

It is apparent that all rates exhibited for the slope in Theorem 4 satisfy (21). Therefore, we have the following result.

**Corollary 2.** *Strong consistency of the OLS estimator  $\hat{\delta}_T$  of the intercept  $\delta$  holds at the same rates as for the slope.*

### 3.3 Comparison with Lai & Wei

Let us return to the point raised in the introduction that strong consistency may obtain despite the near optimal sufficient condition established by Lai & Wei (1982a) being violated. Denote by  $\lambda_{\max}(T)$  and  $\lambda_{\min}(T)$  the maximal and the minimal eigenvalue, respectively, of the second moment matrix of the regressors  $(1, a_{t-1})$ . Applied to our simple regression model (5), the Lai-Wei condition then amounts to

$$\ln \lambda_{\max}(T) = o(\lambda_{\min}(T)), \quad (22)$$

under the additional assumption that  $\mathbf{E}|\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ .

It is shown in Section B.4 that, for constant gain learning, condition (22) is satisfied in the stable and unit root case. In the explosive case, however, the condition is violated:

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} \rightarrow 2 \ln |\beta|.$$



Nevertheless, strong consistency is established in Theorems 2 and 3.

For decreasing gain, (22) is satisfied for  $c < 1/2$  but violated for  $c > 1/2$ . In the latter case,

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} \rightarrow \frac{2c - 1}{\gamma^2 \sigma^2},$$

see Section C.4. Yet Theorem 4 shows that the OLS estimator is strongly consistent.

## 4 Conclusion

This paper derives the strong consistency of the OLS estimators in a stereotypical macroeconomic model with constant and decreasing gain adaptive learning. The rates are obtained at which the estimators converge. In addition, the asymptotic behaviour of the learning recursion is examined in the process: For decreasing gain learning, the almost sure rates at which the learning recursion converges to the rational expectations equilibrium are derived, while it is argued that for constant gain learning no convergence to the equilibrium takes place. Interestingly, the best sufficient conditions for strong consistency of the OLS estimator currently available in the literature are not met in some of the models we consider.

## References

- Adam, K., A. Marcet & J. P. Nicolini (2016). “Stock market volatility and learning”. *Journal of Finance* vol. 71, pp. 33–82.
- Apostol, T. M. (1974). *Mathematical Analysis*. 2nd ed. Reading, MA: Addison-Wesley.
- Benveniste, A., M. Métivier & P. Priouret (1990). *Adaptive Algorithms and Stochastic Approximation*. Originally published in French in 1987. Berlin: Springer.
- Bray, M. M. & N. E. Savin (1986). “Rational expectations equilibria, learning and model specification”. *Econometrica* vol. 54, pp. 1129–1160.
- Brock, W. A. & C. H. Hommes (1997). “A rational route to randomness”. *Econometrica* vol. 65, pp. 1059–1095.
- Chan, N. H. & C. Z. Wei (1988). “Limiting distributions of least squares estimates of unstable autoregressive processes”. *Annals of Statistics* vol. 16, pp. 367–401.
- Chevillon, G., M. Massmann & S. Mavroeidis (2010). “Inference in models with adaptive learning”. *Journal of Monetary Economics* vol. 57, pp. 341–351.
- Chow, Y. S. (1965). “Local convergence of martingales and the law of large numbers”. *Annals of Mathematical Statistics* vol. 36, pp. 552–558.
- Chow, Y. S. & H. Teicher (1973). “Iterated logarithm laws for weighted averages”. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* vol. 26, pp. 87–94.
- Christopeit, N. (1986). “Quasi-least-squares estimation in semimartingale regression models”. *Stochastics* vol. 16, pp. 255–278.
- Christopeit, N. & M. Massmann (2010). “Consistent estimation of structural parameters in regression models with adaptive learning”. TI Discussion Paper 10-077/4. URL: <http://www.tinbergen.nl/discussionpaper/?paper=1633>.
- Christopeit, N. & M. Massmann (2012). “Strong consistency of the least-squares estimator in simple regression models with stochastic regressors”. TI Discussion Paper 12-109/III. URL: <http://www.tinbergen.nl/discussionpaper/?paper=1988>.
- Christopeit, N. & M. Massmann (2017). “Estimating structural parameters in regression models with adaptive learning”. *Econometric Theory*. DOI: 10.1017/S0266466616000529.
- Clarida, R., J. Galí & M. Gertler (2000). “Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory”. *Quarterly Journal of Economics* vol. 115, pp. 147–180.

- Dieppe, A., A. González Pandiella, S. Hall & A. Willman (2013). “Limited information minimal state variable learning in a medium-scale multi-country model”. *Economic Modelling* vol. 33, pp. 808–825.
- Donsker, M. D. & S. R. S. Varadhan (1977). “On the laws of the iterated logarithm for local times”. *Communications on Pure and Applied Mathematics* vol. 30, pp. 707–753.
- Evans, G. W. & S. Honkapohja (2001). *Learning and Expectations in Macroeconomics*. Princeton: Princeton University Press.
- Evans, G. W., S. Honkapohja, T. J. Sargent & N. Williams (2013). “Bayesian model averaging, learning and model selection”. *Macroeconomics at the Service of Public Policy*. Ed. by T. J. Sargent & J. Vilmunen. Oxford, UK: Oxford University Press. Chap. 6, pp. 99–119.
- Grenander, U. & M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*. New York: Wiley.
- Kottmann, T. (1990). “Learning Procedures and Rational Expectations in Linear Models with Forecast Feedback”. PhD thesis. University of Bonn.
- Kushner, H. (2010). “Stochastic approximation: a survey”. *Wiley Interdisciplinary Reviews: Computational Statistics* vol. 2, pp. 87–96.
- Lai, T. L. (2003). “Stochastic approximation”. *Annals of Statistics* vol. 31, pp. 391–406.
- Lai, T. L. & C. Z. Wei (1982a). “Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems”. *Annals of Statistics* vol. 10, pp. 154–166.
- Lai, T. L. & C. Z. Wei (1982b). “Asymptotic properties of projections with applications to stochastic regression problems”. *Journal of Multivariate Analysis* vol. 12, pp. 346–370.
- Lai, T. L. & C. Z. Wei (1983). “Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters”. *Journal of Multivariate Analysis* vol. 13, pp. 1–23.
- Lai, T. L. & C. Z. Wei (1985). “Asymptotic properties of multivariate weighted sums with applications to stochastic linear regression in linear dynamic systems”. *Multivariate Analysis*. Ed. by P. R. Krishnaiah. Vol. VI. Amsterdam: North-Holland, pp. 375–393.
- Ljung, L. (1977). “Analysis of recursive stochastic systems”. *IEEE Transactions on Automatic Control* vol. AC-22, pp. 551–575.
- Lucas, R. E. (1973). “Some international evidence on output-inflation tradeoffs”. *American Economic Review* vol. 63, pp. 326–334.
- Malmendier, U. & S. Nagel (2016). “Learning from inflation experience”. *Quarterly Journal of Economics* vol. 131, pp. 53–87.
- Marcet, A. & T. J. Sargent (1995). “Speed of convergence of recursive least squares: learning with autoregressive moving-average perceptions”. *Learning and Rationality in Economics*. Ed. by A. Kirman & M. Salmon. Blackwell. Chap. 6, pp. 179–215.
- Milani, F. (2007). “Expectations, learning and macroeconomic persistence”. *Journal of Monetary Economics* vol. 54, pp. 2065–2082.
- Nielsen, B. (2005). “Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms”. *Econometric Theory* vol. 21, pp. 534–561.
- Phillips, P. C. B. (1987). “Towards a unified asymptotic theory for autoregression”. *Biometrika* vol. 74, pp. 535–547.
- Phillips, P. C. B. (2007). “Regression with slowly varying regressors and nonlinear trends”. *Econometric Theory* vol. 23, pp. 557–614.
- Phillips, P. C. B. & T. Magdalinos (2007). “Limit theory for moderate deviations from a unit root”. *Journal of Econometrics* vol. 136, pp. 115–130.
- Phillips, P. C. B. & T. Magdalinos (2008). “Limit theory for explosively cointegrated systems”. *Econometric Theory* vol. 24, pp. 865–887.
- Sargent, T. J. (1993). *Bounded Rationality in Macroeconomics*. Oxford: Clarendon Press.

- Sargent, T. J. (1999). *The Conquest of American Inflation*. Princeton: Princeton University Press.
- Shiryayev, A. N. (1996). *Probability*. 2nd ed. 1st edition 1984. New York: Springer.
- Solo, V. & X. Kong (1995). *Signal Processing Algorithms*. Upper Saddle River: Prentice Hall.
- Stout, W. F. (1970). “The Hartmann-Wintner law of the iterated logarithm for martingales”. *Annals of Mathematical Statistics* vol. 41, pp. 2158–2160.
- Wang, X. & J. Yu (2015). “Limit theory for an explosive autoregressive process”. *Economics Letters* vol. 126, pp. 176–180.
- Wei, C. Z. (1985). “Asymptotic properties of least-squares estimates in stochastic regression models”. *Annals of Statistics* vol. 13, pp. 1498–1508.

## A Proof of Theorem 1

The proofs proceed along similar lines as those followed in CM17, and may be considered almost sure (a.s.) convergence counterparts of the weak convergence results obtained there. In particular, they rely on the decomposition of  $a_t$  exposed in Appendix B.1 *loc. cit.*. Yet instead of central limit theorems (CLTs) use will be made of appropriate strong laws of large numbers (LLNs) and laws of the iterated logarithm (LILs).

Reconsider the recursion of  $a_t$  in (10). Passing from  $a_t$  to  $a_t^\# = a_t - \alpha$  and remembering that  $\alpha = \delta / (1 - \beta) = \gamma\delta/c$ , it follows that  $a_t^\#$  obeys the dynamics

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

and the DGP in (5) takes the form

$$y_t = \alpha + \beta a_{t-1}^\# + \varepsilon_t. \quad (\text{A.2})$$

Since, henceforth, we will be working exclusively with  $a_t^\#$ , let us rename  $a_t^\#$  as  $a_t$  for notational simplicity.

### A.1 $c > 1/2$

For didactic reasons it seems favourable to split the proof into two cases:  $c > 1$  and  $1/2 < c \leq 1$ .

#### A.1.1 $c > 1$

It is shown in Appendix B.1.2 of CM17 that, for  $c > 1$ , we may write the solution of (A.1) in the form

$$a_t = O(t^{-c}) + \gamma(\xi_t + \eta_t), \quad (\text{A.3})$$

cf. (B.21) *loc. cit.*. Here

$$\xi_t = \frac{1}{t^c} v_t, \quad \eta_t = \frac{1}{t^c} w_t, \quad (\text{A.4a})$$

$$v_t = \sum_{i=1}^t \frac{\varepsilon_i}{i^{1-c}}, \quad w_t = \sum_{i=1}^t \frac{O_{ti}(1)}{i^{2-c}} \varepsilon_i. \quad (\text{A.4b})$$

The  $O_{ti}(1)$ -terms are deterministic. As to  $\xi_t$ , it follows from the integral comparison theorem (ICT), see Apostol (1974, Proposition 8.23), that the predictable quadratic variation of  $v_t$  is given by

$$\langle v \rangle_t = \sigma^2 \sum_{i=1}^t i^{2(c-1)} = \frac{\sigma^2}{2c-1} t^{2c-1} + O(1). \quad (\text{A.5})$$

Hence  $\langle v \rangle_\infty = \lim_{t \rightarrow \infty} \langle v \rangle_t = \infty$  a.s.. Therefore, by the LIL for sums of independent random variables proved in Chow & Teicher (1973),

$$\limsup_{t \rightarrow \infty} \frac{|v_t|}{\sqrt{2 \langle v \rangle_t \ln_2 \langle v \rangle_t}} = 1.$$

As a consequence,

$$\limsup_{t \rightarrow \infty} \frac{|v_t|}{\sqrt{t^{2c-1} \ln_2 t}} = \sigma \sqrt{\frac{2}{2c-1}}$$

so that

$$\limsup_{t \rightarrow \infty} \sqrt{\frac{t}{\ln_2 t}} |\xi_t| = \sigma \sqrt{\frac{2}{2c-1}} \quad (\text{A.6})$$

and hence  $\xi_t \rightarrow 0$ .

Turning to  $\eta_t$ , it follows again from the ICT that

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

Therefore, by monotone convergence, for  $c > 1$ ,

$$\mathbf{E} \sum_{t=1}^{\infty} t \eta_t^2 < \infty. \quad (\text{A.7})$$

In particular, this means that

$$\sqrt{t} |\eta_t| = o(1). \quad (\text{A.8})$$

(A.3) together with (A.6) and (A.8) then shows that

$$\limsup_{t \rightarrow \infty} \sqrt{\frac{t}{\ln_2 t}} |a_t| = \sigma \gamma \sqrt{\frac{2}{2c-1}}. \quad (\text{A.9})$$

### A.1.2 $1/2 < c \leq 1$

For this case, we need the more refined decomposition

$$a_t = O(t^{-c}) + \gamma (\xi'_t + \eta'_t) + O(t^{-(1+c)}), \quad (\text{A.10})$$

where now

$$\xi'_t = \frac{1}{t^c} v'_t, \quad \eta'_t = \frac{1}{t^c} w'_t, \quad (\text{A.11a})$$

$$v'_t = \sum_{i=1}^t \theta_i \frac{\varepsilon_i}{i^{1-c}}, \quad w'_t = \frac{1}{t} \sum_{i=1}^t \frac{O_{ti}(1)}{i^{1-c}} \varepsilon_i. \quad (\text{A.11b})$$

Here the  $\theta_i$  are deterministic coefficients satisfying  $\lim_{t \rightarrow \infty} \theta_i = 1$ . In Appendix B.1 of CM17 the two different representations (A.3) and (A.10) were introduced to treat the two cases  $c < 1/2$  and  $c \geq 1/2$  separately. Actually, (A.10) remains valid for  $c \leq 1$ , and we need it here to handle this case, since the approach taken above for  $c > 1$ , which is based on the representation (A.3), does not work.

For the proof we need the following lemma.

**Lemma 1.** *Consider the sums*

$$R_t = \sum_{i=1}^t \sigma_i^2 \quad \text{and} \quad S_t = \sum_{i=1}^t \theta_i^2 \sigma_i^2.$$

Suppose that  $\theta_i \rightarrow 1$  and  $R_\infty = \infty$ . Then

$$\frac{S_t}{R_t} \rightarrow 1.$$

The proof runs along familiar lines like, e.g., that of the Kronecker lemma.

The predictable quadratic variation of  $v'_t$  is given by

$$\langle v' \rangle_t = \sigma^2 \sum_{i=1}^t \theta_i^2 i^{2(c-1)}. \quad (\text{A.12})$$

Since  $\langle v \rangle_\infty = \infty$  a.s. continues to hold for  $1/2 < c \leq 1$  and  $\theta_i \rightarrow 1$ , it follows from Lemma 1 that

$$\lim_{t \rightarrow \infty} \frac{\langle v' \rangle_t}{\langle v \rangle_t} = 1, \quad (\text{A.13})$$

so that

$$\lim_{t \rightarrow \infty} \frac{\langle v' \rangle_t}{t^{2c-1}} = \frac{\sigma^2}{2c-1}.$$

Hence, arguing as above, we find that (A.6) remains valid with  $\xi_t$  replaced by  $\xi'_t$ . Turning to  $\eta'_t$ , since

$$\mathbf{E} w_t'^2 = \frac{1}{t^2} O(1) \sum_{i=1}^t i^{2(c-1)} = O(t^{2c-3}),$$

we find that

$$\mathbf{E} \eta_t'^2 = \mathbf{E} \frac{w_t'^2}{t^{2c}} = O(t^{-3}).$$

Therefore

$$\mathbf{E} \sum_{t=1}^{\infty} t \eta_t'^2 = O(1) \sum_{t=1}^{\infty} t^{-2} < \infty,$$

and, a fortiori, (A.7) and (A.8) remain valid for  $\eta'$ . This proves (A.9) for  $1/2 < c \leq 1$ .

For later use, cf. the proof of Theorem 4, we need the asymptotic behaviour of  $\bar{a}_T$ . But this follows easily from (A.9) since

$$|\bar{a}_T| \leq \frac{1}{T} \sum_{t=1}^T |a_t| = O(1) \frac{1}{T} \sum_{t=t_0}^T \sqrt{\frac{\ln_2 t}{t}} = O\left(\sqrt{\frac{\ln_2 T}{T}}\right) \quad (\text{A.14})$$

since

$$\int_{t_0}^T \sqrt{\frac{\ln_2 t}{t}} dt = 2\sqrt{T \ln_2 T} + O(\sqrt{T}).$$

## A.2 $c = 1/2$

We go back to the decomposition (A.10). Comparing (A.12), for  $c = 1/2$ , with

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

it follows from Lemma 1 that

$$\langle v' \rangle_t = \sigma^2 \ln t + O(1).$$

Hence, by the LIL,

$$\limsup_{t \rightarrow \infty} \frac{|v'_t|}{\sqrt{\ln t \ln_3 t}} = \sigma\sqrt{2} \quad (\text{A.15})$$

and

$$\limsup_{t \rightarrow \infty} \sqrt{\frac{t}{\ln t \ln_3 t}} |\xi'_t| = \sigma\sqrt{2}. \quad (\text{A.16})$$

As for  $\eta'_t$ ,

$$\mathbf{E}w_t'^2 = O(1) \frac{1}{t^2} \sum_{i=1}^t \frac{1}{i} = O\left(\frac{\ln t}{t^2}\right).$$

Therefore

$$\mathbf{E} \sum_{t=1}^{\infty} t\eta_t'^2 = \mathbf{E} \sum_{t=1}^{\infty} t \left(\frac{w'_t}{t^{1/2}}\right)^2 = O(1) \sum_{t=1}^{\infty} \frac{\ln t}{t^2} < \infty,$$

so that

$$\sqrt{t}\eta'_t = o(1). \quad (\text{A.17})$$

It then follows from (A.10) together with (A.16) and (A.17) that

$$\limsup_{t \rightarrow \infty} \sqrt{\frac{t}{\ln t \ln_3 t}} |a_t| = \sigma\gamma\sqrt{2}.$$

### A.3 $c < 1/2$

Our starting point is again (A.10). By Kolmogorov's LLN,

$$\lim_{t \rightarrow \infty} t^c \xi'_t = \lim_{t \rightarrow \infty} v'_t = \sum_{i=1}^{\infty} \theta_i \frac{\varepsilon_i}{i^{1-c}} = v' \quad (\text{A.18})$$

is finite with probability one. As to  $\eta'_t$ ,

$$\mathbf{E}w_t'^2 = O(1) \frac{1}{t^2} \sum_{i=1}^t i^{2(c-1)} = O(t^{-2}),$$

so that

$$\mathbf{E} \sum_{t=1}^{\infty} (t^c \eta'_t)^2 = \mathbf{E} \sum_{t=1}^{\infty} w_t'^2 < \infty.$$

Therefore, with probability one,

$$\lim_{t \rightarrow \infty} t^c \eta'_t = 0. \quad (\text{A.19})$$

The  $O(t^{-c})$ -term in (A.10) is actually of the form  $a_0 B_0 t^{-c}$ , see equation (B.18) of CM17, where  $a_0$  is the initial value of  $a_t^\#$  and  $B_0 = e^{-(cE_0 + C_0)}$ . Here  $E_0$  is the Euler constant, and  $C_0$  may be calculated from

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

Summarising, we find that

$$\lim_{t \rightarrow \infty} t^c a_t = u = a_0 B_0 + \gamma v'. \quad (\text{A.20})$$

The limit also takes place in  $L^2$ , so that  $u$  is an  $L^2$ -variable with mean  $a_0 B_0$ . Moreover,  $v'$  and hence  $u$  has a continuous distribution function, cf. CM17 on this issue.

As to  $\bar{a}_T$ , we have

$$\begin{aligned} \bar{a}_T &= \frac{1}{T} \sum_{t=1}^T a_t = \frac{1}{T} \sum_{t=1}^T t^{-c} (t^c a_t + o(1)) \\ &= \frac{u}{1-c} \frac{1}{T^c} + o(T^{-c}). \end{aligned} \quad (\text{A.21})$$

## B Proof of Theorems 2 and 3

### B.1 Prerequisites

Most of the proofs in Appendix B are based on the following two *martingale convergence theorems* (MCTs). They are valid for all solutions (independent of the initial value) of the AR(1)-model with drift, whatever the slope and the intercept may be. Since some of the results presented below seem not to be completely standard and might be of some interest also outside the scope of our focus we formulate them in a neutral terminology, starting with the AR(1)-model

$$y_n = \mu + \lambda y_{n-1} + \varepsilon_n \quad (\text{B.1})$$

which will replace (13) throughout Appendix B. The MCTs concern the almost sure asymptotic behaviour of (local) martingales of the form

$$u_n = \sum_{k=1}^n y_{k-1} \varepsilon_k, \quad (\text{B.2})$$

when the  $y_n$  are predictable with respect to the past  $\varepsilon_n$ . The decisive point is that the *predictable quadratic variation*

$$A_n^0 = \sum_{k=1}^n y_{k-1}^2 \quad (\text{B.3})$$

of  $u_n$  satisfies

$$\lim_{n \rightarrow \infty} A_n^0 = \infty \quad (\text{B.4})$$

with probability one. Since in our context we only deal with i.i.d.  $\varepsilon_n$  and consider only predictable solutions  $y_n$  of (B.1) for which (B.4) holds, we will cite the MCTs only for this special case. Actually, they are valid for more general martingale difference sequences  $\varepsilon_n$ , in which case the expectations in the assumptions of the theorems below have to be replaced by the corresponding conditional expectations.

The following two results are given in Chow (1965) and Wei (1985), respectively. See also Lai & Wei (1982a).

**MCT 1 (Chow (1965)).**

$$\sum_{k=1}^n y_{k-1} \varepsilon_k = o \left[ \sqrt{A_n^0 (\ln A_n^0)^{1+\eta}} \right] \quad (\text{B.5a})$$

for all  $\eta > 0$ . If  $\mathbf{E} |\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ , this may be sharpened to

$$\sum_{k=1}^n y_{k-1} \varepsilon_k = O \left[ \sqrt{A_n^0 \ln A_n^0} \right]. \quad (\text{B.5b})$$

**MCT 2 (Wei (1985)).** If, in addition to  $\mathbf{E} |\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ ,

$$y_n^2 = o \left[ (A_n^0)^\gamma \right] \quad (\text{B.6})$$

for some  $0 < \gamma < 1$ , then

$$\sum_{k=1}^n y_{k-1} \varepsilon_k = O \left[ \sqrt{A_n^0 \ln_2 A_n^0} \right]. \quad (\text{B.7})$$

In the proofs of Theorems 2 and 3, the a.s. behaviour of the *basic statistics*

$$\bar{y}_n = \frac{1}{n} \sum_{k=1}^n y_{k-1}, \quad (\text{B.8})$$

$$A_n^0 = \sum_{k=1}^n y_{k-1}^2, \quad (\text{B.9})$$

$$A_n = \sum_{k=1}^n (y_{k-1} - \bar{y}_n)^2 \quad (\text{B.10})$$

will be of fundamental importance. In particular, they determine the behaviour of the *derived statistics*

$$\frac{A_n^0}{A_n}, \varphi(A_n^0), \psi(A_n^0) \text{ and } \frac{\bar{y}_n^-}{A_n}.$$

The latter appear explicitly in our representation of the OLS estimator. The functions  $\varphi$  and  $\psi$  are given by

$$\varphi(x) = \sqrt{\frac{x}{(\ln x)^{1+\eta}}} \text{ and } \psi(x) = \sqrt{\frac{x}{\ln_2 x}}. \quad (\text{B.11})$$

In passing, note the trivial but useful formula

$$A_n = A_n^0 - n(\bar{y}_n^-)^2. \quad (\text{B.12})$$

### B.1.1 Path properties

To evaluate the a.s. asymptotic behaviour of these statistics we need to know the path properties up to second order of the solution process  $y_n$ . These will be reported below together with sketches of their proofs.

#### B.1.1.1 Stable case.

(i) *Basic statistics:*

$$\lim_{n \rightarrow \infty} \bar{y}_n^- = \lim_{n \rightarrow \infty} \bar{y}_n = \frac{\mu}{1-\lambda}, \quad (\text{B.13a})$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} A_n^0 = \tau^2, \text{ with } \tau^2 = \frac{\sigma^2}{1-\lambda^2} + \frac{\mu^2}{(1-\lambda)^2}, \quad (\text{B.13b})$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} A_n = \frac{\sigma^2}{1-\lambda^2}. \quad (\text{B.13c})$$

(ii) *Derived statistics:*

$$\lim_{n \rightarrow \infty} \frac{A_n^0}{A_n} = \frac{\tau^2}{\sigma^2/(1-\lambda^2)} = 1 + \frac{\mu^2}{\sigma^2} \frac{1+\lambda}{1-\lambda} = r, \quad (\text{B.14a})$$

$$\lim_{n \rightarrow \infty} \frac{\varphi(n)}{\varphi(A_n^0)} = \tau^{-1} \text{ for all } \eta \geq 0, \quad (\text{B.14b})$$

$$\lim_{n \rightarrow \infty} \frac{\psi(n)}{\psi(A_n^0)} = \tau^{-1}, \quad (\text{B.14c})$$

$$\lim_{n \rightarrow \infty} n \frac{\bar{y}_n^-}{A_n} = \frac{\mu}{\sigma^2} \frac{1+\lambda}{1-\lambda} = r - 1. \quad (\text{B.14d})$$

These properties follow readily from the well-known ergodic behaviour of the stationary solution to (B.1) and carry over to any other (causal) solution.

**Remark B.1.** *For the stable case, condition (B.6) is satisfied provided  $\mathbf{E}|\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ . This can be seen as follows. By Lai & Wei (1985, Theorem 1), any solution  $y_n^0$  of the homogeneous model satisfies*

$$\left(y_n^0\right)^2 = o\left(n^{2\beta}\right) \text{ for every } \beta > 1/\alpha.$$

*Since the inhomogeneous solution  $y_n$  differs from  $y_n^0$  at most by a constant, this remains true for  $y_n$ . On the other hand, by (B.13b),  $A_n^0 = n\tau^2(1 + o(1))$ . Hence, for every  $\gamma$ ,*

$$\frac{y_n^2}{(A_n^0)^\gamma} = n^{2\beta-\gamma} o(1).$$



Letting  $\beta \searrow 1/\alpha$ , we find that for all  $2/\alpha < \gamma < 1$  finally  $2/\alpha < 2\beta < \gamma < 1$ , so that  $2\beta - \gamma < 0$ . As a consequence, the LIL-type MCT 2 is valid:

$$\sum_{k=1}^n y_{k-1} \varepsilon_k = O\left[\sqrt{n \ln_2 n}\right].$$

**B.1.1.2 Explosive case.** The causal solution is

$$y_n = \lambda^n y_0 + \mu \frac{\lambda^n - 1}{\lambda - 1} + \lambda^n m_n,$$

with

$$m_n = \sum_{i=1}^n \lambda^{-i} \varepsilon_i.$$

By Kolmogorov's LLN, the martingale  $m_n$  converges a.s. and in  $L^2$  to some finite limit  $m$ :

$$m = \lim_{n \rightarrow \infty} m_n = \sum_{i=1}^{\infty} \lambda^{-i} \varepsilon_i, \quad (\text{B.15})$$

and

$$\mathbf{Var}(m) = \frac{\sigma^2}{\lambda^2 - 1}.$$

**Remark B.2.**  $m$  has a continuous distribution, cf. Remark A.1 in CM17.

The following path properties are then immediate consequences.

(i) With probability one and in  $L^2$

$$\lim_{n \rightarrow \infty} \lambda^{-n} y_n = y_0 + m + \frac{\mu}{\lambda - 1}. \quad (\text{B.16})$$

If  $y_0$  is independent of  $(\varepsilon_n)_{n \geq 1}$ , then the distribution of the limit is continuous.

(ii) *Basic statistics:*

$$n \lambda^{-n} \overline{y_n} \rightarrow \frac{\lambda}{\lambda - 1} (y_0 + m) + \frac{\lambda \mu}{(\lambda - 1)^2} = \kappa, \quad (\text{B.17a})$$

$$n \lambda^{-2n} \overline{y_n^2} \rightarrow \frac{\lambda^2}{\lambda^2 - 1} \left[ y_0 + m + \frac{\mu}{\lambda - 1} \right]^2 = v^2. \quad (\text{B.17b})$$

Note that  $v^2$  is a random variable  $> 0$  a.s..

The proof is obtained from (B.16) by applying the Toeplitz Lemma to  $\xi_n = \lambda^{-n} y_n$  (for (B.17a)) and to  $\xi_n^2 = \lambda^{-2n} y_n^2$  (for (B.17b)).

(iii) *Derived statistics:*

$$\lim_{n \rightarrow \infty} \lambda^{-2n} A_n^0 = v^2, \quad (\text{B.18a})$$

$$\lim_{n \rightarrow \infty} \frac{A_n^0}{A_n} = 1, \quad (\text{B.18b})$$

$$\lim_{n \rightarrow \infty} \frac{\varphi(\lambda^{2n})}{\varphi(A_n^0)} = \frac{1}{v} \text{ for all } \eta \geq 0, \quad (\text{B.18c})$$

$$\lim_{n \rightarrow \infty} n \lambda^n \frac{\overline{y_n}}{A_n} = \frac{\kappa}{v^2}. \quad (\text{B.18d})$$

Sketch of proof. (B.18a) is just (B.17b). (B.18b) follows from  $A_n = A_n^0 - n(\bar{y}_n^-)^2$  and  $n(\bar{y}_n^-)^2 = O(\lambda^{2n}/n)$ . (B.18c) is a consequence of

$$A_n^0 = v^2 \lambda^{2n} (1 + o(1)), \quad \ln A_n^0 = (1 + o(1)) \ln \lambda^{2n},$$

$$\frac{1}{\varphi(A_n^0)^2} = \frac{(\ln A_n^0)^{1+\eta}}{A_n^0} = \frac{(\ln \lambda^{2n})^{1+\eta}}{\lambda^{2n}} \frac{1}{v^2} (1 + o(1)).$$

(B.18d) follows from (B.17a) together with (B.18a) and (B.18b).

**Remark B.3.** *Unlike in the stable case, (B.6) does not hold. This is clear since  $y_n^2 \sim \lambda^{2n}$ ,  $A_n^0 \sim \lambda^{2n}$ , so that*

$$\frac{y_n^2}{(A_n^0)^\gamma} \sim \lambda^{2n(1-\gamma)},$$

with the exponent on the right hand side being positive for all  $0 < \gamma < 1$ . Therefore there is no need to consider the statistic  $\psi(A_n^0)$ .

### B.1.1.3 Unit root case.

Case 1:  $\lambda = 1, \mu \neq 0$ .

The solution to (B.1) in this case is the *random walk with drift*

$$y_n = y_0 + n\mu + \sum_{k=1}^n \varepsilon_k. \quad (\text{B.19})$$

(i) *Basic statistics:*

$$\lim_{n \rightarrow \infty} \frac{y_n}{n} = \mu, \quad (\text{B.20a})$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \bar{y}_n = \frac{\mu}{2}, \quad (\text{B.20b})$$

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \bar{y}_n^2 = \frac{\mu^2}{3}. \quad (\text{B.20c})$$

The proof is again a direct consequence of (B.19) and the Toeplitz Lemma.

(ii) *Derived statistics:*

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} A_n^0 = \frac{\mu^2}{3}, \quad (\text{B.21a})$$

$$\lim_{n \rightarrow \infty} \frac{A_n^0}{A_n} = 4, \quad (\text{B.21b})$$

$$\lim_{n \rightarrow \infty} \sqrt{\frac{n^3}{(\ln n)^{1+\eta}}} \frac{1}{\varphi(A_n^0)} = \frac{3^{1+\eta/2}}{\mu^2} \quad \text{for all } \eta \geq 0, \quad (\text{B.21c})$$

$$\lim_{n \rightarrow \infty} \sqrt{\frac{n^3}{\ln_2 n}} \frac{1}{\psi(A_n^0)} = \frac{\sqrt{3}}{\mu^2}, \quad (\text{B.21d})$$

$$\lim_{n \rightarrow \infty} n^2 \frac{\bar{y}_n^-}{A_n} = \frac{3}{2\mu}. \quad (\text{B.21e})$$

Sketch of proof. (B.21a) is just (B.20c). The rest follows from

$$A_n^0 = n^3 \frac{\mu^2}{3} (1 + o(1)),$$

$$\ln A_n^0 = 3 \ln n + O(1) = (1 + o(1)) 3 \ln n,$$

$$\ln_2 A_n^0 = (1 + o(1)) \ln_2 n,$$

$$n^2 \frac{\bar{y}_n^-}{A_n} = \frac{\bar{y}_n^- / n}{A_n / n^3}$$

together with (B.20b).

**Remark B.4.** For  $\lambda = 1, \mu \neq 0$ , condition (B.6) is fulfilled since

$$\frac{y_n^2}{(A_n^0)^\gamma} \sim \frac{n^2}{n^{3\gamma}} = n^{2-3\gamma}$$

tends to 0 for every  $2/3 < \gamma < 1$ . Therefore MCT 2 is valid.

Case 2:  $\lambda = 1, \mu = 0$ .

In this case,  $y_n$  is a random walk. For the first two moments, we have the following estimates:

$$\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{2n \ln_2 n}} |\bar{y}_n| \leq \frac{2}{3} \sigma \quad (\text{B.22})$$

and

$$\limsup_{n \rightarrow \infty} \frac{1}{2n^2 \ln_2 n} A_n^0 \leq \sigma^2, \quad (\text{B.23a})$$

$$\liminf_{n \rightarrow \infty} \frac{\ln_2 n}{2n^2} A_n^0 = \frac{\sigma^2}{8}. \quad (\text{B.23b})$$

(B.22) follows from the LIL by applying a straightforward extension of the Toeplitz Lemma (replacing ‘lim’ by ‘lim sup’) together with the ICT, partial integration and a calculus version of the Toeplitz Lemma. As to (B.23), both properties are cited in Lai & Wei (1982a, Example 2). The first is a consequence of the LIL, whereas the second is based on a theorem by Donsker & Varadhan (1977, page 751). The problem is that  $1/A_n^0 = O(n^{-2} \ln_2 n)$  and  $n |\bar{y}_n|^2 = O(n^2 \ln_2 n)$ , so that  $Q_n = n |\bar{y}_n|^2 / A_n^0 = O((\ln_2 n)^2)$ . This makes it impossible to determine the behaviour of  $A_n = A_n^0 (1 - Q_n)$ .

Case 3:  $\lambda = -1$ .

As is easily shown,  $y_n$  differs from the alternating random walk  $y_n^0$  without drift only by a constant:

$$y_n = \begin{cases} y_n^0 + \mu & \text{for } n \text{ odd,} \\ y_n^0 & \text{for } n \text{ even.} \end{cases} \quad (\text{B.24})$$

Apparently, the a.s. asymptotic behaviour of the paths is governed by the LIL.

(i) *Mean:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n y_{k-1} = \begin{cases} \frac{\mu}{2} & \text{for } n \text{ odd,} \\ 0 & \text{for } n \text{ even.} \end{cases} \quad (\text{B.25})$$

*Proof.* The proof takes up an idea in the proof of Theorem 2 in Appendix A.3 of CM17. Since the initial value does not play any role, we assume that  $y_0 = 0$ . Then

$$y_n^0 = (-1)^n \tilde{S}_n,$$

where we have introduced the random walk

$$\tilde{S}_n = \sum_{k=1}^n \tilde{\varepsilon}_k \quad \text{with} \quad \tilde{\varepsilon}_k = (-1)^k \varepsilon_k.$$

Then

$$\sum_{k=1}^n y_{k-1}^0 = \sum_{k=1}^n (-1)^{k-1} \tilde{S}_{k-1} = \sigma_n \tilde{S}_n - \sum_{k=1}^n \sigma_k \tilde{\varepsilon}_k.$$

The last equality follows by partial summation, with

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

Then, by the LLN,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n y_{k-1}^0 = 0$$

The assertion then follows from (B.24).

(ii) 2nd moments:

$$\limsup_{n \rightarrow \infty} \frac{1}{2n^2 \ln_2 n} A_n^0 \leq \sigma^2, \quad (\text{B.26a})$$

$$\liminf_{n \rightarrow \infty} \frac{\ln_2 n}{2n^2} A_n^0 \geq \frac{\sigma^2}{2}. \quad (\text{B.26b})$$

We encountered (B.26) essentially in (B.23) where, however,  $y_n$  was a random walk:  $y_n = S_n = \sum_{k=1}^n \varepsilon_k$ . In the present case, since the asymptotic behaviour of the second moments is independent of the drift and the sign, the behaviour of  $A_n^0$  is governed by that of  $\tilde{S}_n$ . (B.26a) is then a consequence of the LIL, which continues to hold when passing from  $S_n$  to  $\tilde{S}_n$  by the already mentioned LIL for martingale difference sequences of Chow & Teicher (1973) and thus remains valid for  $\tilde{S}_n$ . As to (B.26b), it is based on a theorem of Donsker & Varadhan (1977, page 751), which assumes *i.i.d.* shocks, in which case

$$\liminf_{n \rightarrow \infty} \frac{\ln_2 n}{2n^2} A_n^0 = \frac{\sigma^2}{8}. \quad (\text{B.27})$$

At any rate, for *symmetric*  $\varepsilon_n$ ,  $\tilde{S}_n$  is again a random walk of *i.i.d.* shocks so that (B.27) holds for such error terms. It can, however, be shown that at least (B.26b) is satisfied for non-symmetric  $\tilde{\varepsilon}_k$ .

(B.26) carries over to  $A_n$ . For the first inequality, this follows trivially from  $A_n \leq A_n^0$ . For (B.26b), it is a consequence of (B.25), which implies that  $|\bar{y}_n^-| = O(1)$  and therefore

$$\frac{\ln_2 n}{2n^2} A_n = \frac{\ln_2 n}{2n^2} A_n^0 - \frac{\ln_2 n}{2n} |\bar{y}_n^-|^2 = \frac{\ln_2 n}{2n^2} A_n^0 + o(1). \quad (\text{B.28})$$

(iii) Derived statistics:

$$\frac{A_n^0}{A_n} = O\left[(\ln_2 n)^2\right], \quad (\text{B.29a})$$

$$\frac{\varphi_n}{\varphi(A_n^0)} = O(1) \text{ with } \varphi_n = \frac{n}{\sqrt{(\ln n)^{1+\eta} \ln_2 n}}, \text{ for all } \eta \geq 0, \quad (\text{B.29b})$$

$$\frac{\psi_n}{\psi(A_n^0)} = O(1) \text{ with } \psi_n = \frac{n}{\ln_2 n}, \quad (\text{B.29c})$$

$$\frac{|\bar{y}_n^-|}{A_n} = O\left(\frac{\ln_2 n}{n^2}\right). \quad (\text{B.29d})$$

Proof. *Ad (B.29a).*

$$\begin{aligned} \frac{1}{(\ln_2 n)^2} \frac{A_n^0}{A_n} &= \frac{\frac{1}{2n^2 \ln_2 n} A_n^0}{\frac{\ln_2 n}{2n^2} A_n} = \frac{P_n}{Q_n}, \\ \limsup_{n \rightarrow \infty} \frac{1}{(\ln_2 n)^2} \frac{A_n^0}{A_n} &\leq \frac{\limsup_{n \rightarrow \infty} P_n}{\liminf_{n \rightarrow \infty} Q_n} \leq 8. \end{aligned}$$

*Ad (B.29b).* Denote  $\alpha_n = 2n^2 \ln_2 n$ ,  $\beta_n = (\ln_2 n)/2n^2$ . Then

$$\ln A_n^0 = \ln \alpha_n + \ln(\alpha_n^{-1} A_n^0) = (1 + o(1)) 2 \ln n + \ln(\alpha_n^{-1} A_n^0)$$

or

$$\frac{\ln A_n^0}{2 \ln n} = (1 + o(1)) + \frac{\ln(\alpha_n^{-1} A_n^0)}{2 \ln n} \quad (\text{B.30})$$

By (B.26a),

$$\limsup_{n \rightarrow \infty} \ln(\alpha_n^{-1} A_n^0) \leq \ln \sigma^2, \quad (\text{B.31})$$

so that  $\ln(\alpha_n^{-1}A_n^0)$  is bounded from above. On the other hand, since the left hand side of (B.30) is positive for  $n$  large enough,  $\ln(\alpha_n^{-1}A_n^0)/2\ln n$  is also bounded from below. Therefore

$$\frac{\ln A_n^0}{2\ln n} = O(1). \quad (\text{B.32})$$

Coming to (B.29b) and making use of (B.32) and (B.26b), we may write

$$\begin{aligned} \frac{1}{\varphi(A_n^0)^2} &= \frac{(\ln A_n^0)^{1+\eta}}{A_n^0} = \beta_n^{-1} \frac{(\ln A_n^0)^{1+\eta}}{\beta_n^{-1}A_n^0} \\ &= \beta_n^{-1} (2\ln n)^{1+\eta} \left( \frac{\ln A_n^0}{2\ln n} \right)^{1+\eta} O(1) \\ &= \frac{(2\ln n)^{1+\eta} \ln_2 n}{2n^2} O(1). \end{aligned}$$

This shows (B.29b) with

$$\varphi_n = \frac{n}{\sqrt{(\ln n)^{1+\eta} \ln_2 n}}.$$

*Ad (B.29c).* By (B.32), denoting the  $O(1)$ -term by  $C_n$  and noting that  $C_n > 0$  for  $n$  large enough,

$$\ln \frac{\ln A_n^0}{2\ln n} = \ln_2 A_n^0 - \ln_2 n - \ln 2 = \ln C_n$$

or

$$\frac{\ln_2 A_n^0}{2\ln n} = 1 + \ln 2 + \ln C_n.$$

Since the left hand side is positive for  $n$  large enough,  $\liminf_{n \rightarrow \infty} \ln C_n \geq -(1 + \ln 2)$ . As a consequence,

$$\frac{\ln_2 A_n^0}{2\ln_2 n} = O(1). \quad (\text{B.33})$$

Making use of (B.32) and (B.26b), we may then write

$$\begin{aligned} \frac{1}{\psi(A_n^0)^2} &= \frac{\ln_2 A_n^0}{A_n^0} = \beta_n^{-1} \frac{\ln_2 A_n^0}{\beta_n^{-1}A_n^0} \\ &= \beta_n^{-1} (2\ln_2 n) \frac{\ln_2 A_n^0}{2\ln_2 n} O(1) \\ &= \frac{(\ln_2 n)^2}{n^2} O(1). \end{aligned}$$

This shows (B.29c) with

$$\psi_n = \frac{n}{\ln_2 n}.$$

*Ad (B.29d).* This is a straightforward consequence of (B.25) and (B.26b) together with (B.28).

**Remark B.5.** If  $\mathbf{E}|\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ , it follows from (B.26b) that

$$\frac{1}{A_n^0} = O\left[\frac{\ln_2 n}{n^2}\right].$$

On the other hand, by the LIL,  $y_n^2 = O(n \ln_2 n)$ . Therefore

$$\frac{y_n^2}{(A_n^0)^\gamma} = \frac{(\ln_2 n)^{1+\gamma}}{n^{2\gamma-1}} O(1),$$

so that for every  $1/2 < \gamma < 1$  (B.6) will be satisfied.

### B.1.2 Eigenvalues of the moment matrix

Consider the second moment matrix of the regressor  $(1, y_{n-1})$  in (B.1):

$$M_n = \begin{pmatrix} n & n\bar{y}_n^- \\ n\bar{y}_n^- & A_n^0 \end{pmatrix}$$

Its eigenvalues (EVs) are given by

$$\lambda_{\max} = \frac{n + A_n^0}{2} \left[ 1 + \sqrt{1 - 4D_n} \right], \quad (\text{B.34a})$$

$$\lambda_{\min} = \frac{n + A_n^0}{2} \left[ 1 - \sqrt{1 - 4D_n} \right], \quad (\text{B.34b})$$

with

$$D_n = \frac{nA_n^0 - (n\bar{y}_n^-)^2}{(n + A_n^0)^2}.$$

Note that both eigenvalues are real, so that  $0 \leq D_n \leq 1/4$ .

In the following, we will report the eigenvalues for the different cases and give brief sketches of the proofs. We will make use of the corresponding path properties established in Section B.1.1.

**B.1.2.1 Stable case.** Then

$$\frac{1}{n}A_n^0 \rightarrow \tau^2, \quad \bar{y}_n^- \rightarrow \frac{\mu}{1 - \lambda}$$

and

$$D_n = \frac{\frac{1}{n}A_n^0 - (\bar{y}_n^-)^2}{\left(1 + \frac{1}{n}A_n^0\right)^2} \rightarrow D = \frac{\sigma^2 / (1 - \lambda)^2}{(1 + \tau^2)^2}.$$

Therefore

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \lambda_{\max} &= \lambda_+ = \frac{1 + \tau^2}{2} \left[ 1 + \sqrt{1 - 4D} \right], \\ \lim_{n \rightarrow \infty} \frac{1}{n} \lambda_{\min} &= \lambda_- = \frac{1 + \tau^2}{2} \left[ 1 - \sqrt{1 - 4D} \right]. \end{aligned}$$

**B.1.2.2 Explosive case.** Then

$$\lambda^{-2n} A_n^0 \rightarrow v^2, \quad n^2 \lambda^{-2n} (\bar{y}_n^-)^2 \rightarrow \kappa^2.$$

Hence  $A_n^0 = v^2 \lambda^{2n} (1 + o(1))$ ,  $(n\bar{y}_n^-)^2 = \lambda^{2n} (\kappa^2 + o(1))$ , and

$$nD_n = \frac{A_n^0 - n(\bar{y}_n^-)^2}{(1 + A_n^0/n)^2} = \frac{\lambda^{2n} v^2}{(1 + A_n^0/n)^2} (1 + o(1)) \sim \frac{v^2}{(\lambda^{-n} + \lambda^{-n} A_n^0/n)^2} \rightarrow 0$$

since

$$\frac{1}{n} |\lambda|^{-n} A_n^0 = \frac{|\lambda|^n}{n} \lambda^{-2n} A_n^0 \rightarrow \infty.$$

As a consequence, the *maximal EV* is

$$\lim_{n \rightarrow \infty} \lambda^{-2n} \lambda_{\max} = v^2. \quad (\text{B.35})$$

For the minimal EV, use the square root expansion

$$\sqrt{1 + x} = 1 + \frac{x}{2} + O(x^2)$$

to obtain

$$\begin{aligned}\lambda_{\min} &= \frac{n + A_n^0}{2} \left[ 1 - \sqrt{1 - 4D_n} \right] = \frac{n + A_n^0}{2} \left[ 1 - \left( 1 - 2D_n + O(D_n^2) \right) \right] \\ &= D_n (n + A_n^0) + (n + A_n^0) O(D_n^2).\end{aligned}\tag{B.36}$$

Making use of the path properties,

$$\begin{aligned}\frac{1}{n} (n + A_n^0) D_n &= \frac{A_n^0 - n(\bar{y}_n^-)^2}{n + A_n^0} = \frac{\lambda^{2n} v^2}{n + A_n^0} (1 + o(1)) = \frac{v^2}{\lambda^{-2n} n + \lambda^{-2n} A_n^0} (1 + o(1)) \rightarrow 1, \\ \frac{1}{n} (n + A_n^0) D_n^2 &\rightarrow 0,\end{aligned}$$

(B.36) yields, for the *minimal EV*,

$$\frac{1}{n} \lambda_{\min} \rightarrow 1.$$

**Remark B.6.** (B.35) implies that  $\lim_{n \rightarrow \infty} n^{-1} \ln \lambda_{\max} \rightarrow 2 \ln |\lambda|$ , which is in accordance with Nielsen (2005, Corollary 7.2) and Lai & Wei (1985, Corollary 2). Note, however, that these results concern the explosive process  $y_n$  without intercept, so that, transferred to our case, they just say something about the asymptotic behaviour of the one-dimensional matrix  $A_n^0$ . Therefore, Nielsen's result that also  $\lim_{n \rightarrow \infty} n^{-1} \ln \lambda_{\min} \rightarrow 2 \ln |\lambda|$  applies only to an at least 2-dimensional explosive process.

### B.1.2.3 Unit root case.

Case 1:  $\lambda = 1, \mu \neq 0$ .

In this case,

$$n^2 D_n = \frac{\frac{1}{n^3} A_n^0 - \left( \frac{1}{n} \bar{y}_n^- \right)^2}{\left( \frac{1}{n^2} + \frac{1}{n^3} A_n^0 \right)^2} \rightarrow \frac{\frac{\mu^2}{3} - \left( \frac{\mu}{2} \right)^2}{\left( \frac{\mu^2}{3} \right)^2} = \frac{3}{4} \mu^{-2}$$

In particular, since  $D_n \rightarrow 0$ ,

$$\frac{1}{n^3} \frac{n + A_n^0}{2} \left[ 1 + \sqrt{1 - 4D_n} \right] = \frac{1}{n^3} A_n^0 (1 + o(1))$$

and the *maximal EV* satisfies

$$\frac{1}{n^3} \lambda_{\max} \rightarrow \frac{\mu^2}{3}.$$

For the minimal EV, note that

$$\begin{aligned}\frac{1}{n} (n + A_n^0) D_n &= \left( \frac{2}{n^3} \frac{n + A_n^0}{2} \right) n^2 D_n \rightarrow \frac{1}{4}, \\ \frac{1}{n} (n + A_n^0) D_n^2 &\rightarrow 0.\end{aligned}$$

It then follows from (B.36) that the *minimal eigenvalue* satisfies

$$\frac{1}{n} \lambda_{\min} \rightarrow \frac{1}{4}.$$

Case 2:  $\lambda = 1, \mu = 0$ .

In this case, the formulas (B.34) fail since the behaviour of  $D_n$  cannot be derived from the path properties in B.1.1.3. We therefore pass to the equivalent formulas

$$\lambda_{\pm} = \frac{1}{2} \left[ A_n^0 + n \pm \sqrt{(A_n^0 - n)^2 + 4p_n^2} \right],$$

where we have put  $p_n = n\bar{y}_n^-$ . Since  $n/A_n^0 = O(\ln_2 n/n)$  and  $p_n/A_n^0 = O(\sqrt{(\ln_2 n)^3/n})$  by virtue of (B.22) and (B.23b), we may write

$$\begin{aligned}\lambda_{\pm} &= \frac{1}{2} \left[ A_n^0 + n \pm A_n^0 \sqrt{\left(1 - \frac{n}{A_n^0}\right)^2 + 4 \left(\frac{p_n}{A_n^0}\right)^2} \right] \\ &= \frac{1}{2} \left[ A_n^0 + n \pm A_n^0 \sqrt{1 + o(1)} \right] \\ &= \frac{1}{2} \left[ A_n^0 + n \pm A_n^0 (1 + o(1)) \right].\end{aligned}$$

Hence

$$\begin{aligned}\lambda_{\max} &= A_n^0 (1 + o(1)), \\ \lambda_{\min} &= \frac{n}{2} (1 + o(1)).\end{aligned}$$

Case 3:  $\lambda = -1$ .

Since  $A_n^0/n \rightarrow \infty$ ,

$$D_n \leq \frac{nA_n^0}{(n + A_n^0)^2} = \frac{A_n^0/n}{(1 + A_n^0/n)^2} \rightarrow 0$$

and

$$\lambda_{\max} = (n + A_n^0) (1 + o(1)).$$

Therefore, making use of the results in Section B.1.1.3, the *maximal EV* satisfies

$$\begin{aligned}\limsup_{n \rightarrow \infty} \frac{1}{2n^2 \ln_2 n} \lambda_{\max} &\leq \sigma^2, \\ \liminf_{n \rightarrow \infty} \frac{\ln_2 n}{2n^2} \lambda_{\max} &\geq \frac{\sigma^2}{2}.\end{aligned}$$

In (B.36), we have

$$\frac{1}{n} D_n (n + A_n^0) = \frac{A_n^0 - n(\bar{y}_n^-)^2}{n + A_n^0} = \frac{1 - \frac{n}{A_n^0} (\bar{y}_n^-)^2}{1 + \frac{n}{A_n^0}}.$$

Since  $n/A_n^0 \rightarrow 0$  and  $\bar{y}_n^- = O(1)$ ,

$$\frac{1}{n} D_n (n + A_n^0) \rightarrow 1.$$

As a consequence, for the *minimal EV*,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \lambda_{\min} = 1.$$

### B.1.3 Generalities

As announced in Section 3.1, we are concerned with the estimation of the parameters in the transformed model (13). In the notation adopted for Appendix B, this means OLS estimation of the parameters in (B.1). Above, we distinguished two alternative approaches:

The *separate approach* makes use of the representations of the OLS estimator for the single parameters already introduced in Section 3.1.1. In the notation adopted for this Appendix, these are

$$\begin{aligned}\hat{\lambda}_n &= \frac{\sum_{k=1}^n (y_{k-1} - \bar{y}_n^-) (y_k - \bar{y}_n)}{\sum_{k=1}^n (y_{k-1} - \bar{y}_n^-)^2}, \\ \hat{\mu}_n &= \bar{y}_n - \hat{\lambda}_n \bar{y}_n^-\end{aligned}$$



or, in the form to be used below,

$$\widehat{\lambda}_n - \lambda = \frac{u_n}{A_n} - \frac{\bar{y}_n}{A_n} \sum_{k=1}^n \varepsilon_k, \quad (\text{B.38a})$$

$$\widehat{\mu}_n - \mu = (a - \widehat{a}_n) \bar{y}_n + \bar{\varepsilon}_n, \quad (\text{B.38b})$$

with  $u_n$  as in (B.2).

The *joint approach* investigates the 2-dimensional OLS estimator  $\widehat{\theta}_n = (\widehat{\mu}_n, \widehat{\lambda}_n)'$  of  $\theta = (\mu, \lambda)'$ . As carried out in Section 3.1.2, it is based on recent convergence results of its *studentised version*  $\tau_n$ , in terms of which

$$\widehat{\theta}_n - \theta = M_n^{-1/2} \tau_n. \quad (\text{B.39})$$

Here  $M_n$  is the moment matrix

$$M_n = \begin{pmatrix} n & n\bar{y}_n \\ n\bar{y}_n & A_n^0 \end{pmatrix}. \quad (\text{B.40})$$

The first approach will be treated in Section B.2, the second in Section B.3.

## B.2 Proof of Theorem 2

### B.2.1 Estimation of the slope

**B.2.1.1 Generalities.** Starting with (B.38a), write  $\widehat{\lambda}_n - \lambda$  in the form

$$\widehat{\lambda}_n - \lambda = \frac{u_n}{A_n^0} \frac{A_n^0}{A_n} - \frac{\bar{y}_n}{A_n} \sum_{k=1}^n \varepsilon_k = \varphi^{-1} \left( A_n^0 \right) U_n \frac{A_n^0}{A_n} - V_n, \quad (\text{B.41})$$

where we have introduced

$$U_n = \frac{u_n}{\sqrt{A_n^0 (\ln A_n^0)^{1+\eta}}}, \quad V_n = \frac{\bar{y}_n}{A_n} \sum_{k=1}^n \varepsilon_k$$

and  $\varphi(x)$  as in (B.11). We may then apply MCT 1 to obtain that  $U_n = O(1)$  for every  $\eta > 0$ . If  $\mathbf{E}|\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ , then  $\eta = 0$  is admitted. If even (B.6) is satisfied, then we may apply the stronger MCT 2 to find that

$$U_n = \frac{u_n}{\sqrt{A_n^0 \ln_2 A_n^0}}$$

is  $O(1)$ . In this case,  $\varphi(x)$  can be replaced by  $\psi(x)$  from (B.11).

As a consequence, if one is interested in *deterministic* convergence rates for  $\widehat{\lambda}_n - \lambda$ , then one should investigate the sequence  $\chi(A_n^0)$ , with  $\chi$  either of the functions  $\varphi$  or  $\psi$ , and try to find a sequence of numbers  $\chi_n$  s.t.

$$\frac{\chi_n}{\chi(A_n^0)} \frac{A_n^0}{A_n} = O(1), \quad (\text{B.42a})$$

$$\chi_n V_n = O(1). \quad (\text{B.42b})$$

Then it will hold that

$$\chi_n (\widehat{\lambda}_n - \lambda) = U_n \frac{\chi_n}{\chi(A_n^0)} \frac{A_n^0}{A_n} - \chi_n V_n = O(1). \quad (\text{B.43})$$

Sufficient conditions for (B.42a) are

$$\frac{\chi_n}{\chi(A_n^0)} = O(1), \quad (\text{B.44a})$$

$$\frac{A_n^0}{A_n} = O(1). \quad (\text{B.44b})$$

As to (B.42b), write

$$\chi_n V_n = \chi_n \frac{\bar{y}_n}{A_n} \sum_{k=1}^n \varepsilon_k = \chi_n \sqrt{n \ln_2 n} \frac{\bar{y}_n}{A_n} \sqrt{\frac{1}{n \ln_2 n}} \sum_{k=1}^n \varepsilon_k.$$

By the LIL, a sufficient condition for (B.42b) to hold is

$$\chi_n \sqrt{n \ln_2 n} \frac{\bar{y}_n}{A_n} = O(1). \quad (\text{B.45})$$

**B.2.1.2 Stable case.** Here we will make use of the results in Section B.1.1.1. If  $\mathbf{E}|\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ , (B.6) is satisfied, cf. Remark B.1, and we may apply MCT 2 and  $\chi = \psi$ . Making use of (B.14) we see that  $\chi_n = \psi(n)$  will satisfy both (B.44) and (B.45).

If one assumes only finite 2nd moments, one has to make use of MCT 1 and  $\chi_n = \varphi(n)$  will do.

**B.2.1.3 Explosive case.** Here the results in Section B.1.1.2 come into action. According to Remark B.3, we cannot apply MCT 2 but must be content with MCT 1. Making use of (B.18), it is easily shown that

$$\chi_n = \varphi(\lambda^{2n}) = \frac{|\lambda|^n}{n^{1/2+\eta}}$$

satisfies all conditions (B.44) and (B.45). If  $\mathbf{E}|\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ , we may even choose  $\eta = 0$ .

**B.2.1.4 Unit root case.** Here we will make use of the result in Section B.1.1.3.

Case 1:  $\lambda = 1, \mu \neq 0$ .

By Remark B.4, (B.6) is satisfied if higher moments exist. By (B.21),

$$\chi_n = \sqrt{\frac{n^3}{\ln_2 n}}$$

will satisfy (B.44) and (B.45).

If only  $\alpha = 2$  is assumed, then we have to use MCT 1 and

$$\chi_n = \sqrt{\frac{n^3}{(\ln n)^{1+\eta}}}, \quad \eta > 0.$$

Case 2:  $\lambda = 1, \mu = 0$ .

Cf. the discussion in Section B.1.1.3.

Case 3:  $\lambda = -1$ .

Since (B.6) is satisfied, cf. Remark B.5, for  $\alpha > 2$ , (B.29c) yields

$$\frac{n}{\ln_2 n} \frac{1}{\psi(A_n^0)} = O(1).$$

This suggests

$$\chi'_n = \frac{n}{\ln_2 n}$$

as a candidate for the normalising sequence. However, in view of (B.29a), (B.44b) is not satisfied. This may be compensated by choosing

$$\chi_n = \frac{n}{(\ln_2 n)^3} = \frac{\chi'_n}{(\ln_2 n)^2},$$

which will satisfy (B.42a) and (B.42b) (with  $\chi = \psi$ ).

If only second moments are assumed,

$$\chi_n = \frac{n}{\sqrt{(\ln n)^{1+\eta} \ln_2 n}}$$

will do.

Collecting the results, we obtain, in the notation used in this appendix, Theorem 2.

**Theorem 2.** *Strong consistency of the OLS estimator  $\hat{\lambda}_n$  of the slope parameter  $\lambda$  holds at the following rates:*

(i) *Stable case:  $|\lambda| < 1$ . If  $\mathbf{E} |\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ ,*

$$\sqrt{\frac{n}{\ln_2 n}} (\hat{\lambda}_n - \lambda) = O(1). \quad (\text{B.46})$$

*If only second moments exist, then*

$$\sqrt{\frac{n}{(\ln n)^{1+\eta}}} (\hat{\lambda}_n - \lambda) = o(1).$$

*for all  $\eta > 0$ .*

(iia) *Unit root case:  $\lambda = 1$  and  $\mu \neq 0$ . If  $\mathbf{E} |\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ ,*

$$\sqrt{\frac{n^3}{\ln_2 n}} (\hat{\lambda}_n - \lambda) = O(1).$$

*If only second moments exist, then*

$$\sqrt{\frac{n^3}{(\ln n)^{1+\eta}}} (\hat{\lambda}_n - \lambda) = o(1).$$

*for all  $\eta > 0$ .*

(iib) *Unit root case:  $\lambda = -1$ . If  $\mathbf{E} |\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ ,*

$$\frac{n}{(\ln_2 n)^3} (\hat{\lambda}_n - \lambda) = O(1).$$

*If only second moments exist, then*

$$\frac{n}{\sqrt{(\ln n)^{1+\eta} \ln_2 n}} (\hat{\lambda}_n - \lambda) = o(1).$$

*for all  $\eta > 0$ .*

(iii) *Explosive case:  $|\lambda| > 1$ . Assuming only 2nd moments,*

$$\frac{|\lambda|^n}{n^{1/2+\eta}} (\hat{\lambda}_n - \lambda) = o(1) \quad (\text{B.47})$$

*for all  $\eta > 0$ . If  $\mathbf{E} |\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ , (B.47) remains valid, with  $O(1)$  instead of  $o(1)$  for  $\eta = 0$ .*

## B.2.2 Estimation of the intercept

Remember that

$$\widehat{\mu}_n - \mu = (\widehat{\lambda}_n - \lambda)\overline{y}_n + \overline{\varepsilon}_n.$$

Hence any rate  $\phi_n$  for which  $\phi_n(\widehat{\mu}_n - \mu) = O(1)$  should satisfy

$$\phi_n \overline{\varepsilon}_n = O(1), \tag{B.48a}$$

$$\phi_n(\widehat{\lambda}_n - \lambda)\overline{y}_n = O(1). \tag{B.48b}$$

In view of the LIL, (B.48a) means that

$$\frac{\phi_n}{\psi(n)} = O(1). \tag{B.49a}$$

(B.49a) rules out all rates tending faster to infinity than  $\psi_n = \psi(n)$ . Also, if  $\chi_n$  is the rate for  $\widehat{\lambda}_n$  according to Theorem 2, then (B.48b) becomes

$$\phi_n(\widehat{\lambda}_n - \lambda)\overline{y}_n = \chi_n(\widehat{\lambda}_n - \lambda)\frac{\phi_n}{\chi_n}\overline{y}_n = O(1).$$

Therefore a second sufficient condition for  $\phi_n$  to qualify as rate is

$$\frac{\phi_n}{\chi_n}\overline{y}_n = O(1). \tag{B.49b}$$

**B.2.2.1 Stable case.** Since  $\overline{y}_n = O(1)$ , both  $\phi_n = \varphi(n)$  and  $\phi_n = \psi(n)$  from Theorem 2(i) will do according to the dichotomy established there.

**B.2.2.2 Explosive case.** By Theorem 2(iii),  $\chi_n = |\lambda|^n/n^{1/2+\eta}$ . Since  $n\lambda^{-n}\overline{y}_n \rightarrow \kappa$ , cf. (B.17a),

$$\phi_n = n^{1/2-\eta}$$

will do for every  $\eta \geq 0$ :

$$\frac{\phi_n}{\chi_n}\overline{y}_n = n^{1/2-\eta}\frac{n^{1/2+\eta}}{|\lambda|^n}\frac{\lambda^n}{n}(\kappa + o(1)) = O(1),$$

which shows (B.49b). (B.49a) is trivially satisfied.

### B.2.2.3 Unit root case.

Case 1:  $\lambda = 1, \mu \neq 0$ .

Since  $\overline{y}_n/n \rightarrow \mu/2$ , cf. (B.20b),  $\phi_n = \psi_n$  will do:

$$\frac{\phi_n}{\chi_n}\overline{y}_n = \sqrt{\frac{n}{\ln_2 n}}\sqrt{\frac{\ln_2 n}{n^3}}\frac{n}{2}\mu(1 + o(1)) = O(1).$$

If only 2nd moments exist,

$$\phi_n = \sqrt{\frac{n}{(\ln n)^{1+\eta}}}. \tag{B.50}$$

Case 2:  $\lambda = -1$ .

The optimal rate  $\chi_n = n/(\ln_2 n)^3$  from Theorem 2(iib) does not satisfy (B.49a). However, since  $\overline{y}_n = O(1)$ , (B.50) will do.

Gathering the results, we arrive at Corollary 1, in the notation used in this appendix.

**Corollary 1.** *Strong consistency of the OLS estimator  $\hat{\mu}_T$  of the intercept  $\mu$  holds at the following rates.*

(i) *Stable case: If  $\mathbf{E}|\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ ,*

$$\sqrt{\frac{n}{\ln_2 n}} (\hat{\mu}_n - \mu) = O(1).$$

*If only second moments exist, then*

$$\sqrt{\frac{n}{(\ln n)^{1+\eta}}} (\hat{\mu}_n - \mu) = o(1).$$

*for all  $\eta > 0$ .*

(iia) *Unit root case:  $\lambda = 1$  and  $\mu \neq 0$ . If  $\mathbf{E}|\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ ,*

$$\sqrt{\frac{n}{\ln_2 n}} (\hat{\mu}_n - \mu) = O(1).$$

*If only second moments exist, then*

$$\sqrt{\frac{n}{(\ln n)^{1+\eta}}} (\hat{\mu}_n - \mu) = o(1).$$

*for all  $\eta > 0$ .*

(iib) *Unit root case:  $\lambda = -1$ . Same as in case (iia).*

(iii) *Explosive case: Assuming only 2nd moments,*

$$n^{1/2-\eta} (\hat{\mu}_n - \mu) = o(1). \tag{B.51}$$

*for all  $\eta > 0$ . If  $\mathbf{E}|\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ , (B.51) remains valid, with  $O(1)$  instead of  $o(1)$  for  $\eta = 0$ .*

### B.2.3 Stable case revisited

The question whether in some special cases a LIL-type convergence theorem may be available despite the nonexistence of higher moments seems to have a positive answer at least in the stable case. Basically, it is the problem of finding LIL-like laws for martingales of the form

$$M_n = \sum_{k=1}^n y_{k-1} \varepsilon_k$$

if only second moments for  $\varepsilon_n$  are available. We cite here one reference which is appropriate for our purposes, namely the following version of a LIL for stationary ergodic processes due to Stout (1970).

**Result 2 (Stout (1970)).** *Let  $(Y_i)_{i \geq 1}$  be a stationary ergodic stochastic sequence with  $\mathbf{E}\{Y_i | Y_1, Y_2, \dots, Y_{i-1}\} = 0$  a.s. for all  $i \geq 2$  and  $\mathbf{E}Y_1^2 = 1$ . Then, with probability one,*

$$\limsup_{n \rightarrow \infty} \frac{\sum_{i=1}^n Y_i}{\sqrt{2n \ln_2 n}} = 1. \tag{B.52}$$

In our stable case:  $Y_i = y_{i-1}^0 \varepsilon_i$ , where  $y_n^0$  is the *stationary* solution to (B.1). Note that the stationary solution is used in order to ensure that the resulting process  $(Y_i)$  is stationary ergodic. Then  $\mathbf{E}Y_i^2 = \sigma^4 / (1 - \lambda^2)$ . Since  $y_{n-1}^0$  and  $\varepsilon_n$  are independent, only  $\varepsilon_n \in L^2$  needs to be required, cf. Shiryaev (1996, Chapter II, §6, Theorem 6).

(B.52) then ensures that MCT 2 holds without any further assumptions:

$$\limsup_{n \rightarrow \infty} \frac{\sum_{k=1}^n y_{k-1}^0 \varepsilon_k}{\sqrt{2n \ln_2 n}} = \frac{\sigma^2}{\sqrt{1 - \lambda^2}}. \quad (\text{B.53})$$

Actually, (B.53) is valid for *any* (possibly nonstationary) solution to (B.1). This is so since the difference between the two solutions of (B.1) is given by

$$y_n - y_n^0 = \lambda^n u$$

with  $u = y_0 - y_0^0$ , implying that

$$\sum_{k=1}^n y_{k-1} \varepsilon_k - \sum_{k=1}^n y_{k-1}^0 \varepsilon_k = u \sum_{k=1}^n \lambda^{k-1} \varepsilon_k = O(1).$$

Finally, passing to  $(-Y_i)$  in (B.52) shows the same behaviour for the  $\liminf_{n \rightarrow \infty}$  or, put in a form needed for the OLS estimator,

$$\limsup_{n \rightarrow \infty} \frac{|\sum_{k=1}^n y_{k-1} \varepsilon_k|}{\sqrt{2n \ln_2 n}} = \frac{\sigma^2}{\sqrt{1 - \lambda^2}}.$$

This is just a form of the MCT 2 needed to calculate the rate of convergence in Theorem 2(i). Hence (B.46) remains valid even if *only 2nd moments are required*.

### B.3 Proof of Theorem 3

Our starting point will be Nielsen (2005, Theorem 2.4), which we cite here because it is of interest in its own right.

**Result 3 (Nielsen (2005)).** *Assume that  $\mathbf{E}|\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ . Then, except for the case of  $\lambda = 1 \wedge \mu = 0$ , the following holds with probability one:*

$$\tau_n = \begin{cases} O[(\ln_2 n)^{1/2}] & \text{for } |a| < 1, \\ O[(\ln n)^{1/2}] & \text{for } |a| = 1, \\ o[n^\rho] & \text{for } |a| > 1, \end{cases} \quad (\text{B.54})$$

with the last line being valid for all  $\rho > 1/\alpha$ .

The exclusion of the aforementioned case is due to the fact that it violates Assumption 2.3 of Nielsen (2005).

As elaborated in Section 3.1.2, this approach comes down to investigating the asymptotic behaviour of the minimal eigenvalues  $\lambda_n = \lambda_{\min}(M_n)$  of the moment matrix  $M_n$  and to find sequences of numbers  $\chi_n$  s.t.

$$\chi_n \frac{\|\tau_n\|}{\sqrt{\lambda_n}} = O(1). \quad (\text{B.55})$$

The minimal eigenvalues  $\lambda_n$  of  $M_n$  are calculated in Section B.1.2. In the consideration of the individual cases below we will therefore only cite the corresponding result without further explicit reference.

### B.3.1 Stable case

In this case,

$$\frac{1}{n}\lambda_n \rightarrow \lambda_-$$

for some positive constant  $\lambda_-$ . Taking account of (B.54), we obtain

$$\frac{\|\tau_n\|}{\sqrt{\lambda_n}} = \sqrt{\frac{\ln_2 n}{n}} O(1),$$

so that

$$\chi_n = \sqrt{\frac{n}{\ln_2 n}}$$

will satisfy (B.55).

**Remark B.7.** *In the stable case, both eigenvalues diverge at the same rate, so that both components of  $\hat{\theta}_n$  will have the same rate of convergence.*

### B.3.2 Explosive case

In this case,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \lambda_{\min} = 1.$$

Going back to (B.54),

$$\frac{\|\tau_n\|}{\sqrt{\lambda_n}} = \frac{n^\rho}{\sqrt{n}} O(1).$$

As a consequence,

$$\chi_n = n^{1/2-\rho}$$

will satisfy (B.55).

### B.3.3 Unit root case

Case 1:  $\lambda = 1, \mu \neq 0$ .

It follows that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \lambda_{\min} = \frac{1}{4}$$

and, by (B.54),

$$\frac{\|\tau_n\|}{\sqrt{\lambda_n}} = \sqrt{\frac{\ln n}{n}} O(1).$$

As a consequence,

$$\chi_n = \sqrt{\frac{n}{\ln n}}$$

will satisfy (B.55).

Case 2:  $\lambda = 1, \mu = 0$ .

In this case Assumption 2.3 in Nielsen (2005) is violated, so that we cannot apply Result 3. Nevertheless, the results for Case 2 in Section B.1.2.3 in connection with (B.23a) show that

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} = O(1).$$

Therefore the fundamental sufficient condition of Lai & Wei (1982a) is satisfied and therefore their Theorem 1 shows that

$$\sqrt{\frac{\lambda_{\min}}{\ln \lambda_{\max}}} (\hat{\theta}_n - \theta) = O(1).$$

For more details cf. the discussion in Section B.4. Since  $\ln A_n^0 / \ln n = O(1)$ , see (B.32), it follows that

$$\sqrt{\frac{n}{\ln n}} (\hat{\theta}_n - \theta) = O(1).$$

Case 3:  $\lambda = -1$ .

Since

$$\lim_{n \rightarrow \infty} \frac{1}{n} \lambda_{\min} = 1$$

and  $\|\tau_n\| = O(n^\rho)$ ,  $\chi_n$  may be chosen as in case  $\lambda = 1$ .

Gathering the results we obtain the following theorem.

**Theorem 3.** *Assume that  $\mathbf{E}|\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ . Then strong consistency of the joint OLS estimator  $\hat{\theta}_n$  holds at the following rates.*

(i) *Stable case:*

$$\sqrt{\frac{n}{\ln_2 n}} (\hat{\theta}_n - \theta) = O(1).$$

(ii) *Unit root case:* For both  $\lambda = 1$  and  $\lambda = -1$ ,

$$\sqrt{\frac{n}{\ln n}} (\hat{\theta}_n - \theta) = O(1).$$

(iii) *Explosive case:*

$$n^{1/2-\rho} (\hat{\theta}_n - \theta) = o(1)$$

for every  $\rho > 1/\alpha$ .

**Remark B.8.** *One may be tempted to apply the results in Lai & Wei (1985) by writing the AR(1)-model with intercept in augmented form as*

$$x_n = Bx_{n-1} + e_n,$$

with

$$x_n = \begin{pmatrix} y_n \\ 1 \end{pmatrix}, \quad B = \begin{pmatrix} \lambda & \mu \\ 0 & 1 \end{pmatrix}, \quad e_n = \begin{pmatrix} \varepsilon_n \\ 0 \end{pmatrix}.$$

For the case of  $\lambda = 1$ ,  $B$  has the double root  $\lambda = 1$ . A formal application of Theorem 3 loc. cit. would then yield that  $\liminf_{n \rightarrow \infty} n^{-1} \lambda_{\min} > 0$ . However, Lai & Wei's crucial assumption (3.11) (with  $p = 2$ ) is not satisfied. Namely,

$$E_n = \frac{1}{n} \sum_{i=1}^n e_i e_i' = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

so that

$$E_n + B E_n B' = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 \left[ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \lambda^2 & 0 \\ 0 & 0 \end{pmatrix} \right] \rightarrow \sigma^2 \begin{pmatrix} 1 + \lambda^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

The limit matrix is singular, thus violating the requirement that it be regular.



## B.4 Comparison with Lai & Wei

It will be of interest to compare our results with those that can be obtained on the basis of the seminal paper by Lai & Wei (1982a). They consider the general regression model

$$y_n = \theta' x_n + \varepsilon_n$$

with predetermined regressors  $x_n$  and martingale difference errors  $\varepsilon_n$ . Specialised to our case of i.i.d. errors it is assumed that  $\mathbf{E} |\varepsilon_n|^\alpha < \infty$  for some  $\alpha > 2$ . Lai & Wei's basic Theorem 1 states that a sufficient condition for strong consistency of the joint OLS estimator  $\hat{\theta}_n$  is that

$$\ln \lambda_{\max}(n) = o(\lambda_{\min}(n)), \quad (\text{B.56})$$

where  $\lambda_{\max}(n)$  and  $\lambda_{\min}(n)$  denote again the maximal and the minimal eigenvalue, respectively, of the moment matrix

$$M_n = X_n' X_n,$$

where  $X_n' = (x_1, \dots, x_n)$ . In this case, it will hold that

$$\|\hat{\theta}_n - \theta\| = O\left(\sqrt{\frac{\ln \lambda_{\max}(n)}{\lambda_{\min}(n)}}}\right). \quad (\text{B.57})$$

Applied to our scenario,  $x_n = (1, y_{n-1})'$  and  $\theta = (\mu, \lambda)'$ , so that  $M_n$  is just the matrix in (B.40). Its eigenvalues have been calculated in Section B.1.2, and we will make use of these results to check condition (B.56) for the various cases.

### B.4.1 Stable case

In this case, (B.56) is satisfied and

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} = \frac{\ln n}{n} (1 + o(1)).$$

Hence (B.56) is satisfied and

$$\chi_n = \sqrt{\frac{n}{\ln n}}$$

is a rate corresponding to (B.57). This rate is slower than that in Corollary 1 and Theorem 3.

### B.4.2 Explosive case

In this case,

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} \rightarrow 2 \ln |\lambda|.$$

Hence (B.56) is violated. Nevertheless, it is shown in Theorems 2 and 3 that the OLS estimator is strongly consistent. This shows that condition (B.56) is indeed not necessary. The explosive case of our model may hence be seen as a counterpart to Lai & Wei's Example 1.

**Remark B.9.** *Note that in the explosive case the condition*

$$\frac{A_n}{\ln n} \rightarrow \infty$$

*in Lai & Wei (1982b), valid for simple regression models, is satisfied in view of the result in Section B.1.1.2.*

### B.4.3 Unit root case

Case 1:  $\lambda = 1$ .

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} = O\left(\frac{\ln n}{n}\right).$$

Therefore (B.56) is satisfied and a corresponding rate is

$$\chi_n = \sqrt{\frac{n}{\ln n}}.$$

This is the rate appearing in Theorem 3(ii). Given that the error terms have a moment somewhat higher than the second, it is weaker than the corresponding rate in Corollary 1.

Case 2:  $\lambda = -1$ .

It was shown in Section B.1.2.3 that

$$\lambda_{\max} = \left(n + A_n^0\right) (1 + o(1)) = A_n^0 \left(1 + \frac{n}{A_n^0}\right) (1 + o(1)).$$

Hence, since  $A_n^0/n \rightarrow \infty$ ,

$$\ln \lambda_{\max} = \ln A_n^0 + o(1).$$

Making use of (B.32), we find that

$$\frac{\ln \lambda_{\max}}{\ln n} = \frac{\ln A_n^0}{\ln n} + o(1) = O(1).$$

Since  $\lambda_{\min} = n(1 + o(1))$ , (B.56) is satisfied.

**Remark B.10.** *Lai & Wei (1982a, Theorem 2) may be applied to both cases, showing that the rate for the joint estimator  $\hat{\theta}_T$  is  $\sqrt{n/\ln n}$ .*

## C Proof of Theorem 4

### C.1 Generalities

As in Appendix A, we will make the calculations in terms of the centred process  $a_t^\# = a_t - \alpha$ , for which the corresponding dynamics and the DGP are given by (A.1) and (A.2). As is readily seen, the OLS estimator  $\hat{\beta}_T$  is the same whether calculated with the original  $a_t$  or the transformed  $a_t^\#$ . Using again the convention of renaming  $a_t^\#$  as  $a_t$ , we are thus from now on working with the DGP

$$y_t = \alpha + \beta a_{t-1} + \varepsilon_t, \tag{C.1}$$

and the dynamics

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

Note that, with this notational convention,  $\lim_{T \rightarrow \infty} a_t = 0$ .

Our procedure will be the same as in Section B.2. In particular, the OLS estimator of  $\beta$  is

$$\hat{\beta}_T - \beta = \frac{\sum_{t=1}^T (a_{t-1} - \bar{a}_T) \varepsilon_t}{A_T} = R_T - V_T,$$

with

$$\begin{aligned} \bar{a}_T &= \frac{1}{T} \sum_{t=1}^T a_{t-1}, & A_T^0 &= \sum_{t=1}^T a_{t-1}^2, \\ A_T &= \sum_{t=1}^T (a_{t-1} - \bar{a}_T)^2 = A_T^0 - T (\bar{a}_T)^2 \end{aligned} \tag{C.3}$$

and

$$\begin{aligned} u_T &= \sum_{t=1}^T a_{t-1} \varepsilon_t, \\ R_T &= \frac{u_T}{A_T}, \quad V_T = \frac{\bar{a}_T}{A_T} \sum_{t=1}^T \varepsilon_t, \end{aligned} \tag{C.4}$$

cf. (B.38a). As in Section B.2, the main tool in analysing  $R_T$  and  $V_T$  will be the martingale convergence theorems stated in Section B.1. For the reader's convenience we restate them here in the notation for the process  $a_t$ .

**MCT 1 (Chow (1965)).**

$$\sum_{t=1}^T a_{t-1} \varepsilon_t = o \left[ \sqrt{A_T^0 (\ln A_T^0)^{1+\eta}} \right] \tag{C.5a}$$

for all  $\eta > 0$ . If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ , this may be sharpened to

$$\sum_{t=1}^T a_{t-1} \varepsilon_t = O \left[ \sqrt{A_T^0 \ln A_T^0} \right]. \tag{C.5b}$$

**MCT 2 (Wei (1985)).** If, in addition to  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ ,

$$a_T^2 = o \left[ (A_T^0)^\gamma \right] \tag{C.6}$$

for some  $0 < \gamma < 1$ , then

$$\sum_{t=1}^T a_{t-1} \varepsilon_t = O \left[ \sqrt{A_T^0 \ln_2 A_T^0} \right].$$

Arguing as in Section B.2.1.1, we rewrite  $R_T$  in the form

$$R_T = \frac{u_T}{A_T} = \chi^{-1} \left( A_T^0 \right) U_T \frac{A_T^0}{A_T} \tag{C.7}$$

with

$$(x) = \varphi(x) = \sqrt{\frac{x}{(\ln x)^{1+\eta}}} \quad \text{or} \quad \chi(x) = \psi(x) = \sqrt{\frac{x}{\ln_2 x}}, \tag{C.8}$$

$$U_T = \frac{u_T}{\sqrt{A_T^0 (\ln A_T^0)^{1+\eta}}} \quad \text{or} \quad U_T = \frac{u_T}{\sqrt{A_T^0 \ln_2 A_T^0}}, \tag{C.9}$$

respectively, according to whether MCT 1 or MCT 2 is applicable.

Following the same logic as in Section B.2.1.1, in order to obtain deterministic convergence rates for the OLS estimator, our task is then to find sequences of numbers  $\chi_T$  such that

$$\frac{\chi_T}{\chi(A_T^0)} = O(1) \tag{C.10}$$

and

$$\chi_T \sqrt{T \ln_2 T} \frac{\bar{a}_T}{A_T} = O(1) \tag{C.11}$$

are satisfied. Then, if in addition

$$A_T^0 / A_T = O(1), \tag{C.12}$$

it will hold that

$$\chi_T (\widehat{\beta}_T - \beta) = O(1). \tag{C.13}$$

(C.10) to (C.12) are just the conditions (B.44) and (B.45) in Section B.2.1.1.

## C.2 Asymptotics of $A_T^0$

### C.2.1 $c > 1/2$

We will show that

$$\frac{A_T^0}{\ln T} \rightarrow \frac{\gamma^2 \sigma^2}{2c-1} \quad (\text{C.14})$$

Taking squares in (C.2) (remembering our renaming convention) yields

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

Multiplying by  $t$  and then summing over  $t$  we obtain

$$\sum_{t=1}^T t a_t^2 = \sum_{t=1}^T t \left(1 - \frac{c}{t}\right)^2 a_{t-1}^2 + \gamma^2 \sum_{t=1}^T \frac{1}{t} \varepsilon_t^2 + 2\gamma \sum_{t=1}^T \left(1 - \frac{c}{t}\right) a_{t-1} \varepsilon_t. \quad (\text{C.15})$$

Since

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

we may write the first term in (C.15) as

$$\begin{aligned} \sum_{t=1}^T t \left(1 - \frac{c}{t}\right)^2 a_{t-1}^2 &= \sum_{t=1}^T t a_{t-1}^2 + \left[ -2c \sum_{t=1}^T a_{t-1}^2 + c^2 \sum_{t=1}^T \frac{1}{t} a_{t-1}^2 \right] \\ &= \sum_{t=1}^T (t-1) a_{t-1}^2 + \sum_{t=1}^T a_{t-1}^2 + [\dots] \\ &= \sum_{t=1}^T t a_t^2 - T a_T^2 + A_T^0 + \left[ -2c A_T^0 + c^2 \sum_{t=1}^T \frac{1}{t} a_{t-1}^2 \right] \\ &= \sum_{t=1}^T t a_t^2 - T a_T^2 + (1-2c) A_T^0 + c^2 \sum_{t=1}^T \frac{1}{t} a_{t-1}^2. \end{aligned}$$

Returning to (C.15), this means that

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

Solving for  $A_T^0$  yields

$$(2c-1) A_T^0 = T a_T^2 + c^2 \sum_{t=1}^T \frac{1}{t} a_{t-1}^2 + \gamma^2 \sum_{t=1}^T \frac{1}{t} \varepsilon_t^2 + 2\gamma u_T - 2\gamma c \sum_{t=1}^T \frac{1}{t} a_{t-1} \varepsilon_t \quad (\text{C.16})$$

(remember (C.4)). Now bring in the asymptotic behaviour of  $a_t$  established in Theorem 1(i):

$$a_t = O\left(\sqrt{\frac{\ln_2 t}{t}}\right), \quad (\text{C.17})$$

to analyse the individual terms on the right hand side of (C.16). The following properties all hold with probability one:

(i)

$$T a_T^2 = O(\ln_2 T).$$

(ii)

$$\sum_{t=1}^T \frac{1}{t} a_{t-1}^2 = O(1) \sum_{t=1}^T \frac{\ln_2 t}{t^2} = O(1).$$

(iii)

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

This follows from the decomposition  $\nu_t = \varepsilon_t^2 - \sigma^2$ , applying Kolmogorov's LLN to  $\nu_t$ :

$$\sum_{t=1}^T \frac{1}{t} \varepsilon_t^2 = \sum_{t=1}^T \frac{1}{t} \sigma^2 + \sum_{t=1}^T \frac{\nu_t}{t} = \sigma^2 \ln T + O(1).$$

(iv)

$$\sum_{t=1}^T \frac{1}{t} a_{t-1} \varepsilon_t = O(1)$$

by the martingale convergence theorem.

Hence

$$(2c-1) A_T^0 = \gamma^2 \sigma^2 \ln T + O(\ln_2 T) + 2\gamma u_T. \quad (\text{C.18})$$

Noting that  $\langle u \rangle_T = \sigma^2 A_T^0$ , we then argue as follows. Suppose that  $A_\infty^0 < \infty$  on some set  $\Gamma$  of positive measure. Then, by the martingale convergence theorem,  $u_T$  converges a.s. on  $\Gamma$  to some finite limit. Dividing (C.18) by  $A_T^0$ , we obtain

$$\begin{aligned} (2c-1) &= \gamma^2 \sigma^2 \frac{\ln T}{A_T^0} + \frac{O(\ln_2 T)}{A_T^0} + O(1) \\ &= \gamma^2 \sigma^2 \frac{\ln T}{A_T^0} \left[ 1 + O\left(\frac{\ln_2 T}{\ln T}\right) \right] + O(1). \end{aligned}$$

On  $\Gamma$ , the right hand side converges to  $\infty$ , which is impossible since the left hand side is finite. As a consequence,  $A_\infty^0 = \infty$  with probability one. Again from the martingale convergence theorem (now the version for martingales with unbounded bracket process) it then follows that

$$\frac{u_T}{A_T^0} \rightarrow 0$$

Dividing (C.18) by  $A_T^0$  we now obtain

$$(2c-1) = \gamma^2 \sigma^2 \frac{\ln T}{A_T^0} [1 + o(1)] + o(1).$$

This shows that

$$\frac{A_T^0}{\ln T} \rightarrow \frac{\gamma^2 \sigma^2}{2c-1} \quad (\text{C.19})$$

with probability one. Going back to (C.3) and noting that

$$\bar{a}_T = O\left(\sqrt{\frac{\ln_2 T}{T}}\right), \quad (\text{C.20})$$

cf. (A.14), we find that

$$\frac{A_T}{A_T^0} = 1 + O\left(\frac{\ln_2 T}{\ln T}\right) \quad (\text{C.21})$$

and

$$\frac{\bar{a}_T}{A_T} = O\left(\frac{1}{\ln T} \sqrt{\frac{\ln_2 T}{T}}\right). \quad (\text{C.22})$$

Finally, let us consider condition (C.6). In view of (C.17) and (C.19),

$$\frac{a_T^2}{(A_T^0)^\gamma} = O\left(\frac{\ln_2 T}{T(\ln T)^\gamma}\right) = o(1) \quad (\text{C.23})$$

for all  $\gamma$ .

### C.2.2 $c < 1/2$

Making use of Theorem 1 (iii) and the result in (A.21) in Section A.3 it can easily be calculated that

$$A_T^0 = T^{1-2c} \frac{u^2}{1-2c} (1 + o(1)) \quad (\text{C.24a})$$

$$A_T = T^{1-2c} v^2 (1 + o(1)) \quad (\text{C.24b})$$

with  $v^2 = c^2 u^2 / (1-c)^2 ((1-2c))$ , and

$$\lim_{T \rightarrow \infty} \frac{A_T^0}{A_T} = \left(1 - \frac{1}{c}\right)^2. \quad (\text{C.25})$$

Also, making use again of (A.21) together with (C.24b), it turns out that

$$\frac{\bar{a}_T}{A_T} = \frac{1}{T^{1-c}} w (1 + o(1)) \quad (\text{C.26})$$

with  $w \neq 0$  a.s.. Finally, let us consider condition (C.6). In view of Theorem 1 (iii) and (C.24a),

$$\frac{a_T^2}{(A_T^0)^\gamma} = O\left(\frac{T^{-2c}}{T^{(1-2c)\gamma}}\right) = O\left[T^{-2c-(1-2c)\gamma}\right].$$

Hence (C.6) is fulfilled for all  $\gamma > 0$ .

## C.3 Consistency

### C.3.1 $c > 1/2$

We go back to the procedure announced in Section C.1. Straightforward calculation shows that

$$\varphi(A_T^0) = \sqrt{\frac{A_T^0}{(\ln A_T^0)^{1+\eta}}} = \sqrt{\frac{r \ln T}{(\ln_2 T)^{1+\eta}}} (1 + o(1)),$$

$$\psi(A_T^0) = \sqrt{\frac{A_T^0}{\ln_2 A_T^0}} = \sqrt{\frac{r \ln T}{\ln_3 T}} (1 + o(1)),$$

with  $r = \gamma^2 \sigma^2 / (2c - 1) > 0$ . In view of (C.10) this yields as candidates for the normalising sequences  $\chi_T = \varphi_T = \varphi(\ln T)$  or  $\chi_T = \psi_T = \psi(\ln T)$ . As to (C.11), it follows from (C.22) that

$$\varphi_T \sqrt{T \ln_2 T} \frac{\bar{a}_T}{A_T} = \sqrt{\frac{\ln T}{(\ln_2 T)^{1+\eta}}} \sqrt{T \ln_2 T} \frac{1}{\ln T} \sqrt{\frac{\ln_2 T}{T}} O(1) = O\left(\sqrt{\frac{(\ln_2 T)^{1-\eta}}{\ln T}}\right)$$

and

$$\psi_T \sqrt{T \ln_2 T} \frac{\bar{a}_T}{A_T} = \sqrt{\frac{\ln T}{\ln_3 T}} \sqrt{T \ln_2 T} \frac{1}{\ln T} \sqrt{\frac{\ln_2 T}{T}} O(1) = O\left(\sqrt{\frac{(\ln_2 T)^2}{\ln T \ln_3 T}}\right).$$

Hence condition (C.11) is satisfied for both choices of the normalising sequence  $\chi_T$ . Condition (C.12) is satisfied by virtue of (C.21).

Summarising, we arrive at the following conclusions:

- (i) If  $\varepsilon_t$  has moments up to second order, then the rate of a.s. convergence of the OLS estimator is  $\varphi_T = \sqrt{\ln T / (\ln_2 T)^{1+\eta}}$  for every  $\eta > 0$ .
- (ii) If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ , then also  $\eta = 0$  will do. However, since condition (C.6) is satisfied, we may apply MCT 2 to obtain  $\psi_T = \sqrt{\ln T / \ln_3 T}$  as a normalising sequence.

### C.3.2 $c < 1/2$

From the results in Section C.2.2 it readily follows that

$$\begin{aligned} \varphi(A_T^0) &= \sqrt{\frac{A_T^0}{(\ln A_T^0)^{1+\eta}}} = w \sqrt{\frac{T^{1-2c}}{(\ln T)^{1+\eta}}} (1 + o(1)), \\ \psi(A_T^0) &= \sqrt{\frac{A_T^0}{\ln_2 A_T^0}} = w' \sqrt{\frac{T^{1-2c}}{\ln_2 T}} (1 + o(1)) \end{aligned}$$

for some positive random variables  $w$  and  $w'$ . Hence the deterministic sequences

$$\varphi_T = \sqrt{\frac{T^{1-2c}}{(\ln T)^{1+\eta}}} \quad \text{and} \quad \psi_T = \sqrt{\frac{T^{1-2c}}{\ln_2 T}}$$

both qualify as candidates for the normalisation of the OLS estimator, in the sense that they satisfy (C.10). Condition (C.12) is fulfilled in view of (C.25). It remains to verify (C.11). By (C.26),

$$\psi_T \sqrt{T \ln_2 T} \frac{\bar{a}_T}{A_T} = \sqrt{\frac{T^{1-2c}}{\ln_2 T}} \sqrt{T \ln_2 T} \frac{1}{T^{1-c}} O(1) = O\left(\frac{1}{\sqrt{T^{1-c}}}\right).$$

Similarly for  $\varphi_T$ . Finally, as shown above, condition (C.6) is also fulfilled.

Summarising, we arrive at the following conclusions:

- (i) If  $\varepsilon_t$  has moments up to second order, then the rate of a.s. convergence of the OLS estimator is  $\varphi_T = \sqrt{T^{1-2c} / (\ln T)^{1+\eta}}$  for every  $\eta > 0$ .
- (ii) If  $\mathbf{E} |\varepsilon_t|^\alpha < \infty$  for some  $\alpha > 2$ , then also  $\eta = 0$  will do. However, since condition (C.6) is satisfied, we may apply MCT 2 to obtain  $\psi_T = \sqrt{T^{1-2c} / \ln_2 T}$  as a normalising sequence.

## C.4 Comparison with Lai & Wei

As in Section B, it may be of some interest to check the Lai & Wei condition (B.56). Going back to Section B.1.2, the formulas for the two eigenvalues are given by

$$\lambda_{\pm} = \frac{T + A_T^0}{2} \left[ 1 + \sqrt{1 - 4D_T} \right],$$

with

$$D_T = \frac{TA_T^0 - (T\bar{a}_T)^2}{(T + A_T^0)^2}.$$

**C.4.1**  $c > 1/2$

By (C.19) and (C.20),

$$\begin{aligned} A_T^0 &= (1 + o(1)) r \ln T, \\ \left(T \bar{a}_T\right)^2 &= \frac{\ln_2 T}{T} O(1), \end{aligned}$$

with  $r = \gamma^2 \sigma^2 / (2c - 1)$ . Straightforward calculations show that

$$T + A_T^0 = T(1 + o(1))$$

and

$$\begin{aligned} D_T &= r \frac{\ln T}{T} (1 + o(1)), \\ \left(T + A_T^0\right) D_T &= (1 + o(1)) r \ln T. \end{aligned}$$

Hence, since  $D_T \rightarrow 0$ ,

$$\lambda_{\max} = T(1 + o(1)), \quad \ln \lambda_{\max} = (1 + o(1)) \ln T.$$

For  $\lambda_{\min}$ , the expansion (B.36) yields

$$\lambda_{\min} = (1 + o(1)) r \ln T.$$

As a consequence,

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} \rightarrow r^{-1}.$$

Thus Lai & Wei's condition (B.56) is violated. Yet, the OLS estimator is strongly consistent.

**Remark C.1.** *Note that by virtue of (C.19) and (C.21)*

$$\frac{A_T}{\ln T} \rightarrow r. \tag{C.27}$$

*Therefore the consistency condition of Lai & Wei (1982b) mentioned above in Remark B.9 is not satisfied.*

**C.4.2**  $c < 1/2$

In this case, making use of (C.24a) and (A.21), it turns out that

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} \sim \frac{\ln T}{T^{1-2c}} \kappa$$

for some finite positive random variable  $\kappa$ , so that the Lai & Wei condition in (B.56) is satisfied.