Electronic Journal of Statistics
Vol. 13 (2019) 1646–1693
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1558

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

Norbert Christopeit*

University of Bonn, Adenauerallee 24–42, 53113 Bonn, Germany

Michael Massmann*

WHU – Otto Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany,
and
Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands.
e-mail: michael.massmann@whu.edu

Abstract: This paper looks at the strong consistency of the ordinary least squares (OLS) estimator in linear regression models with adaptive learning. It is a companion to Christopeit & Massmann (2018) 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 (1982a), is not always met.

MSC 2010 subject classifications: Primary 62F10, 62H12, 62J05, 62M10; secondary 91A26, 91B64, 91B84.

Keywords and phrases: Adaptive learning, almost sure convergence, non-stationary regression, ordinary least squares.

Received August 2018.

*We are very grateful to Bent Nielsen for his reading of an earlier version of this article. The paper has also benefited from discussions at seminars at the University of Bonn and at Erasmus University Rotterdam. Finally, we would like to thank the editor and two anonymous referees for their comments and suggestions. All remaining deficiencies or errors are of course ours.

Contents

1. Introduction

This paper looks at the strong consistency of the ordinary least squares (OLS) estimator in a linear model whose regressors are generated by an adaptive learning recursion. In particular, interest lies on the estimation of what we call the structural parameters δ and β in the model

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

where the index $t = 1, 2, \ldots$, the explanatory variable is generated recursively by

$$a_t = a_{t-1} + \gamma_t \left(y_t - a_{t-1} \right) \tag{2}$$

and the error term ε_t is specified below. Of central importance in this model is the so-called weighting, or gain, sequence γ_t which governs the extent to which the previous value a_{t-1} of the regressor is updated, or learned, in the light of its deviation from the present realisation of y_t. We examine two specifications of the gain sequence: For a known parameter $\gamma > 0$,

$$\gamma_t = \begin{cases} \gamma \\ \gamma/t. \end{cases} \tag{3}$$

The former is referred to as *constant* gain, the latter is an instance of a *decreasing* gain sequence since $\gamma_t \to 0$.

The model in (1)-(3) can be seen as a special case of the structural model $y_t = \beta a_{t-1} x_t + \delta x_t + \varepsilon_t$ where x_t is some exogenous covariate and the learning recursion of a_t is given by a stochastic approximation algorithm, see Lai (2003) for an overview. Models of this class have been particularly prominent in the recent macroeconomic literature on bounded rationality which interprets $a_{t-1} x_t = y^e_{t|t-1}$, say, as the expectations economic agents form about y_t by estimating at time $t-1$ the so-called rational expectations equilibrium (REE) $y_t = \alpha x_t + \varepsilon_t$, where

$$\alpha = \frac{\delta}{1-\beta}, \tag{4}$$

cf. Sargent (1993, 1999) and Evans & Honkapohja (2001). In the present paper, we effectively assume $x_t = x$ to be constant and thus focus on (1)-(3) in order to keep the analysis tractable.

Examining the strong consistency of the OLS estimators of β and δ in model (1)-(3) is of interest for two reasons: First, the empirical estimation of adaptive learning models has recently gained popularity amongst researchers and policy makers; see, for instance, Milani (2007) and Chevillon, Massmann & Mavroeidis (2010), Malmendier & Nagel (2016) and Adam, Marcet & Nicolini (2016). Yet, secondly, 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 Christopeit & Massmann (2018), hereafter referred to as CM18, is one of the first comprehensive attempts to examine the asymptotic behaviour of an econometric estimation procedure in an adaptive learning model. There, we derive the limiting distributions of the OLS estimator of δ and β in model (1)-(3) in both the constant and decreasing gain setting. In particular, it is shown that the OLS estimator is weakly consistent, although its asymptotic distribution may be highly non-standard. In contrast, in the present paper we look at strong consistency of the OLS estimator. More ambitiously, our interest lies on rates rather than on the mere fact of convergence.

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 γ_t. It is well known, see e.g. Benveniste, Métivier & Priouret (1990), that with constant gain learning, a_t effectively estimates α in (4) by exponential smoothing and $a_t \nrightarrow \alpha$ in general. As opposed to that, with decreasing gain learning, a_t is a generalised recursive least squares estimator and the convergence $a_t \to \alpha$ does hold with probability one if $\sum_t \gamma_t = \infty$ but $\sum_t \gamma_t^2 \ln^2 t < \infty$, provided that $\beta < 1$, cf. Kottmann (1990) for details. Importantly, 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.

The model in (1) is a linear regression model with predetermined stochastic regressors. There is an extensive literature on parameter estimation in this model class. 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 Christopeit (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 Christopeit & 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 β and δ always obtains.

For constant gain learning, the model to be estimated is basically an autoregressive model of order one with a constant term. Most of the literature on the strong consistency of the OLS estimator in general autoregressions considers models without an intercept, cf., e.g., Lai & Wei (1983a) and Lai & Wei (1985). As will be seen, however, the existence of an unknown intercept can make a considerable difference to the analysis. In particular, we consider the rates of convergence of the *separate* OLS estimators of β and δ. These are compared to the speed of convergence of the norm of the *joint*, i.e. bivariate, OLS estimator of $\theta = (\beta, \delta)$. That part of the analysis builds on a recent treatment by Nielsen (2005) of OLS estimation in vector autoregressive models with general deterministic terms.

For decreasing gain learning, it is interesting to note that the asymptotic second moment matrix is singular. This is due to the fact that the regressor a_t converges a.s. to the constant α. This violation of the so-called Grenander condition may affect the rates of weak convergence of the OLS estimator, see Phillips (2007) and CM18. Yet it does not pose any problem for a.s. convergence.

Reconsider the model in (1)-(3). We will frequently work with an alternative representation of the dynamics of a_t: Substitute (1) into (2) to obtain

$$a_t = (1 - c_t)\, a_{t-1} + \gamma_t\, (\delta + \varepsilon_t)\,, \tag{5}$$

where we have defined $c_t = (1 - \beta)\,\gamma_t$. With our choice of γ_t in (3), c_t becomes

$$c_t = \begin{cases} c \\ c/t \end{cases}$$

where

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

Recall also the parameter spaces: For decreasing gain, $\beta < 1$ and any $\gamma > 0$ are admissible such that $c > 0$. For constant gain, as opposed to that, $\gamma > 0$ while β and, therefore, c may take any value. Finally, we make the following maintained assumptions:

Maintained assumptions. *The ε_t, for $t = 1, 2, \ldots$, are independently and identically distributed (i.i.d.) with mean 0 and variance σ^2. The initial value a_0 is independent of ε_t, $t = 1, 2 \ldots$ and in L^2.*

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

The outline of the paper is as follows: The asymptotics of a_t are examined in Section 2, both for constant and decreasing gain learning. Subsequently, the strong consistency of the OLS estimators of β and δ is derived in Section 3, again for both learning types. Since the constant gain learning model is essentially an autoregression with intercept, the proofs of the results in Section 3.1 are phrased in neutral notation in a self-contained Section 4. The proofs of the results on the decreasing gain model in Sections 2.2 and 3.2 are relegated to Section 6. A conclusion and an outlook is presented in Section 5.

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. The corresponding model is (1)-(5) with $\gamma_t = \gamma$ and $c_t = c = (1 - \beta)\,\gamma$. As a result, the dynamics of a_t, $t = 1, 2, \ldots$, can be written as

$$a_t = (1 - c)\, a_{t-1} + \gamma\, (\delta + \varepsilon_t)\,. \tag{6}$$

The initial value a_0 satisfies the maintained assumptions.

It is well-known in the literature that constant gain recursions do not in general converge to the REE. In particular, the precise limiting behaviour of a_t as given in (6) is derived in Theorem 1 of CM18 and depends crucially on parameter c. In detail,

(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 (6) is an explosive autoregressive process.

Seminal papers on autoregressive processes that CM18 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 CM18 for the reader's convenience.

Theorem 1 (Christopeit & Massmann (2018, 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 α and positive variance.*
(ii) *If $c = 0$ then a_t is a random walk with drift $\delta\gamma$ and*

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

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

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

(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 $\beta \in (-\infty, \infty)$ and $\gamma > 0$, does a_t converge to the REE α 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. $\gamma_t = \gamma/t$ and $c_t = c/t = (1-\beta)\gamma/t$. Consequently, the recursion of a_t, $t = 1, 2, \ldots$, in (5) becomes

$$a_t = \left(1 - \frac{c}{t}\right) a_{t-1} + \frac{\gamma}{t}\left(\delta + \varepsilon_t\right) \tag{7}$$

where the initial value a_0 satisfies the maintained assumptions.

As mentioned in the introduction, for $\beta < 1$ and $\gamma > 0$, the mere convergence of a_t to α 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 (7) 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 δ and of the disturbance ε_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 2 below that a_t converges almost surely to the REE α for all combinations of β and γ. The *rates* of convergence are, however, different for the three regimes $c > 1/2$, $c = 1/2$ and $0 < c < 1/2$. The proof is relegated to Section 6.1.

Theorem 2. *For decreasing gain with gain sequence $\gamma_t = \gamma/t$, strong convergence of a_t to α holds at the following rates.*

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

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

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

$$\limsup_{t\to\infty} \sqrt{\frac{t}{\ln t \ln_3 t}}\,|a_t - \alpha| = \sigma\gamma\sqrt{2}.$$

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

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

where u has a continuous distribution function.

It is plain that, as c decreases, the convergence of a_t to α gets progressively slower. The value $c = 1/2$ can be interpreted as a boundary separating 'good' 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 CM18. 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 et al. (2013).

It is of interest to compare the convergence rates in Theorem 2 with those valid for weak convergence of $a_t - \alpha$, cf. Theorem 3 in CM18. 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 rates are identical.

3. Strong consistency of the OLS estimator

3.1. Constant gain

In this section we are concerned with the OLS estimation of β and δ in

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t, \tag{8}$$

$t = 1, 2, \ldots$, under constant gain learning. As argued in Section 2.1,

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

does not converge to the REE α for any value of $c = (1 - \beta)\gamma$. There is hence no issue of asymptotic collinearity in (8)-(9).

It is shown in Section 2.2 of CM18 that the OLS estimator $\widehat{\theta}_T = (\widehat{\delta}_T, \widehat{\beta}_T)'$ of $\theta = (\delta, \beta)'$ in (8) is, up to a constant of proportionality, equal to that of $\theta^* = (\delta^*, \beta^*)'$ in the autoregressive model

$$a_t^* = \delta^* + \beta^* a_{t-1}^* + \gamma \varepsilon_t, \tag{10}$$

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,

$$\widehat{\theta}_T - \theta = \gamma^{-1} \left(\widehat{\theta}_T^* - \theta^* \right). \tag{11}$$

By transforming the model in this fashion, we arrive at a first order autoregressive model *with intercept.* 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. In general, however, this literature only provides rates for the *bivariate* (henceforth called *joint*) estimator $\widehat{\theta}_T$, i.e. a rate for its *norm.* 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). Needless to say that this approach works only for autoregressive models with lag one.

In the sequel, we will start with the separate approach in Section 3.1.1. The joint approach will be sketched in Section 3.1.2, followed by a comparison of the two approaches.

3.1.1. Separate estimation of the parameters

Consider the separate OLS estimators of β^* and δ^* in (10), namely

$$\widehat{\beta}_T^* = \frac{\sum_{t=1}^T (a_{t-1} - \overline{a}_T^-)\,(a_t - \overline{a}_T)}{A_T},$$
$$\widehat{\delta}_T^* = \overline{a}_T - \widehat{\beta}_T^* \overline{a}_T^-,$$

where

$$\overline{a}_T = \frac{1}{T} \sum\nolimits_{t=1}^T a_t, \qquad \overline{a}_T^- = \frac{1}{T} \sum\nolimits_{t=1}^T a_{t-1}, \tag{12a}$$

$$A_T^0 = \sum\nolimits_{t=1}^{T} a_{t-1}^2, \qquad A_T = \sum\nolimits_{t=1}^{T} \left(a_{t-1} - \overline{a}_T^-\right)^2 = A_T^0 - T\left(\overline{a}_T^-\right)^2. \tag{12b}$$

Theorem 3 and Corollary 1 below will summarise the properties of the OLS estimators of the original slope β and intercept δ, see (8). The proofs, however, are conducted in terms of the starred model, see Section 4. The main argument of the proofs consists in determining the rate of convergence of the slope estimator.

The case distinctions in Theorem 3 and Corollary 1 are phrased in terms of the parameter c and correspond to the original a_t in (9) 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 (10), as indicated by the parameter β^*:

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

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 4.2.3.

Theorem 3. *Strong consistency of the OLS estimator $\widehat{\beta}_T$ of the slope parameter β holds at the following rates.*

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

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

If only second moments exist, then

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

for all $\eta > 0$.

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

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

If only second moments exist, then

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

for all $\eta > 0$.

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

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

If only second moments exist, then

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

for all $\eta > 0$.

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

$$\frac{\left|1-c\right|^T}{T^{1/2+\eta}}\left(\widehat{\beta}_T - \beta\right) = o(1) \tag{13}$$

for all $\eta > 0$. *If* $\mathbf{E}\left|\varepsilon_t\right|^p < \infty$ *for some* $p > 2$, *(13) 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* δ *holds at the following rates.*

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

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

If only second moments exist, then

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

for all $\eta > 0$.

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

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

If only second moments exist, then

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

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,*

$$T^{1/2-\eta}\left(\widehat{\delta}_T - \delta\right) = o(1) \tag{14}$$

for all $\eta > 0$. *If* $\mathbf{E}\left|\varepsilon_t\right|^p < \infty$ *for some* $p > 2$, *(14) 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.

For the proof of Theorem 3 and Corollary 1 we will make essential use of the observation made at the beginning of this section, namely that the OLS estimator $\widehat{\theta}_T = (\widehat{\delta}_T, \widehat{\beta}_T)'$ of the parameters $\theta = (\delta, \beta)'$ in (8) may equally well be obtained as the OLS estimator of the parameters in a first order autoregressive model with intercept, namely (10). The study of the latter seems, however, to be of some interest of its own – independent of its appearance in our learning model. Section 4 therefore investigates the properties of the paths and of the OLS estimator in a general AR(1)-model with intercept. Theorem 3 and Corollary 1 above are then just Theorem 3* and Corollary 1* in Section 4.4, respectively.

3.1.2. Joint estimation of the parameters

The second approach for estimating $\theta = (\delta, \beta)'$ in (8) employs recent results derived in Nielsen (2005) on the rates of convergence of the *studentised version*

$$\tau_T = M_T^{1/2} \left(\widehat{\theta}_T^* - \theta^* \right) \tag{15}$$

of the OLS estimator of θ^* in (10), 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 χ_T s.t.

$$\chi_T \left\| M_T^{-1/2} \right\| \|\tau_T\| = O\left(1\right) \tag{16}$$

where $\|A\| = \lambda_{\max}^{1/2}\left(A'A\right)$ denotes the spectral norm of A. In view of (11) and (15) the sequence of numbers χ_T will then satisfies

$$\chi_T (\widehat{\theta}_T - \theta) = O(1). \tag{17}$$

Note that (16) involves calculating norms of the *inverse* $M_T^{-1/2}$. This amounts to estimating the *minimal* eigenvalue of M_T since

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

so that (16) turns into

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

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 χ_T s.t.

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

As a further complication, the rates of the two components of $\widehat{\theta}_T$ can (and will in the majority of cases) be different, so that (17) 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 again conducted in terms of the AR(1)-model with intercept in Section 4.

Theorem 4. *Assume that* $\mathbf{E}|\varepsilon_t|^p < \infty$ *for some* $p > 2$. *Then strong consistency of the joint OLS estimator* $\widehat{\theta}_T$ *holds at the following rates.*

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

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

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

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

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

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

for every $\rho > 1/p$.

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 does not seem to be included, however, in (Nielsen, 2005, Theorem 2.5).

3.1.3. 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 (1) under the assumption that $\mathbf{E}|\varepsilon_t|^p < \infty$ for some $p > 2$, the Lai-Wei condition then amounts to

$$\ln \lambda_{\max}(T) = o(\lambda_{\min}(T)), \tag{18}$$

cf. Theorem 1 loc. cit.. Then the joint OLS estimator $\widehat{\theta}_T$ will converge a.s. to θ at rate $(\ln \lambda_{\max}(T)/\lambda_{\min}(T))^{1/2}$.

The following expressions for $\ln \lambda_{\max}/\lambda_{\min}$ for the stable, unit root and explosive cases of the constant gain model follow immediately from the results in Section 4.3.

(i) In the stable case,

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

(18) is thus satisfies. The convergence rate of the OLS estimator is given by $(T/\ln T)^{1/2}$. It is slower than that in Corollary 1 and Theorem 4*(i)*.

(ii) In the unit root case, in view of the Remarks 11 to 13 in Section 4.3.3,

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

Therefore (18) is again satisfied and the corresponding convergence rate is, as in the stable case, $(T/\ln T)^{1/2}$. This the rate appearing in Theorem 4*(ii)*. Given that the error terms have a moment somewhat higher than the second, it is somewhat weaker than the corresponding rate in Corollary 1.

(iii) In the explosive case,

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

Hence (18) is violated. Nevertheless, it is shown in Theorems 3 and 4 that the OLS estimator is strongly consistent. This shows that condition (18) is indeed not necessary. The explosive case of our model may hence be seen as a counterpart to Example 1 in Lai & Wei (1982a).

Remark 1. *Note, however, that the condition*

$$\frac{A_T}{\ln T} \to \infty$$

in Lai & Wei (1982b), valid for simple regression models, is satisfied in the explosive case, in view of the result in Section 4.2.2.

3.2. Decreasing gain

3.2.1. Main result

Consider now OLS estimation of δ and β in

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t \tag{19}$$

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

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

see (7). Recall that the strong consistency of a_t is given by Theorem 2. That of the OLS estimator of β in (19) is presented in the following theorem, whose proof can be found in Section 6.2. As in the context of weak consistency of $\widehat{\beta}_T$ in CM18, 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 5. *For decreasing gain with gain sequence $\gamma_t = \gamma/t$, strong consistency of the OLS estimator $\widehat{\beta}_T$ of the slope parameter β holds at the following rates.*

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

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

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

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

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

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

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

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

Let us compare the convergence rates in Theorem 5 to those established in the context of the weak consistency of $\widehat{\beta}_T$ in CM18. There it was found, cf. Theorem 4 loc. cit., 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 5 above that the 'path taming' sequences are given by $(\ln_2 T)^{-(1+\eta)}$ and $(\ln T)^{-(1+\eta)}$, respectively.

A comparison of Theorems 2 and 5 reveals that there is a trade-off between the convergence rates of a_t and the $\widehat{\beta}_T$. For a further discussion of this issue, see CM18.

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

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

In view of the LIL, any normalising sequence ψ_T should satisfy

$$\psi_T \sqrt{\frac{\ln_2 T}{T}} = O(1). \tag{20}$$

It is apparent that all rates exhibited for the slope in Theorem 5 satisfy (20). Therefore, we have the following result.

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

3.2.2. Comparison with Lai & Wei

As in the constant gain setting, cf. Section 3.1.3, it may be of some interest to check the Lai-Wei condition (18) in the decreasing gain model, too. Since the behaviour of the basic statistics is different from that in the constant gain case, the eigenvalues of the second moment matrix of the regressor $(1, a_{t-1})$ have to calculated anew. We start with the basic formula

$$\lambda_{\pm} = \frac{T + A_T^0}{2} \left[1 \pm \sqrt{1 - 4D_T}\right] \quad \text{with} \quad D_T = \frac{TA_T^0 - \left(T\overline{a}_T^-\right)^2}{\left(T + A_T^0\right)^2}.$$

The square root expansion $\sqrt{1+x} = 1 + \frac{x}{2} + O\left(x^2\right)$ of the smaller eigenvalue is given by

$$\lambda_{\min} = \frac{T + A_T^0}{2} \left[1 - \left(1 - 2D_T + O\left(D_T^2\right)\right)\right] = D_T \left(T + A_T^0\right)\left(1 + O\left(D_T\right)\right). \tag{21}$$

See also Section 4.3.

Case (i): $c > 1/2$. By (82) and (90),

$$A_T^0 = (1 + o(1))\, r \ln T, \quad \left(T\overline{a}_T^-\right)^2 = O\left(T \ln_2 T\right),$$

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

$$TA_T^0 - \left(T\overline{a}_T^-\right)^2 = (1 + o(1))\, rT \ln T, \quad T + A_T^0 = T\left(1 + o(1)\right)$$

so that

$$D_T = r\frac{\ln T}{T}\left(1 + o(1)\right).$$

Hence, since $D_T \to 0$,

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

For $\lambda_{\min}$, the expansion (21) yields

$$\lambda_{\min} = D_T\left(T + A_T^0\right)\left(1 + O\left(D_T\right)\right) = (1 + o(1))\, r \ln T.$$

As a consequence,

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

Thus the Lai-Wei condition (18) is marginally violated in the same way as in (Lai & Wei, 1982a, Example 1). Yet, the OLS estimator is strongly consistent, as shown in Theorem 5.

Remark 2. *Note that by virtue of* (90) *and* (95)

$$\frac{A_T}{\ln T} \to r.$$

Therefore the consistency condition of Lai & Wei (1982b) mentioned above in Remark 1 is not satisfied.

Case (ii): $c < 1/2$. In this case, making use of (83) and (98a), it turns out that

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

for some finite positive random variable κ, so that the Lai-Wei condition is satisfied.

4. OLS estimation in AR(1) models with intercept

As pointed out at the end of Section 3.1.1, Theorem 3 is essentially a statement about the asymptotic properties of the OLS estimator in general AR(1)-models with intercept. Continuing the discussion at the beginning of Section 3.1, our main focus will be on what we call *separate estimation* of the parameters. The *joint approach*, though inferior for most parameter constellations, is treated because it provides a result for the special case of a unit root model without drift, in which case our separate estimation approach is not conclusive. Needless to say that a distinction between separate and joint estimation is sensible, and a gain in estimation accuracy feasible, only for autoregressive models of order one, since in this case tractable separate expressions for the two parameter estimators are available. Starting with these formulae, Section 4.1 exhibits the basic structure of the proofs. It turns out that the main prerequisites are the asymptotic path properties in conjunction with two basic martingale convergence laws. They are presented in Section 4.2. Eigenvalues of the regressor second moment matrix are computed in Section 4.3 before all ingredients are used in Section 4.4 to prove Theorem 3 and Corollary 1.

In order to make clear that the contents of this section are of interest on their own and may be seen as independent of the constant gain model, we use neutral notation. Re-consider to start with the starred model in (10) with its definition of $\beta^* = 1 - c$ as well as $\delta^* = \gamma\delta$, and recall that $\gamma > 0$. We then set

$$\lambda = \beta^* \quad \text{and} \quad \mu = \delta^*$$

Moreover, the index is now $i = 1, \ldots, n$, instead of $t = 1, \ldots, T$. Theorem 3* and Corollary 1* in Section 4.4.1 are then proved in this neutral notation. They can be translated back to the underlying (non-starred) notation of the gain model in (8) by noting the identities $\beta = 1 - (1 - \lambda)/\gamma$ and $\delta = \mu/\gamma$ and by recalling the classification of cases at the beginning of Section 3.1.1. In particular, Theorem 3 and Corollary 1 distinguish between the three cases according to values of c to facilitate comparison with the corresponding results in CM18. Since the material in this section is self-contained, we incorporate the proofs of all auxiliary as well as main results.

4.1. Prerequisites

The basic model for this section is the AR(1) model with intercept

$$y_n = \mu + \lambda y_{n-1} + \varepsilon_n, \tag{22}$$

where the ε_n, $n = 1, 2, \ldots$, are i.i.d. disturbances. Some comments on this assumption will be made later on.

4.1.1. Separate approach

The standard textbook formulae for the OLS estimators of the two parameters λ and μ are

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

where

$$\overline{y}_n = \frac{1}{n}\sum_{k=1}^{n} y_k, \quad \overline{y}_n^- = \frac{1}{n}\sum_{k=1}^{n} y_{k-1.}.$$

Or, in the form to be used below,

$$\widehat{\lambda}_n - \lambda = \frac{u_n}{A_n} - \frac{\overline{y}_n^-}{A_n}\sum_{k=1}^{n} \varepsilon_k, \tag{23a}$$

$$\widehat{\mu}_n - \mu = \left(\lambda - \widehat{\lambda}_n\right)\overline{y}_n^- + \overline{\varepsilon}_n, \tag{23b}$$

where we have put

$$u_n = \sum_{k=1}^{n} y_{k-1}\varepsilon_k \quad \text{and} \quad A_n = \sum_{k=1}^{n} \left(y_{k-1} - \overline{y}_n^-\right)^2.$$

For later use, introduce

$$A_n^0 = \sum_{k=1}^{n} y_{k-1}^2$$

and note the trivial but useful formula

$$A_n = A_n^0 - n\left(\overline{y}_n^-\right)^2.$$

4.1.1.1. Estimation of the slope Our procedure to establish strong convergence rates for the slope will be as follows. Introduce functions

$$\varphi_1(x) = \sqrt{\frac{x}{(\ln x)^{1+\eta}}} \quad \text{and} \quad \varphi_2(x) = \sqrt{\frac{x}{\ln_2 x}}$$

(for $\eta \geq 0$ and for x large enough). Then we may write (23a) in the form

$$\widehat{\lambda}_n - \lambda = \frac{u_n}{A_n^0}\frac{A_n^0}{A_n} - \frac{\overline{y_n^-}}{A_n}\sum\nolimits_{k=1}^{n} \varepsilon_k = \varphi_i^{-1}\left(A_n^0\right) U_n^i \frac{A_n^0}{A_n} - V_n, \tag{24}$$

where we have introduced

$$U_n^i = u_n \frac{\varphi_i\left(A_n^0\right)}{A_n^0}, \quad V_n = \frac{\overline{y_n^-}}{A_n}\sum\nolimits_{k=1}^{n} \varepsilon_k.$$

Remark 3. *For all cases considered, it will turn out that* $A_\infty^0 = \lim_{n\to\infty} = \infty$, *so that for n large enough the expressions* $\varphi_i\left(A_n^0\right)$ *are well defined.*

Remark 4. *The distinction between the two cases* $i = 1$ *or* 2 *is introduced to take account of the strength of assumptions imposed on the* ε_n, *cf. the three scenarios below.*

The U_n^i are of the form

$$U_n^1 = \frac{u_n}{\sqrt{A_n^0\left(\ln A_n^0\right)^{1+\eta}}} \quad \text{and} \quad U_n^2 = \frac{u_n}{\sqrt{A_n^0 \ln_2 A_n^0}}, \tag{25}$$

respectively. The decisive point is that A_n^0 is the *predictable quadratic variation* of u_n. This calls for some sharpened martingale convergence theorem (MCT), ideally of the LIL type. The following well-known MCTs are fundamental to our approach. They hold for $A_\infty^0 = \infty$, cf. Remark 3 above. For the result in (26), see also Chow (1965).

MCT 1 (Lai & Wei (1982a))**.**

$$\sum_{k=1}^{n} y_{k-1}\varepsilon_k = o\left(\sqrt{A_n^0\left(\ln A_n^0\right)^{1+\eta}}\right) \tag{26}$$

for all $\eta > 0$. *If* $\mathbf{E}\left|\varepsilon_n\right|^p < \infty$ *for some* $p > 2$, *this may be sharpened to* $\eta = 0$, *but with* $o(\cdot)$ *replaced by* $O(\cdot)$.

MCT 2 (Wei (1985))**.** *If, in addition to* $\mathbf{E}\left|\varepsilon_n\right|^p < \infty$ *for some* $p > 2$, *it holds that*

$$y_n^2 = o\left[\left(A_n^0\right)^\gamma\right]$$

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

$$\sum_{k=1}^{n} y_{k-1}\varepsilon_k = O\left(\sqrt{A_n^0 \ln_2 A_n^0}\right). \tag{27}$$

Remark 5. *Actually, the MCTs are valid for martingale difference sequences (MDS)* ε_n *with respect to some filtration* $\mathcal{F}_n$ *and some predetermined sequence* y_{n-1}. *The integrability conditions to be introduce below then have to be replaced by corresponding conditions on the conditional moments of the form* $\mathbf{E}\left\{\left|\varepsilon_n\right|^p \mid \mathcal{F}_{n-1}\right\} < \infty$. *We come back to this point in Section 5.*

We will henceforth distinguish between three scenarios. They determine which MCT and hence, in view of (24), which statistic $\varphi_i(A_n^0)$ may be used (at best).

($\mathcal{S}1$) The ε_n are i.i.d. with finite second moments. MCT 1 is valid, use statistic φ_1 with $\eta > 0$.

($\mathcal{S}1+$) The ε_n are i.i.d. and $\mathbf{E}\left|\varepsilon_n\right|^p < \infty$ for some $p > 2$. MCT 1 is valid for $\eta = 0$, use statistic φ_1 with $\eta = 0$.

($\mathcal{S}2$) In addition to ($\mathcal{S}1+$), (2) holds for some for some $0 < \gamma < 1$. MCT 2 is valid, use statistic φ_2.

It will turn out that $A_\infty^0 = \infty$, holds in every scenario.

The basic building block of our analysis will be that

$$U_n^i = O\left(1\right) \tag{28}$$

in each scenario. Independently of the scenario, what we are actually looking for are *deterministic* convergence rates for $\widehat{\lambda}_n - \lambda$, i.e. a sequence of numbers φ_n s.t.

$$\varphi_n(\widehat{\lambda}_n - \lambda) = O(1). \tag{29}$$

In view of (24) and (28), letting φ denote any of the functions φ_i,

$$\varphi_n(\widehat{\lambda}_n - \lambda) = \frac{\varphi_n}{\varphi\left(A_n^0\right)} U_n \frac{A_n^0}{A_n} - \varphi_n V_n. \tag{30}$$

In view of (28), a set of *sufficient conditions* for (29) to hold is

$$\frac{\varphi_n}{\varphi\left(A_n^0\right)} \frac{A_n^0}{A_n} = O(1), \tag{31a}$$

$$\varphi_n V_n = O(1). \tag{31b}$$

To verify (31a), it is often easier to establish the sufficient conditions

$$\frac{\varphi_n}{\varphi\left(A_n^0\right)} = O(1), \tag{32a}$$

$$\frac{A_n^0}{A_n} = O(1). \tag{32b}$$

As to (31b), write

$$\varphi_n V_n = \varphi_n \frac{\overline{y}_n^-}{A_n} \sum\nolimits_{k=1}^n \varepsilon_k = \varphi_n \sqrt{n \ln_2 n} \frac{\overline{y}_n^-}{A_n} \sqrt{\frac{1}{n \ln_2 n}} \sum\nolimits_{k=1}^n \varepsilon_k. \tag{33}$$

By the LIL, a sufficient condition for (31b) to hold is then

$$\varphi_n \sqrt{n \ln_2 n} \frac{\overline{y}_n^-}{A_n} = O\left(1\right). \tag{34}$$

Collecting the conditions established so far, what remains to be done is to consider the asymptotic behaviour of the *basic* statistics $\overline{y}_n, \overline{y}_n^-, A_n^0, A_n$ as well that of the *derived* statistics

$$\varphi_i(A_n^0), \frac{A_n^0}{A_n} \quad \text{and} \quad \frac{\overline{y}_n^-}{A_n}.$$

This will be done in the next Section 4.2. The behaviour will be different depending on whether the stable case, the explosive case or the unit root case is considered.

4.1.1.2. Estimation of the intercept In view of (23b), a set of sufficient conditions for any rate ψ_n satisfying $\psi_n(\widehat{\mu}_n - \mu) = O(1)$ is

$$\psi_n \overline{\varepsilon}_n = O(1), \tag{35a}$$

$$\psi_n(\widehat{\lambda}_n - \lambda)\overline{y}_n^- = O(1). \tag{35b}$$

Writing

$$\psi_n \overline{\varepsilon}_n = \psi_n \sqrt{\frac{\ln_2 n}{n}} \frac{1}{\sqrt{n \ln_2 n}} \sum\nolimits_{k=1}^n \varepsilon_k = \frac{\psi_n}{\varphi_2\left(n\right)} \frac{1}{\sqrt{n \ln_2 n}} \sum\nolimits_{k=1}^n \varepsilon_k$$

shows, cf. the LIL, that

$$\frac{\psi_n}{\varphi_2\left(n\right)} = O(1) \tag{36a}$$

is necessary and sufficient for (35a). (36a) rules out all rates tending faster to infinity than $\psi_n = \varphi_2\left(n\right)$. Also, if φ_n is the rate for $\widehat{\lambda}_n$ according to (29), then (35b) becomes

$$\psi_n(\widehat{\lambda}_n - \lambda)\overline{y}_n^- = \varphi_n(\widehat{\lambda}_n - \lambda)\frac{\psi_n}{\varphi_n}\overline{y}_n^- = O(1).$$

Therefore a sufficient condition for (35b) is

$$\frac{\psi_n}{\varphi_n}\overline{y}_n^- = O(1). \tag{36b}$$

Our procedure will therefore be to find sequences of numbers ψ_n that satisfy (36a) and (36b).

4.1.2. Joint approach

The joint approach works with the usual multivariate (here: bivariate) formulation of (22). Then the textbook formula for the OLS estimator of the two-dimensional parameter vector $\theta = (\mu, \lambda)'$ is given by

$$\widehat{\theta}_n - \theta = M_n^{-1} w_n,$$

where

$$M_n = \begin{pmatrix} n & \sum_{k=1}^n y_{k-1} \\ \sum_{k=1}^n y_{k-1} & \sum_{t=1}^T y_{k-1}^2 \end{pmatrix} \quad \text{and} \quad w_n = \begin{pmatrix} \sum_{k=1}^n \varepsilon_k \\ \sum_{k=1}^n y_{k-1}\varepsilon_k \end{pmatrix}.$$

The usual approach would be to estimate the (Euclidean) *norm* $\|\widehat{\theta}_n - \theta\|$ by

$$\left\|\widehat{\theta}_n - \theta\right\| \leq \left\|M_n^{-1}\right\| \left\|w_n\right\| \tag{37}$$

and then try to obtain rates of convergence for both quantities on the right hand side of (37). Note that this involves the computation of the norm of the inverse M_n^{-1}, which is tantamount to calculating the *minimal eigenvalue* $\lambda_{\min}(M_n)$ since

$$\left\|M_n^{-1}\right\| = \left\|M_n^{-1/2}\right\|^2 = \lambda_{\max}\left(M_n^{-1}\right) = \frac{1}{\lambda_{\min}\left(M_n\right)}. \tag{38}$$

For the convergence rate of $\widehat{\theta}_n$, one is therefore left with the task of finding a sequence of numbers χ_n s.t.

$$\chi_n \frac{\|w_n\|}{\lambda_{\min}\left(M_n\right)} = O\left(1\right). \tag{39}$$

This would make use of martingale convergence theorems.

An alternative approach was recently proposed by Nielsen (2005), who derives rates of convergence for the *studentised version*

$$\tau_n = M_n^{1/2}\left(\widehat{\theta}_n - \theta\right) \tag{40}$$

of the OLS estimator for stable, explosive and unit root vector autoregressive models. Convergence rates for the OLS estimator itself may then be obtained as follows: Given the rates for $\|\tau_n\|$, find sequences of numbers χ_n s.t.

$$\chi_n \left\|M_n^{-1/2}\right\| \|\tau_n\| = O\left(1\right) \tag{41}$$

or, equivalently,

$$\chi_n \frac{\|\tau_n\|}{\sqrt{\lambda_{\min}\left(M_n\right)}} = O\left(1\right), \tag{42}$$

see (38). The computation of the convergence rate χ_n hence hinges on the calculation of $\lambda_{\min}(M_n)$. Since $\widehat{\theta}_n - \theta = M_n^{-1/2}\tau_n$, it then follow from (41) that χ_n satisfies

$$\chi_n \left\|\widehat{\theta}_n - \theta\right\| = O(1).$$

As pointed out in Section 3, the rates for the OLS estimator obtained by the joint approach cannot be better than those for the intercept obtained by the separate approach. Actually, they turn out basically the same, except for the *unit root case* $\lambda = 1, \mu = 0$. In this case, the separate approach does not lead to a result, whereas the joint approach does. Therefore our focus in Section 4.3 will be on this case. Also, as we can build on Nielsen (2005), we will use the second approach based on (42).

4.2. Path behaviour

4.2.1. Stable case

The following path properties follow readily from the well-known ergodic behaviour of the stationary solution to (22) and carry over to any other (causal) solution.

1. *Basic statistics:*

$$\lim_{n\to\infty} \overline{y_n^-} = \lim_{n\to\infty} \overline{y}_n = \frac{\mu}{1-\lambda}, \tag{43a}$$

$$\lim_{n\to\infty} \frac{1}{n} A_n^0 = \tau^2, \quad \text{with} \quad \tau^2 = \frac{\sigma^2}{1-\lambda^2} + \frac{\mu^2}{(1-\lambda)^2}, \tag{43b}$$

$$\lim_{n\to\infty} \frac{1}{n} A_n = \frac{\sigma^2}{1-\lambda^2}. \tag{43c}$$

2. *Derived statistics:*

$$\lim_{n\to\infty} \frac{A_n^0}{A_n} = \frac{\tau^2}{\sigma^2/(1-\lambda^2)} = 1 + \frac{\mu^2}{\sigma^2}\frac{1+\lambda}{1-\lambda} = r, \tag{44a}$$

$$\lim_{n\to\infty} \frac{\varphi_1(n)}{\varphi_1(A_n^0)} = \tau^{-1} \quad \text{for all } \eta \geq 0, \tag{44b}$$

$$\lim_{n\to\infty} \frac{\varphi_2(n)}{\varphi_2(A_n^0)} = \tau^{-1}, \tag{44c}$$

$$\lim_{n\to\infty} n\frac{\overline{y_n^-}}{A_n} = \frac{\mu}{\sigma^2}(1+\lambda). \tag{44d}$$

Remark 6. *For the stable case, condition (2) is satisfied provided that* $\mathbf{E}|\varepsilon_n|^p < \infty$ for some $p > 2$. *This can be seen as follows. By (Lai & Wei, 1985, Theorem 1), any solution y_n^0 of the model in (22) with $\mu = 0$, i.e. of the homogeneous model, satisfies*

$$\left(y_n^0\right)^2 = o(n^{2q}) \text{ for every } q > 1/p. \tag{45}$$

Since the inhomogeneous solution y_n differs from y_n^0 at most by a constant, the statement (45) remains true for y_n. On the other hand, by (43b), $A_n^0 = n\tau^2(1+o(1))$. Hence, for every γ,

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

Letting $q \searrow 1/p$, we find that for all $2/p < \gamma < 1$ finally $2/p < 2q < \gamma < 1$, so that $2q - \gamma < 0$.

4.2.2. Explosive case

The causal solution is

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

with

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

By the theorem of Kolmogorov and Khinchine, see (Shiryaev, 1996, Part IV, §2, Theorem 1), the martingale m_n converges a.s. and in L^2 to some finite limit m:

$$m = \lim_{n\to\infty} m_n = \sum\nolimits_{i=1}^\infty \lambda^{-i}\varepsilon_i,$$

and

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

Remark 7. *m has a continuous distribution, cf. Remark A.1 in CM18. See also (Lai & Wei, 1983b, Corollary 3 and 4) and (Lai & Wei, 1985, Lemma 2).*

The following path properties are then immediate consequences.

1. With probability one and in L^2

$$\lim_{n\to\infty} \lambda^{-n} y_n = y_0 + m + \frac{\mu}{\lambda - 1} = b. \tag{46}$$

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

2. *Basic statistics:*

$$n\,|\lambda|^{-n}\,\overline{y}_n = O\,(1)\,, \quad n\,|\lambda|^{-n}\,\overline{y}_n^- = O\,(1)\,. \tag{47}$$

$$\lim_{n\to\infty} n\lambda^{-2n}\overline{y_n^2} = \frac{\lambda^2}{\lambda^2 - 1}\left[y_0 + m + \frac{\mu}{\lambda - 1}\right]^2 = v^2. \tag{48}$$

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

Proof. (46) is obvious. As to (47), for $\lambda > 1$, the Toeplitz Lemma applied to $\xi_n = \lambda^{-n} y_n$ yields $\lim_{n\to\infty} n\lambda^{-n}\overline{y}_n = \lambda b/\left(\lambda - 1\right)$. For $\lambda < 1$, the Toeplitz Lemma cannot be applied since the λ^n alternate in sign. Writing

$$|\lambda|^{-n}\,y_n = \left(\frac{\lambda}{|\lambda|}\right)^n \lambda^{-n} y_n = (-1)^n\,\lambda^{-n} y_n$$

shows that $|\lambda|^{-n}\,y_n$ does not converge (except for $b = 0$) but is, in any event, $O(1)$. Therefore, with $\xi_k = |\lambda|^{-k}\,y_k$,

$$\frac{|\lambda|}{|\lambda| - 1}\frac{1}{|\lambda|^n - 1}\sum\nolimits_{k=1}^n y_k = \left[\sum\nolimits_{k=1}^n |\lambda|^k\right]^{-1}\sum\nolimits_{k=1}^n |\lambda|^k\,\xi_k = O\,(1)\,.$$

This shows (47). (Since $\overline{y}_n$ and $\overline{y}_n^-$ differ only by $n^{-1}(y_0 - y_n)$, the means behave the same way.) For (48), apply again the Toeplitz Lemma to $\xi_n^2 = \lambda^{-2n} y_n^2$. together with.$A_n = A_n^0 - n\left(\overline{y}_n^-\right)^2$ and $n\left(\overline{y}_n^-\right)^2 = O\left(\lambda^{2n}/n\right)$. □

3. *Derived statistics:*

$$\lim_{n\to\infty} \lambda^{-2n} A_n^0 = v^2, \tag{49a}$$

$$\lim_{n\to\infty} \frac{A_n^0}{A_n} = 1, \tag{49b}$$

$$\lim_{n\to\infty} \frac{\varphi_1\left(\lambda^{2n}\right)}{\varphi_1\left(A_n^0\right)} = \frac{1}{v} \quad \text{for all } \eta \geq 0, \tag{49c}$$

$$n\,|\lambda|^n\,\frac{\overline{y}_n^-}{A_n} = O(1). \tag{49d}$$

Sketch of proof. (49a) is just (48). (49b) follows from $A_n = A_n^0 - n\left(\overline{y}_n^-\right)^2$ and $n\left(\overline{y}_n^-\right)^2 = O\left(\lambda^{2n}/n\right)$. (49c) is a consequence of

$$A_n^0 = v^2\lambda^{2n}\left(1+o(1)\right), \quad \ln A_n^0 = \left(1+o(1)\right)\ln\lambda^{2n},$$
$$\frac{1}{\varphi_1\left(A_n^0\right)^2} = \frac{\left(\ln A_n^0\right)^{1+\eta}}{A_n^0} = \frac{\left(\ln\lambda^{2n}\right)^{1+\eta}}{\lambda^{2n}}\frac{1}{v^2}\left(1+o(1)\right).$$

(49d) follows from (47) together with (49a) and (49b). □

Remark 8. *Unlike in the stable case, (2) does not hold. This is clear since* $y_n^2 \sim \lambda^{2n}, A_n^0 \sim \lambda^{2n}$, *so that*

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

with the exponent on the right hand side being positive for all $0 < \gamma < 1$. *Therefore there is no need to consider the statistic* $\varphi_2\left(A_n^0\right)$.

4.2.3. Unit root case

Case $\lambda = 1, \mu \neq 0$. The solution to (22) in this case is the *random walk with drift*

$$y_n = y_0 + n\mu + \sum\nolimits_{k=1}^n \varepsilon_k. \tag{50}$$

1. *Basic statistics:*

$$\lim_{n\to\infty}\frac{y_n}{n} = \mu, \tag{51a}$$
$$\lim_{n\to\infty}\frac{1}{n}\overline{y}_n = \lim_{n\to\infty}\frac{1}{n}\overline{y}_n^- = \frac{\mu}{2}, \tag{51b}$$
$$\lim_{n\to\infty}\frac{1}{n^2}\overline{y_n^2} = \frac{\mu^2}{3}, \tag{51c}$$
$$\lim_{n\to\infty}\frac{1}{n^3}A_n^0 = \frac{\mu^2}{3}, \quad \lim_{n\to\infty}\frac{1}{n^3}A_n = \frac{\mu^2}{12} \tag{51d}$$

Sketch of proof. The proof is again a direct consequence of (50) and the Toeplitz Lemma. For the last line, note that $n\left(\overline{y}_n^-\right)^2 \sim n^3\mu^2/4$. □

2. *Derived statistics:*

$$\lim_{n\to\infty}\frac{A_n^0}{A_n} = 4, \tag{52a}$$
$$\lim_{n\to\infty}\sqrt{\frac{n^3}{\left(\ln n\right)^{1+\eta}}}\frac{1}{\varphi_1\left(A_n^0\right)} = \frac{3^{1+\eta/2}}{\mu^2} \quad \text{for all } \eta \geq 0, \tag{52b}$$
$$\lim_{n\to\infty}\sqrt{\frac{n^3}{\ln_2 n}}\frac{1}{\varphi_2\left(A_n^0\right)} = \frac{\sqrt{3}}{\mu^2}, \tag{52c}$$

$$\lim_{n\to\infty} n^2 \frac{\overline{y_n^-}}{A_n} = \frac{3}{2\mu}. \tag{52d}$$

Sketch of proof.

$$\begin{aligned} A_n^0 &= n^3 \frac{\mu^2}{3} (1 + o(1)), \\ \ln A_n^0 &= 3 \ln n + O(1) = (1 + o(1)) \, 3 \ln n, \\ \ln_2 A_n^0 &= (1 + o(1)) \ln_2 n, \\ n^2 \frac{\overline{y_n^-}}{A_n} &= \frac{\overline{y_n^-}/n}{A_n/n^3} \to \frac{\mu/2}{\mu^2/3}. \end{aligned}$$

□

Remark 9. *For $\lambda = 1$, $\mu \neq 0$, condition (2) 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 $\lambda = 1, \mu = 0$. In this case, y_n is the random walk $S_n = \sum_{k=1}^n \varepsilon_k$. For the first two moments, we have the following estimates:

$$\limsup_{n\to\infty} \frac{1}{\sqrt{2n \ln_2 n}} |\overline{y}_n| \leq \frac{2}{3}\sigma \tag{53}$$

and

$$\limsup_{n\to\infty} \frac{1}{2n^2 \ln_2 n} A_n^0 \leq \sigma^2, \tag{54a}$$

$$\liminf_{n\to\infty} \frac{\ln_2 n}{2n^2} A_n^0 = \frac{\sigma^2}{8}. \tag{54b}$$

Proof. (53) 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 (54), 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 $1/A_n^0 = O\left(n^{-2} \ln_2 n\right)$ and $n \left|\overline{y_n^-}\right|^2 = O\left(n^2 \ln_2 n\right)$, so that $Q_n = n \left|\overline{y_n^-}\right|^2 / A_n^0 = O((\ln_2 n)^2)$. This makes it impossible to determine the behaviour of $A_n = A_n^0 (1 - Q_n)$. □

Case $\lambda = -1$. y_n^0 is the alternating random walk without drift, i.e. the solution to (22) with $\mu = 0$ and $\lambda = -1$. For arbitrary μ, the corresponding solution y_n (with the same initial value y_0) differs from y_n^0 only by a constant:

$$y_n = \begin{cases} y_n^0 + \mu & \text{for } n \text{ odd}, \\ y_n^0 & \text{for } n \text{ even}. \end{cases} \tag{55}$$

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

1. *Mean:*

$$\lim_{n\to\infty} \frac{1}{n} \sum\nolimits_{k=1}^{n} y_{k-1} = \begin{cases} \frac{\mu}{2} & \text{for} \quad n \text{ odd}, \\ 0 & \text{for} \quad n \text{ even}. \end{cases} \tag{56}$$

Proof. The proof takes up an idea in the proof of Theorem 2 in Appendix A.3 of CM18. Since the initial value does not play any role, we assume that $y_0 = 0$. Then

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

where we have introduced the random walk

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

Then

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

The last equality follows by partial summation, with

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

Then, by the law of large numbers (LLN),

$$\lim_{n\to\infty} \frac{1}{n} \sum\nolimits_{k=1}^{n} y_{k-1}^0 = 0.$$

The assertion then follows from (55). □

2. *2nd moments:*

$$\limsup_{n\to\infty} \frac{1}{2n^2 \ln_2 n} A_n^0 \le \sigma^2, \tag{57a}$$

$$\liminf_{n\to\infty} \frac{\ln_2 n}{2n^2} A_n^0 \ge \frac{\sigma^2}{16}. \tag{57b}$$

The same holds true for A_n.

Proof. By (55) and (56),

$$A_n^0 = \sum\nolimits_{k=1}^{n} \left(y_{k-1}^0\right)^2 + O(n) = \sum\nolimits_{k=1}^{n} \left(\widetilde{S}_{k-1}\right)^2 + O(n).$$

The right hand side of (57) is the same as in (54), where, however, y_n was a random walk: $y_n = S_n$. If S_n is replaced by $\widetilde{S}_n$, (57a) remains valid by the LIL for sums of weighted i.i.d. sequences by Chow & Teicher (1973). As to (57b), the above mentioned theorem of (Donsker & Varadhan, 1977, page 751) assumes *i.i.d.* shocks, in which case (54b) is true. At any rate, for *symmetric* ε_n, $\widetilde{S}_n$ is again a random walk of i.i.d. shocks so that (54b) holds for such error terms. It can, however, be shown that it remains valid at least in the weaker form (57b) also for non-symmetric ε_n. To see this, introduce random variables $\varepsilon_k^* = \varepsilon_{2k} - \varepsilon_{2k-1}$.Then the ε_n^* are i.i.d. with variance $2\sigma^2$ and

$$\widetilde{S}_{2n} = \sum\nolimits_{k=1}^{n} \varepsilon_k^* = S_n^*,$$

a random walk of the ε_k^*. Let $[n/2]$ denote the largest integer $\leq n/2$. Then $A_n^0 \geq \sum_{k=1}^{[n/2]-1} \widetilde{S}_{2k}^2 + O(n) = \sum_{k=1}^{[n/2]-1} (S_k^*)^2 + O(n) = A_{[n/2]}^* + O(n)$. But

$$\liminf_{n\to\infty} \frac{\ln_2 [n/2]}{2\,[n/2]^2} A_{[n/2]}^* = \frac{2\sigma^2}{8},$$

so that

$$\liminf_{n\to\infty} \frac{\ln_2 n}{2n^2} A_{[n/2]}^* \geq \frac{\sigma^2}{16}.$$

(57) carries over to A_n. For the first inequality, this follows trivially from $A_n \leq A_n^0$. For (57b), it is a consequence of (56), which implies that $\left|\overline{y}_n^-\right| = O(1)$ and therefore

$$\frac{\ln_2 n}{2n^2} A_n = \frac{\ln_2 n}{2n^2} A_n^0 - \frac{\ln_2 n}{2n} \left|\overline{y}_n^-\right|^2 = \frac{\ln_2 n}{2n^2} A_n^0 + o(1). \tag{58}$$

□

3. *Derived statistics:*

$$\frac{A_n^0}{A_n} = O\left[(\ln_2 n)^2\right], \tag{59a}$$

$$\frac{\varphi_{1n}}{\varphi_1(A_n^0)} = O(1) \text{ with } \varphi_{1n} = \frac{n}{\sqrt{(\ln n)^{1+\eta} \ln_2 n}}, \text{ for all } \eta \geq 0, \tag{59b}$$

$$\frac{\varphi_{2n}}{\varphi_2(A_n^0)} = O(1) \text{ with } \varphi_{2n} = \frac{n}{\ln_2 n}, \tag{59c}$$

$$\frac{\left|\overline{y}_n^-\right|}{A_n} = O\left(\frac{\ln_2 n}{n^2}\right). \tag{59d}$$

Proof. Ad (59a).

$$\frac{1}{(\ln_2 n)^2} \frac{A_n^0}{A_n} = \frac{\frac{1}{2n^2 \ln_2 n} A_n^0}{\frac{\ln_2 n}{2n^2} A_n} = \frac{P_n}{Q_n},$$

$$\limsup_{n\to\infty} \frac{1}{(\ln_2 n)^2} \frac{A_n^0}{A_n} \leq \frac{\limsup_{n\to\infty} P_n}{\liminf_{n\to\infty} Q_n} \leq 2.$$

Ad (59b). Denote $\alpha_n = 2n^2 \ln_2 n$. Then

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

or

$$\frac{\ln A_n^0}{2 \ln n} = (1 + o(1)) + \frac{\ln\left(\alpha_n^{-1} A_n^0\right)}{2 \ln n} \tag{60}$$

By (57a), $\limsup_{n\to\infty} \ln\left(\alpha_n^{-1} A_n^0\right) \leq \ln \sigma^2$, so that

$$\lim_{n\to\infty} \frac{\ln A_n^0}{2 \ln n} = 1. \tag{61}$$

On the other hand, making use of (61) and (57b), we may write

$$\frac{1}{\varphi_1\left(A_n^0\right)^2}=\frac{\left(\ln A_n^0\right)^{1+\eta}}{A_n^0}=\frac{2n^2}{\ln_2 n}\frac{1}{A_n^0}\left(\frac{\ln A_n^0}{2\ln n}\right)^{1+\eta}\frac{\left(2\ln n\right)^{1+\eta}\ln_2 n}{2n^2}.$$

Since $\left(\ln_2 n/n^2\right)\left(1/A_n^0\right)=O\left(1\right)$ by (57b) this shows (59b) with

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

Ad (59c). By (61), denoting the $O\left(1\right)$-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, $\liminf_{n\to\infty}\ln C_n\geq-\left(1+\ln 2\right)$. As a consequence,

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

Making use of (61) and (57b),

$$\begin{aligned}\frac{1}{\varphi_2\left(A_n^0\right)^2}=\frac{\ln_2 A_n^0}{A_n^0}&=\frac{2n^2}{\ln_2 n}\frac{1}{A_n^0}\left(\frac{\ln_2 A_n^0}{2\ln_2 n}\right)\frac{\left(\ln_2 n\right)^2}{n^2}\\&=\frac{\left(\ln_2 n\right)^2}{n^2}O\left(1\right).\end{aligned}$$

This shows (59c) with

$$\varphi_{2n}=\frac{n}{\ln_2 n}.$$

Ad (59d). This is a straightforward consequence of (56) and (57b) together with (58). □

Remark 10. *If* $\mathbf{E}\left|\varepsilon_n\right|^p<\infty$ *for some* $p>2$, *it follows from (57b) that*

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

On the other hand, by the LIL (cf. e.g. MCT 2), $y_n^2=O\left(n\ln_2 n\right)$. *Therefore*

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

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

4.3. Eigenvalues of the moment matrix

For the *joint approach* to the OLS estimator, consider the second moment matrix of the regressor $(1, y_{n-1})$ in (22):

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

Its eigenvalues are given by

$$\lambda_{\max} = \frac{n + A_n^0}{2}\left[1 + \sqrt{1 - 4D_n}\right], \tag{62a}$$

$$\lambda_{\min} = \frac{n + A_n^0}{2}\left[1 - \sqrt{1 - 4D_n}\right], \tag{62b}$$

with

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

Note that both eigenvalues are real, so that $0 \le D_n \le 1/4$. These formulae will be evaluated for the single cases by making use of the path properties established above. For the minimal eigenvalue, which actually is of interest to us in the joint approach (cf. (39) and (42)), it turns out that in the explosive and the unit root case $D_n = o(1)$, so that (62b) is not conclusive. We therefore use the square root expansion

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

to obtain

$$\lambda_{\min} = \frac{n + A_n^0}{2}\left[1 - \left(1 - 2D_n + O\left(D_n^2\right)\right)\right] = D_n\left(n + A_n^0\right)\left(1 + O\left(D_n\right)\right). \tag{63}$$

In the following, we will report the eigenvalues for the different cases. Except for the unit root case $\lambda = 1, \mu = 0$ the proofs are on the basis of (62) or (63), using the path properties. They are rather straightforward and/or the results can be found elsewhere, see e.g. Nielsen (2005). We therefore desist from reproducing them here.

4.3.1. Stable case

$$\lim_{n\to\infty} \frac{1}{n}\lambda_{\max} = \lambda_+ = \frac{1+\tau^2}{2}\left[1 + \sqrt{1-4D}\right],$$

$$\lim_{n\to\infty} \frac{1}{n}\lambda_{\min} = \lambda_- = \frac{1+\tau^2}{2}\left[1 - \sqrt{1-4D}\right]$$

with $D = \lim_{n\to\infty D_n} = \sigma^2/\left(1-\lambda^2\right)\left(1+\tau^2\right)^2$ and τ^2 as in (43b).

4.3.2. Explosive case

$$\lim_{n\to\infty} \lambda^{-2n}\lambda_{\max} = v^2,$$
$$\lim_{n\to\infty} \frac{1}{n}\lambda_{\min} = 1,$$

with v^2 as in (48).

4.3.3. Unit root case

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

$$\lim_{n\to\infty} \frac{1}{n^3}\lambda_{\max} = \frac{\mu^2}{3},$$
$$\lim_{n\to\infty} \frac{1}{n}\lambda_{\min} = \frac{1}{2}.$$

Remark 11. *For future reference note that*

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} = 6\frac{\ln n}{n}\left(1 + o(1)\right) = o(1).$$

Case $\lambda = -1$.

$$\limsup_{n\to\infty} \frac{1}{2n^2 \ln_2 n}\lambda_{\max} \leq \sigma^2,$$
$$\liminf_{n\to\infty} \frac{\ln_2 n}{2n^2}\lambda_{\max} \geq \frac{\sigma^2}{2},$$
$$\lim_{n\to\infty} \frac{1}{n}\lambda_{\min} = 1.$$

Remark 12. *For future reference, note that*

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

Hence, since $A_n^0/n \to \infty$*, making use of (61),* $\ln \lambda_{\max} = \ln A_n^0 + o\left(1\right) = \left(1 + o\left(1\right)\right) 2\ln n$*, so that*

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} = \frac{2\ln n}{n}\left(1 + o(1)\right) = o\left(1\right).$$

We have singled out the case $\lambda = 1, \mu = 0$, because of its importance due to two reasons. First, our separate estimation approach does not lead to a result in this case. Secondly, even in the joint approach it appears to be a white spot in the literature, to the best of our knowledge. For instance, it is not covered by (Nielsen, 2005, Theorem 2.5).

Case $\lambda = 1, \mu = 0$. In this case, the formulae (62) turn out not to be particularly useful since the behaviour of D_n cannot be derived from the path properties in Section 4.2.3. We therefore pass to the equivalent formulae

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

where we have put $p_n = n\overline{y}_n^-$. Since $n/A_n^0 = O(\ln_2 n/n)$ and $p_n/A_n^0 = O(((\ln_2 n)^3/n)^{1/2})$ by virtue of (54b) and (53), we may write

$$\begin{aligned}\lambda_{\pm} &= \frac{1}{2}\left[A_n^0 + n \pm A_n^0\sqrt{\left(1 - \frac{n}{A_n^0}\right)^2 + 4\left(\frac{p_n}{A_n^0}\right)^2}\right] \\ &= \frac{1}{2}\left[A_n^0 + n \pm A_n^0\left(1 + o(1)\right)\right].\end{aligned} \tag{64}$$

For $\lambda_{\max} = \lambda_+$, this means that

$$\lambda_{\max} = \frac{A_n^0}{2}\left[1 + \frac{n}{A_n^0} + (1 + o(1))\right] = A_n^0\left(1 + o(1)\right).$$

By virtue of (54a), it follows that

$$\lambda_{\max} = O\left(n^2 \ln_2 n\right). \tag{65}$$

For $\lambda_{\min} = \lambda_-$, we write

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

This shows that a more detailed analysis of the $o(1)$-term is necessary in order to capture the asymptotic behaviour of $\left(A_n^0/n\right)o(1)$. For the purpose of establishing our result it suffices, however, to appeal to standard results from the general theory of autoregressive processes *without* intercept. (Lai & Wei, 1985, Theorem 3 for $p = 1$), for instance, show that

$$\liminf_{n\to\infty} \frac{1}{n}\lambda_{\min} > 0, \tag{66}$$

cf. also Lai & Wei (1983a).

Actually, a close look at (64) reveals that $A_n^0 n^{-1} o(1) = O(1)$ so that $\liminf_{n\to\infty} \frac{1}{n}\lambda_{\min} < \infty$. Consequently, the rate in (66) cannot in fact be improved upon.

Remark 13. *(65) and (66) together show that*

$$\frac{\ln \lambda_{\max}}{\lambda_{\min}} = \frac{(2\ln n + \ln_3 n)}{nc_n}O(1) = \frac{\ln n}{n}O(1) = o(1)$$

since $\lambda_{\min} = nc_n$ *with* $\liminf_{n\to\infty} c_n > 0$ *and hence* $c_n^{-1} = O(1)$.

4.4. Consistency of the OLS estimator

We are now ready to go back the OLS estimator discussed in Section 4.1. As before, we will distinguish the separate and the joint approach.

4.4.1. Separate approach

All we need to do is to verify the conditions established in Section 4.1.1 for the individual cases making use of the results in Section 4.2. For each case, the kind of scenario assumed (i.e. the conditions imposed on the ε_n) will determine which statistic φ may be used at best, and the corresponding rates φ_n are obtained form the path properties. Our main concern is the slope.

Theorem 3*. *Strong consistency of the OLS estimator $\widehat{\lambda}_n$ of the slope parameter λ holds at the following rates:*

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

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

If only second moments exist, then

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

for all $\eta > 0$.

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

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

If only second moments exist, then

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

for all $\eta > 0$.

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

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

If only second moments exist, then

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

for all $\eta > 0$.

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

$$\frac{|\lambda|^n}{n^{1/2+\eta}}(\widehat{\lambda}_n - \lambda) = o(1) \tag{67}$$

for all $\eta > 0$. *If* $\mathbf{E}|\varepsilon_n|^p < \infty$ *for some* $p > 2$, *(67) remains valid, with* $O(1)$ *instead of* $o(1)$ *for* $\eta = 0$.

Proof. Ad (i). If $\mathbf{E}|\varepsilon_n|^p < \infty$ for some $p > 2$, (2) is satisfied, cf. Remark 6, so that we have automatically scenario ($\mathcal{S}2$) Making use of (44) we see that $\varphi_n = \varphi_2(n)$ will satisfy both (32) and (34).

If one assumes only finite 2nd moments, one has to make use of MCT 1 and $\varphi_n = \varphi_1(n)$ with $\eta > 0$ will do.

Ad (iia). By Remark 10, (2) is satisfied if higher moments exist. By (52),

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

will satisfy (32) and (34).

If only second moments are assumed, then we have to use MCT 1 and

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

Ad (iib). By Remark 10, (2) is satisfied if higher moments exist.

$$\varphi_n = \frac{n}{(\ln_2 n)^3}$$

will satisfy (31). If only second moments are assumed,

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

will do.

Ad (iii). According to Remark 8, we are at best in scenario ($\mathcal{S}1+$). Then in view of (49) and noting that $\varphi_1\left(\lambda^{2n}\right) = (2\ln|\lambda|)^{-(1+\eta)/2}\left[|\lambda|^n n^{-(1+\eta)/2}\right]$, (32) is satisfied for

$$\varphi_n' = \frac{|\lambda|^n}{n^{(1+\eta)/2}}.$$

Making use of (49d), a simple calculation shows that it also satisfies condition (34). Hence φ_n' is a valid rate for the OLS estimator. But since φ_n' Since $\varphi_n'(\widehat{\lambda}_n - \lambda) = o(1)$ is then true for *all* $\eta > 0$, we may as well take $\varphi_n = |\lambda|^n n^{1+\eta/2}$. □

Coming to the intercept, we have the following

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

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

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

If only second moments exist, then

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

for all $\eta > 0$.

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

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

If only second moments exist, then

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

for all $\eta > 0$.

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

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

$$n^{1/2-\eta}\left(\widehat{\mu}_n - \mu\right) = o(1) \tag{68}$$

for all $\eta > 0$. *If* $\mathbf{E}\left|\varepsilon_n\right|^p < \infty$ *for some* $p > 2$, *(68) remains valid for* $\eta = 0$ *and with* $o(1)$ *replaced by* $O(1)$.

Proof. Ad (i). Since $\overline{y}_n^- = O(1)$, both $\psi_n = \varphi_1(n)$ and $\psi_n = \varphi_2(n)$ from Theorem 3*(i)* will do according to the dichotomy established there.
Ad (iii). By Theorem 3*(iii)*, $\varphi_n = \left|\lambda\right|^n / n^{1/2+\eta}$. Since $n\left|\lambda\right|^{-n}\overline{y}_n^- = O(1)$ (cf. (47)),

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

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

$$\frac{\psi_n}{\varphi_n}\overline{y}_n^- = n^{1/2-\eta}\frac{n^{1/2+\eta}}{\left|\lambda\right|^n}\overline{y}_n^- = \frac{n}{\left|\lambda\right|^n}\overline{y}_n^- = O(1),$$

which shows (36b). (36a) is trivially satisfied.
Ad (iia). Since $\overline{y}_n^-/n \to \mu/2$, cf. (51b), $\psi_n = \varphi_2(n)$ will do if higher moments exist:

$$\frac{\psi_n}{\varphi_n}\overline{y}_n^- = \sqrt{\frac{n}{\ln_2 n}}\sqrt{\frac{\ln_2 n}{n^3}}\frac{n}{2}\mu\left(1 + o(1)\right) = O(1).$$

If only 2nd moments exist,

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

Ad (iib). Same as for *(iia).* □

4.4.2. Joint approach

As pointed out in Section 4.1.2 we follow the Nielsen approach based on (42). Our starting point will be (Nielsen, 2005, Theorem 2.4), which we cite here because it is of interest in its own right.

Result 1 (Nielsen (2005)). *Assume that $\mathbf{E}\left|\varepsilon_t\right|^p < \infty$ for some $p > 2$ and recall the definition of τ_n in (40). Then the following holds with probability one:*

$$\tau_n = \begin{cases} O\left[\left(\ln_2 n\right)^{1/2}\right] & \text{for } |\lambda| < 1, \\ O\left[\left(\ln n\right)^{1/2}\right] & \text{for } |\lambda| = 1, \\ o\left[n^{\rho}\right] & \text{for } |\lambda| > 1, \end{cases} \tag{69}$$

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

As elaborated in Section 4.1.2, this approach comes down to investigating the asymptotic behaviour of the minimal eigenvalues $\lambda_n = \lambda_{\min}\left(M_n\right)$ of the moment matrix M_n and to find sequences of numbers χ_n s.t.

$$\chi_n \frac{\left\|\tau_n\right\|}{\sqrt{\lambda_n}} = O\left(1\right). \tag{70}$$

Then it will hold that

$$\chi_n \left\|\widehat{\theta}_n - \theta\right\| = O(1).$$

The minimal eigenvalues λ_n of M_n are calculated in Section 4.3. The proofs are rather straightforward combinations of those results with (69). It will be given only for the critical unit root case $\lambda = 1, \mu = 0$ since the other cases are not surprising in view of Corollary 1* and the discussion in Section 3.1. Note that due to the assumption in (69) we are automatically in scenario $(\mathcal{S}1+)$, in the notation introduced in Section 4.1.1.

The following theorem summarises the rates for the individual cases.

Theorem 4*. *Assume that $\mathbf{E}\left|\varepsilon_n\right|^p < \infty$ for some $p > 2$ Then strong consistency of the joint OLS estimator $\widehat{\theta}_n$ holds at the following rates.*

(i) Stable case: $|\lambda| < 1$.

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

(ii) Unit root case: $\lambda = 1$ or $\lambda = -1$, with μ arbitrary.

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

(iii) Explosive case: $|\lambda| > 1$.

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

for every $\rho > 1/p$.

Remark 14. *In the stable case, both eigenvalues diverge at the same rate, so that both components of $\widehat{\theta}_n$ will have the same rate of convergence.*

Unfortunately in the critical case $\lambda = 1, \mu = 0$, Theorem 4**(ii)* does not say much about the actual rate of convergence of the slope OLS estimator. Actually, looking at the corresponding rates for the unit root case obtained by separate estimation (cf. Theorem 3*) one should expect a much better rate.

Proof. Ad (ii). For both $\lambda = 1$ and $\lambda = -1$,

$$\liminf_{n\to\infty} \frac{1}{n}\lambda_{\min} > 0,$$

so that

$$\lambda_{\min}^{-1/2} = n^{-1/2}O(1).$$

Using (69),

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

it follows that

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

will satisfy (70). □

5. Conclusion and outlook

5.1. Summary

This paper considers the question of strongly consistent OLS estimation in regression models with adaptive learning. In particular, it makes three contributions to the literature: First, we derive rates at which a_t converges almost surely to the REE α in the decreasing gain learning model. Secondly, we establish rates for the strong consistency of the OLS estimators of δ and β in the constant and decreasing gain learning models. Interestingly, we find that the near optimal sufficient condition by Lai & Wei (1982a) is not satisfied in some of our models. Thirdly, we present a complete treatment of OLS estimation in an autoregressive model of order one with intercept. In particular, we cover the unit root case with slope one and zero intercept, which to our knowledge has not yet been treated in the literature.

5.2. Refinements

If more powerful convergence result than MCT 1 or MCT 2 are available, the results may be refined in several directions. We consider here one exemplary case, namely the stable constant gain case, in the general notation of Section 4. Other scenarios are beyond the scope of the present paper and are left to future research.

In the stable constant gain case, it can be shown that the rate $\varphi_n = \varphi_2(n)$ remains valid even when the ε_n possess *only 2nd moments.* In addition, the vague $O(1)$ result in Theorem 3* may actually be sharpened to yield bounds for the scaled OLS estimator. The basis of the argument is the following 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 $\mathbf{E}\{Y_i \mid Y_1, Y_2, \ldots, Y_{i-1}\} = 0$ a.s. for all $i \geq 2$ and $\mathbf{E}Y_1^2 = \zeta^2$. Then, with probability one,*

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n Y_i}{\sqrt{n \ln_2 n}} = \zeta\sqrt{2}. \tag{71}$$

We apply Result 2 to $Y_i = y_{i-1}\varepsilon_i$. Actually, this sequence is not stationary ergodic unless y_i is the *stationary* solution y_i^0 to (22). Since, however, the difference between any two solutions is $y_n - y_n^0 = \lambda^n \left(y_0 - y_0^0\right)$, the corresponding numerators in (71) differ by $O(1)$, so that (71) remains valid for *any* solution. As y_{n-1} and ε_n are independent, only $\varepsilon_n \in L^2$ needs to be required, cf. (Shiryaev, 1996, Chapter II, §6, Theorem 6), and $\zeta^2 = \sigma^4/\left(1-\lambda^2\right)$. Passing to $(-Y_i)$, we get a similar result for the lim inf, i.e. with $-\zeta\sqrt{2}$ on the right hand side of (71). Return to the basic formula (30) in Section 4.1.1.1:

$$\varphi_n(\widehat{\lambda}_n - \lambda) = \frac{\varphi_n}{\varphi\left(A_n^0\right)} \frac{A_n^0}{A_n} U_n - \varphi_n V_n, \tag{72}$$

where U_n in (25) can be expressed as

$$U_n = \frac{\sum_{k=1}^n y_{k-1}\varepsilon_k}{\sqrt{A_n^0 \ln_2 A_n^0}} = \frac{\sum_{i=1}^n Y_i}{\sqrt{n \ln_2 n}} \frac{\sqrt{n \ln_2 n}}{\sqrt{A_n^0 \ln_2 A_n^0}}.$$

Using Result 2 and (43b) as well as (44c), we now have

$$\limsup_{n\to\infty} U_n = \frac{\zeta\sqrt{2}}{\tau}, \qquad \liminf_{n\to\infty} U_n = -\frac{\zeta\sqrt{2}}{\tau}.$$

This gives a more precise meaning to (28). Writing (33) in the form

$$\varphi_n V_n = \varphi_n \sqrt{n \ln_2 n} \frac{\overline{y_n^-}}{A_n} L_n,$$

with L_n obeying the LIL for i.i.d. ε_n, and employing (43a) and (43c), we can now compute upper and lower bounds for $\varphi_n(\widehat{\lambda}_n - \lambda)$ in (72):

$$\limsup_{n\to\infty} \varphi_n(\widehat{\lambda}_n - \lambda) \leq \sqrt{2}\kappa, \quad \liminf_{n\to\infty} \varphi_n(\widehat{\lambda}_n - \lambda) \geq -\sqrt{2}\kappa$$

with a constant $\kappa = \sqrt{1-\lambda^2} + \mu(1+\lambda)/\sigma$.

5.3. Extensions

As already pointed out in Remark 5, the MCTs are apt to deal with more general error sequence than just i.i.d. ε_n. In dealing with more general error sequences, the chief problem is to determine the asymptotics of the first and second order empirical moments (i.e. what we call the basic statistics) and to mimic their behaviour by some deterministic sequence φ_n. The moment condition on the ε_n has to be replaced by the corresponding condition $\sup_n \mathbf{E}\left\{|\varepsilon_n|^p \mid \mathcal{F}_{n-1}\right\} < \infty$ (with $p \geq 2$) on the conditional moments. A major issue is that in the MCTs the denominator is $g\left(\langle u\rangle_n\right)^{1/2}$, where $\langle u\rangle_n$ is the predictable quadratic variation

$$\langle u\rangle_n = \sum\nolimits_{i=1}^{n} y_{i-1}^2 \mathbf{E}\left\{\varepsilon_i^2 \mid \mathcal{F}_{i-1}\right\}$$

of $u_n = \sum_{i=1}^n y_{i-1}\varepsilon_i$ and where $g(x)$ is either of the functions $g(x) = x\left(\ln x\right)^{1+\eta}$ or $g(x) = x \ln_2 x$. For i.i.d. errors, we have $\langle u\rangle_n = \sigma^2 A_n^0$, with $A_n^0 = \sum_{i=1}^n y_{i-1}^2$, in accordance with the OLS formula where A_n^0 appears as denominator. For general MDSs, $\langle u\rangle_n$ and $\sigma^2 A_n^0$ will generally not coincide, and the crucial task would be to determine the asymptotics of the ratios $\langle u\rangle_n / A_n^0$ and $g\left(\langle u\rangle_n\right)/g\left(A_n^0\right)$. The following two paragraphs offer some idea of the problems that may arise in the process.

For constant gain learning, the approach in Section 4 seems to go through in the *stable case* for *stationary ergodic* sequences ε_n since both properties are inherited by $y_{n-1}\varepsilon_n$ (at least for stationary initial value y_0 independent of the future errors) and ergodicity provides the LLNs for the first two moments. For *independent* but not identically distributed ε_n, anything may happen, including inconsistency of the OLS estimator. In the *explosive case,* the results seem to carry over to MDS error sequences since the basic building block is the fundamental martingale convergence theorem for martingales with finite variation. As to the *unit root case*, the tools needed for the proofs above are very special results for the random walk of i.i.d. sequences, and extensions to other error sequences will be only available for very specific cases.

For decreasing gain learning, the results of Theorem 2 remain basically valid for MDS ε_n with uniformly bounded second moments. In the case $c \geq 1/2$, the $\limsup$ remains finite, but indefinite. The reason is that, instead of to the powerful LIL by Chow & Teicher (1973), appeal has to be made to MCT 2. For $c < 1/2$, *(iii)* remains true without the additional assertion about the distribution of the limit u. For the proof of Theorem 5, when $c > 1/2$, property 3 of Section 6.2.2, which is actually the one determining the asymptotic behaviour of A_T^0, has to be revisited. The point is that it has to be ensured that

$$\sum\nolimits_{t=1}^{T} \varepsilon_t^2 / t = \mathrm{const} \times \ln T + O(1). \tag{73}$$

So whatever error sequence is considered it should satisfy (73). Otherwise the asymptotic behaviour of A_T^0 might be quite different. The case $c < 1/2$ in Theorem 5 does not seem to be affected.

6. Proofs

6.1. Proof of Theorem 2

The proofs proceed along lines similar to those followed in CM18, 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.. In the present paper, we will use a decomposition applied in CM18 in the case $c < 1$, but which actually remains valid for all $c > 0$. As to the probabilistic tools needed, roughly speaking, whenever a CLT comes into the play in CM18, we will now make use of an appropriate strong LLN and a LIL.

Reconsider the recursion (7) for a_t. 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

$$a_t^{\#} = \left(1 - \frac{c}{t}\right) a_{t-1}^{\#} + \frac{\gamma}{t}\varepsilon_t \tag{74}$$

and the DGP in (1) takes the form

$$y_t = \alpha + \beta a_{t-1}^{\#} + \varepsilon_t. \tag{75}$$

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

The basis of all calculations will be the representation

$$a_t = O\left(t^{-c}\right) + \gamma\left(\xi_t + \eta_t\right) \tag{76}$$

of a_t. In (76),

$$\xi_t = \frac{1}{t^c} v_t, \qquad \eta_t = \frac{1}{t^{1+c}} w_t,$$
$$v_t = \sum_{i=1}^{t} \theta_i \frac{\varepsilon_i}{i^{1-c}}, \quad w_t = \sum_{i=1}^{t} \frac{O_{ti}(1)}{i^{1-c}} \varepsilon_i.$$

Here i_0 is the largest[1] integer less than or equal to c. The θ_i are nonnegative deterministic coefficients satisfying $\lim_{t\to\infty} \theta_i = 1$. The $O_{ti}(1)$-terms are deterministic and uniformly bounded in i, t. This representation is proved in Appendix B.1 of CM18 for the special case $c < 1$ (corresponding to $i_0 = 0$), but an inspection of the proof in CM18 shows that it remains valid for *all* $c > 0$.

For $c < 1$ (i.e. $i_0 = 0$), the $O\left(t^{-c}\right)$-term is of the form $O\left(t^{-c}\right) = a_0 B_0 t^{-c} + O\left(t^{-1}\right)$, where B_0 is some positive constant, cf. Appendix B.1 in CM18. Therefore (76) may be put into the stronger form

$$a_t = a_0 B_0 t^{-c} + \gamma\left(\xi_t + \eta_t\right) + O\left(t^{-1}\right). \tag{77}$$

This is the representation proved in CM18 for $c < 1/2$ and which will be needed below for this case.

[1]In CM18, i_0 was erroneously introduced as the *smallest* integer greater than or equal to c, but in the proof the correct definition given here is used.

Case (i): $c > 1/2$. By Lemma 1 below, the predictable quadratic variation $\langle v \rangle_t$ of v_t is the same as that of

$$v'_t = \sum_{i=1}^{t} \frac{\varepsilon_i}{i^{1-c}}.$$

Hence

$$\langle v \rangle_t = \langle v' \rangle_t = \sigma^2 \sum\nolimits_{i=1}^{t} i^{2(c-1)} = \frac{\sigma^2}{2c-1} t^{2c-1} + O(1)$$

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

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

As a consequence,

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

so that

$$\limsup_{t\to\infty} \sqrt{\frac{t}{\ln_2 t}}\, |\xi_t| = \sigma \sqrt{\frac{2}{2c-1}} \tag{78}$$

and hence $\xi_t \to 0$.

Turning to w_t, it follows from the integral comparison test (ICT), see (Apostol, 1974, Proposition 8.23), that

$$\mathbf{E} w_t^2 = O\left(t^{2c-1}\right) \text{ and } \mathbf{E} t\eta_t^2 = \frac{1}{t^2}.$$

Hence, by monotone convergence, $\mathbf{E} \sum_{t=1}^{\infty} t\eta_t^2 < \infty$, so that $t\eta_t^2 \to 0$. In particular, this means that

$$\sqrt{t}\, |\eta_t| = o(1). \tag{79}$$

(78) and (79) show that

$$\limsup_{t\to\infty} \sqrt{\frac{t}{\ln_2 t}}\, |\xi_t + \eta_t| = \sigma \sqrt{\frac{2}{2c-1}}.$$

In connection with (76) this shows *(i)* of Theorem 2 (remember our transformation).

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 $\theta_i \to 1$ and $R_\infty = \infty$. Then $S_t/R_t \to 1$.

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

Case (ii): $c = 1/2$. We go back to the decomposition (77). Again by Lemma 1, the predictable quadratic variation of v_t is given by

$$\langle v \rangle_t = \langle v' \rangle_t = \sigma^2 \sum\nolimits_{i=1}^{t} i^{-1} = \sigma^2 \ln t + O(1).$$

Hence, by the LIL, cited above,

$$\limsup_{t\to\infty} \frac{|v_t|}{\sqrt{\ln t \ln_3 t}} = \sigma\sqrt{2}$$

and therefore

$$\limsup_{t\to\infty} \sqrt{\frac{t}{\ln t \ln_3 t}}\, |\xi_t| = \sigma\sqrt{2}. \tag{80}$$

As for w_t, $\mathbf{E}w_t^2 = O(\ln t)$. Therefore, $\mathbf{E}\sum_{t=1}^{\infty} t\eta_t^2 = O(1)\sum_{t=1}^{\infty} \frac{\ln t}{t^2} < \infty$, so that

$$\sqrt{t}\eta_t = o(1). \tag{81}$$

Theorem 2*(ii)* then follows from (76) together with (80) and (81).

Case (iii): $c < 1/2$. Our starting point is again (77). By Kolmogorov's LLN,

$$\lim_{t\to\infty} t^c \xi_t = \lim_{t\to\infty} v_t = \sum_{i=1}^{\infty} \theta_i \frac{\varepsilon_i}{i^{1-c}} = v$$

is finite with probability one. As to w_t,

$$\mathbf{E}w_t^2 = O(1) \sum_{i=1}^{t} i^{2(c-1)} = O(1),$$

so that $\mathbf{E}\eta_t^2 = O\left(t^{-2(1+c)}\right)$ and

$$\mathbf{E}\sum_{t=1}^{\infty} \left(t^c \eta_t\right)^2 < \infty.$$

Therefore, with probability one, $\lim_{t\to\infty} t^c \eta_t = 0$. Hence, by (77),

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

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. CM18 on this issue.

Remark 15. *In the proof of Theorem 5 below we will need the asymptotic behaviour of the means*

$$\overline{a}_T = \frac{1}{T}\sum_{t=1}^{T} a_t \quad \textit{and} \quad \overline{a}_T^- = \frac{1}{T}\sum_{t=1}^{T} a_{t-1}$$

in Case (i) and Case (iii) since both appear in the formula for the OLS estimator. For Case (i), it follows from Theorem 2 that

$$|\overline{a}_T| \leq \frac{1}{T}\sum_{t=1}^{T}|a_t| = O(1)\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{\ln_2 t}{t}} = O\left(\sqrt{\frac{\ln_2 T}{T}}\right) \tag{82}$$

since

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

In Case (iii), we have

$$\begin{aligned}\overline{a}_T &= \frac{1}{T}\sum_{t=1}^{T}a_t = \frac{1}{T}\sum_{t=1}^{T}t^{-c}t^{c}a_t = \frac{1}{T}\sum_{t=1}^{T}t^{-c}\left(u+o(1)\right)\\ &= \frac{u}{1-c}\frac{1}{T^c} + o\left(T^{-c}\right).\end{aligned} \tag{83}$$

Since $\overline{a}_T$ and $\overline{a}_T^-$ differ only by $(1/T)\left(a_T - a_0\right)$ and $a_T = o(1)$, the asymptotic behaviour of $\overline{a}_T^-$ is the same as that of $\overline{a}_T$.

6.2. Proof of Theorem 5

6.2.1. Generalities

As in the proof of Theorem 2, 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 (74) and (75). As is readily seen, the OLS estimator $\widehat{\beta}_T$ 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

$$y_t = \alpha + \beta a_{t-1} + \varepsilon_t, \tag{84}$$

and the dynamics

$$a_t = \left(1 - \frac{c}{t}\right)a_{t-1} + \frac{\gamma}{t}\varepsilon_t. \tag{85}$$

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

Formally, (84) resembles (22) in Section 4, apart from the different notation for the time parameter (t instead of n), structural parameters α instead of μ and β instead of λ, and the regressors a_{t-1} instead of y_{n-1}. With these replacement, the OLS estimators may therefore be written

$$\widehat{\beta}_T - \beta = \frac{u_T}{A_T} - \frac{\overline{a}_T^-}{A_T}\sum_{t=1}^{T}\varepsilon_t, \tag{86a}$$

$$\widehat{\alpha}_T - \alpha = \left(\widehat{\beta}_T - \beta\right)\overline{a}_T^- + \overline{\varepsilon}_T, \tag{86b}$$

cf. (23). The statistics appearing in (86) together with those appearing in the formulae below are defined as in (12) in Section 3.1.1, cf. also Section 4.1.1. The goal is again to find deterministic sequences φ_T and ψ_T such that $\varphi_T(\widehat{\beta}_T - \beta) = O(1)$ and $\psi_T(\widehat{\alpha}_T - \alpha) = O(1)$. Only the separate approach will be considered.

The crucial point is that the analysis in Section 4 does not depend on the fact that the regressors are predetermined values of y_n, but only on the behaviour of the basic statistics $\overline{y}_n, \overline{y}_n^-, A_n$ and A_n^0 and the derived ones. In the present model, this corresponds to $\overline{a}_T, \overline{a}_T^-, A_T$ and A_T^0. As a consequence, 'all' we have to do is to verify for the model in (84) with regressors (85) the crucial conditions (32) and (34) from Section 4.1.1.1. Phrased in the notation of the present model for easy reference, we need to check whether

$$\frac{\varphi_T}{\varphi\left(A_T^0\right)} = O\left(1\right), \tag{87a}$$

$$\frac{A_T^0}{A_T} = O\left(1\right) \tag{87b}$$

as well as

$$\varphi_T \sqrt{T \ln_2 T} \frac{\bar{a}_T^-}{A_T} = O\left(1\right) \tag{88}$$

are satisfied. The functions $\varphi = \varphi_i$ are defined as in Section 4.1.1.1.

6.2.2. Asymptotics of the basic statistics

Apart from evaluating the asymptotic behaviour of the basic statistics we will check to validity of condition (2) in Section 4.1.1.1. For reference, we repeat it here in the actual notation:

$$\frac{a_T^2}{\left(A_T^0\right)^{\gamma}} = o(1) \text{ for some } \gamma > 0. \tag{89}$$

Case (i): $c > 1/2$. We will show that

$$\frac{A_T^0}{\ln T} \to \frac{\gamma^2 \sigma^2}{2c-1}. \tag{90}$$

Starting with (85) (remembering our renaming convention) the same algebraic manipulations as in CM18 yield

$$\left(2c-1\right) A_T^0 = -T a_T^2 + c^2 \sum_{t=1}^{T} \frac{1}{t} a_{t-1}^2 + \gamma^2 \sum_{t=1}^{T} \frac{1}{t} \varepsilon_t^2 + 2\gamma u_T - 2\gamma c \sum_{t=1}^{T} \frac{1}{t} a_{t-1} \varepsilon_t. \tag{91}$$

(no probabilistic arguments are involved). Now bring in the asymptotic behaviour of a_t established in Theorem 2*(i)*:

$$a_t = O\left(\sqrt{\frac{\ln_2 t}{t}}\right), \tag{92}$$

to analyse the individual terms on the right hand side of (91). The following four properties all hold with probability one:

1. $Ta_T^2 = O(\ln_2 T)$.
2. $\sum_{t=1}^T a_{t-1}^2/t = O(1)$. This is because, by (92), $M = \sup_t ta_{t-1}^2/\ln_2 t < \infty$ such that $\sum_{t=1}^T a_{t-1}^2/t = \sum_{t=1}^T (\ln_2 t/t^2)\cdot(ta_{t-1}^2/\ln_2 t) \leq M\sum_{t=1}^T \ln_2 t/t^2 = O(1)$.
3. $\sum_{t=1}^T \varepsilon_t^2/t = \sigma^2 \ln T + O(1)$. This follows from the decomposition $\nu_t = \varepsilon_t^2 - \sigma^2$, applying the strong LLN for i.i.d. sequences to ν_t: $\sum_{t=1}^T \varepsilon_t^2/t = \sum_{t=1}^T \sigma^2/t + \sum_{t=1}^T \nu_t/t = \sigma^2 \ln T + O(1)$.
4. $\sum_{t=1}^T \frac{1}{t} a_{t-1}\varepsilon_t = O(1)$. This is due to Chow's local martingale convergence theorem, see (Lai & Wei, 1982a, equation (2.7)).

Hence

$$(2c-1)\, A_T^0 = \gamma^2\sigma^2 \ln T + O(\ln_2 T) + 2\gamma u_T. \tag{93}$$

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 Γ of positive probability. Then, by the martingale convergence theorem, u_T converges a.s. on Γ to some finite limit. Dividing (93) 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 Γ, the right hand side converges to ∞, 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

$$\frac{u_T}{A_T^0} \to 0.$$

Dividing (93) by A_T^0 we now obtain

$$(2c-1) = \gamma^2\sigma^2 \frac{\ln T}{A_T^0}\,[1 + o(1)] + o(1). \tag{94}$$

This shows (90).

Making use of (82), we find that

$$\frac{A_T}{A_T^0} = 1 + O\left(\frac{\ln_2 T}{\ln T}\right) \tag{95}$$

and

$$\frac{\overline{a}_T^-}{A_T} = O\left(\frac{1}{\ln T}\sqrt{\frac{\ln_2 T}{T}}\right). \tag{96}$$

Finally, taking account of (90) and (92),

$$\frac{a_T^2}{\left(A_T^0\right)^\gamma} = O\left(\frac{\ln_2 T}{T\,(\ln T)^\gamma}\right) = o\,(1) \tag{97}$$

for all γ. Therefore condition (89) is satisfied.

Remark 16. *For $c = 1/2$, (94) only shows that $A_T^0/\ln T \to \infty$, so that it does not allow the determination of the exact speed of divergence.*

Case (ii): $c < 1/2$. Recall that $c > 0$. Define $x_t = t^c a_{t-1}$ and $\beta_t = t^{-2c}$. Then by Theorem 2*(iii)*, $x_t^2 \to u^2$ and $b_T = \sum_1^T \beta_t \to \infty$ such that $b_T/T^{1-2c} \to 1/(1-2c)$. Now use the Toeplitz Lemma:

$$\frac{A_T^0}{T^{1-2c}} = \frac{\sum_1^T \beta_t x_t^2}{b_T}\frac{b_T}{T^{1-2c}} \to \frac{u^2}{1-2c}. \tag{98a}$$

Similarly,

$$\frac{A_T}{T^{1-2c}} \to v^2 \tag{98b}$$

where $v^2 = c^2u^2/((1-c)^2(1-2c))$. Consequently,

$$\lim_{T\to\infty}\frac{A_T^0}{A_T} = \left(1 - \frac{1}{c}\right)^2. \tag{99}$$

Also, making use of (83) together with (98b), it turns out that

$$\frac{\overline{a}_T^-}{A_T} = \frac{1}{T^{1-c}}w\,(1 + o(1)) \tag{100}$$

with $w \neq 0$ a.s.. Finally, let us consider condition (89). In view of Theorem 2*(iii)* and (98a),

$$\frac{a_T^2}{\left(A_T^0\right)^\gamma} = O\left(\frac{T^{-2c}}{T^{(1-2c)\gamma}}\right) = O\left[T^{-2c-(1-2c)\gamma}\right]. \tag{101}$$

Hence (89) is fulfilled for all $\gamma > 0$.

6.2.3. Consistency

Case (i): $c > 1/2$. As indicated in Section 6.2.1, what we have to do is to find deterministic sequences φ_T such that conditions (87) and (88) are satisfied. Straightforward calculation shows that

$$\varphi_1\left(A_T^0\right) = \sqrt{\frac{A_T^0}{\left(\ln A_T^0\right)^{1+\eta}}} = \sqrt{\frac{r\ln T}{\left(\ln_2 T\right)^{1+\eta}}}\,(1 + o(1)),$$

$$\varphi_2\left(A_T^0\right) = \sqrt{\frac{A_T^0}{\ln_2 A_T^0}} = \sqrt{\frac{r\ln T}{\ln_3 T}}\,(1 + o(1)),$$

with $r = \gamma^2\sigma^2/(2c-1) > 0$. In view of (87a) this yields as candidates for the normalising sequences $\varphi_T^1 = \varphi_1(\ln T)$ or $\varphi_T^2 = \varphi_2(\ln T)$, according to the prevalent scenario, cf. Section 4.1.1.1. As to (88), it follows from (96) that

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

and

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

Hence condition (88) is satisfied for both choices of the normalising sequence φ_T. Condition (87b) is satisfied by virtue of (99). Summarising, we arrive at the following conclusions:

1. If ε_t has moments up to second order, then the rate of a.s. convergence of the OLS estimator is $\varphi_T = (\ln T/(\ln_2 T)^{1+\eta})^{1/2}$ for every $\eta > 0$.
2. If $\mathbf{E}|\varepsilon_t|^p < \infty$ for some $p > 2$, then also $\eta = 0$ will do. However, in view of (97), we may apply MCT 2 to obtain $\varphi_T = (\ln T/\ln_3 T)^{1/2}$ as a normalising sequence.

Case (ii): $c < 1/2$. From the results in Section 6.2.2 it readily follows that

$$\varphi_1\left(A_T^0\right) = \sqrt{\frac{A_T^0}{(\ln A_T^0)^{1+\eta}}} = w\sqrt{\frac{T^{1-2c}}{(\ln T)^{1+\eta}}}(1+o(1)),$$
$$\varphi_2\left(A_T^0\right) = \sqrt{\frac{A_T^0}{\ln_2 A_T^0}} = w'\sqrt{\frac{T^{1-2c}}{\ln_2 T}}(1+o(1))$$

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

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

both qualify as candidates for the normalisation of the OLS estimator, in the sense that they satisfy (87a). Condition (87b) is fulfilled in view of (99). It remains to verify (88). By (100),

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

Similarly for φ_T. Summarising, we arrive at the following conclusions:

1. If ε_t has moments up to second order, then the rate of a.s. convergence of the OLS estimator is $\varphi_T = (T^{1-2c}/(\ln T)^{1+\eta})^{1/2}$ for every $\eta > 0$.
2. If $\mathbf{E}|\varepsilon_t|^p < \infty$ for some $p > 2$, then also $\eta = 0$ will do. Again, due to (101), we may apply MCT 2 to obtain $\psi_T = (T^{1-2c}/\ln_2 T)^{1/2}$ as a normalising sequence.

References

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

Apostol, T. M. (1974). *Mathematical Analysis*, 2 ed. Addison-Wesley, Reading, MA. MR0344384

Benveniste, A., Métivier, M. and Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximation*. Springer, Berlin. Orginally published in French in 1987. MR1082341

Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* **16** 367–401. MR0924877

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

Chow, Y. S. (1965). Local convergence of martingales and the law of large numbers. *Annals of Mathematical Statistics* **36** 552–558. MR0182040

Chow, Y. S. and Teicher, H. (1973). Iterated logarithm laws for weighted averages. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **26** 87–94. MR0331488

Christopeit, N. (1986). Quasi-least-squares estimation in semimartingale regression models. *Stochastics* **16** 255–278. MR0837614

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

Christopeit, N. and Massmann, M. (2018). Estimating structural parameters in regression models with adaptive learning. *Econometric Theory* **34** 68–111. MR3739236

Donsker, M. D. and Varadhan, S. R. S. (1977). On the laws of the iterated logarithm for local times. *Communications on Pure and Applied Mathematics* **30** 707–753. MR0461682

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

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

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

Lai, T. L. (2003). Stochastic approximation. *Annals of Statistics* **31** 391–406. MR1983535

Lai, T. L. and Wei, C. Z. (1982a). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics* **10** 154–166. MR0642726

Lai, T. L. and Wei, C. Z. (1982b). Asymptotic properties of projections with applications to stochastic regression problems. *Journal of Multivariate Analysis* **12** 346–370. MR0666011

Lai, T. L. and Wei, C. Z. (1983a). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *Journal of Multivariate Analaysis* **13** 1–23. MR0695924

Lai, T. L. and Wei, C. Z. (1983b). A note on martingale difference sequences satisfying the local Marcinkiewicz-Zygmund condition. *Bulletin of the Institute of Mathematics, Academica Sinica* **11** 1–13. MR0718896

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

Ljung, L. (1977). Analysis of recursive stochastic systems. *IEEE Transactions on Automatic Control* **AC-22** 551–575. MR0465458

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

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

Milani, F. (2007). Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics* **54** 2065–2082. MR2457750

Nielsen, B. (2005). Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms. *Econometric Theory* **21** 534–561. MR2162760

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* **74** 535–547. MR0909357

Phillips, P. C. B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory* **23** 557–614. MR2364380

Phillips, P. C. B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics* **136** 115–130. MR2328587

Phillips, P. C. B. and Magdalinos, T. (2008). Limit theory for explosively cointegrated systems. *Econometric Theory* **24** 865–887. MR2428852

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

Sargent, T. J. (1999). *The Conquest of American Inflation*. Princeton University Press, Princeton.

Shiryaev, A. N. (1996). *Probability*, 2 ed. Springer, New York. 1st edition 1984. MR1368405

Solo, V. and Kong, X. (1995). *Signal Processing Algorithms*. Prentice Hall, Upper Saddle River.

Stout, W. F. (1970). The Hartmann-Wintner law of the iterated logarithm for martingales. *Annals of Mathematical Statistics* **41** 2158–2160.

Wang, X. and Yu, J. (2015). Limit theory for an explosive autoregressive process. *Economics Letters* **126** 176–180. MR3298417

Wei, C. Z. (1985). Asymptotic properties of least-squares estimates in stochastic regression models. *Annals of Statistics* **13** 1498–1508. MR0811506