ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Christopeit, Norbert; Massmann, Michael

Working Paper

Consistent Estimation of Structural Parameters in Regression Models with Adaptive Learning

Tinbergen Institute Discussion Paper, No. 10-077/4

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Christopeit, Norbert; Massmann, Michael (2010) : Consistent Estimation of Structural Parameters in Regression Models with Adaptive Learning, Tinbergen Institute Discussion Paper, No. 10-077/4, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/86764

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

TI 2010-077/4
Tinbergen Institute Discussion Paper

Consistent Estimation of Structural Parameters in Regression Models with Adaptive Learning

Norbert Christopeit[1]
Michael Massmann[2]

[1] *University of Bonn;*
[2] *VU University Amsterdam.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

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

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

Consistent estimation of structural parameters in regression models with adaptive learning

Norbert Christopeit
University of Bonn*

Michael Massmann
Vrije Universiteit Amsterdam

August 2010

Abstract

In this paper we consider regression models with forecast feedback. Agents' expectations are formed via the recursive estimation of the parameters in an auxiliary model. The learning scheme employed by the agents belongs to the class of stochastic approximation algorithms whose gain sequence is decreasing to zero. Our focus is on the estimation of the parameters in the resulting actual law of motion. For a special case we show that the ordinary least squares estimator is consistent.

Keywords: Adaptive learning, forecast feedback, stochastic approximation, linear regression with stochastic regressors, consistency

1 Introduction

Rational expectations have been criticised in the economic literature as imputing too much knowledge to economic agents. As an alternative way for agents to form expectations, adaptive learning has been suggested as a form of bounded rationality, see e.g. Sargent (1993). Agents are thereby thought of as econometricians and, in every time period, estimate what they perceive as law of motion of the economy with a view to constructing their forecasts, updating parameter estimates and expectations recursively as new information becomes available. The updating recursion of the parameters is usually specified as an instance of a stochastic recursive algorithm, with recursive least squares and constant gain least squares as prominent examples, cf. Evans and Honkapohja (2001). The learning mechanism creates a so-called forecast feedback in the model which makes the behaviour of the entire system much more complex than may seem at first sight. The literature on adaptive learning has so far to a large extent focused on the question of whether and, if so, under what conditions agents' expectations converge to a rational expectations equilibrium; cf. Evans and Honkapohja (2008).

*Address correspondance to Norbert Christopeit, Department of Economics, University of Bonn, Adenauerallee 24-42, D-53113 Bonn,Germany; e-mail: christopeit@uni-bonn.de.

In the following we will refer to this convergence issue as the 'internal forecasting problem' (IFP). The bottom line of this literature is that, in standard setups, recursive least squares or other decreasing gain learning algorithms allow agents to become rational eventually, while constant gain least squares leads to persistent learning dynamics.

Recently, research interests have shifted to an empirical analysis of models with adaptive learning. For instance, using U.S. data, Milani (2006, 2007) and Chevillon, Massmann, and Mavroeidis (2010) adopt Bayesian and frequentist methods, respectively, to estimate and conduct inference on the structural parameters in New Keynesian Phillips curve models in which agents' expectations are formed by adaptive learning algorithms. So as to differentiate the empirical estimation of structural parameters in models with forecast feedback from the aforementioned IFP, we will henceforth refer to it as the 'external estimation problem' (EEP). In this context, however, at least two crucial issues have not yet been addressed by the literature: First, due to the self-referential, complex dynamics introduced into the models by the forecast feedback little is known about the statistical properties of the estimators of the structural parameters. To our knowledge, the only result available to date is derived by Chevillon et al. (2010) who establish a local asymptotic approximation to the distribution of the ordinary least squares estimator. Secondly, and relatedly, the question of which learning algorithms are compatible with the consistent estimation of the structural parameters has not been conclusively answered yet. In particular, Chevillon et al. (2010) argue that the identification of models with forecast feedback may break down when agents engage in decreasing gain learning. It hence appears as if learning models in which agents become rational eventually cannot be estimated consistently.

The main purpose of the present paper is to address these two issues. We consider a bivariate structural model whose two explanatory variables are a constant and an expectations term, the latter of which arises from agents' adaptive forecasts. Our focus will be on recursive learning schemes with decreasing gain. For a special class of decreasing gain sequences, which includes recursive least squares learning, we establish consistency of the ordinary least squares (OLS) estimator of the two structural parameters. These results are noteworthy in several regards. From an economic point of view, the consistency result is remarkable since research on the IFP appears to imply that when the forecasts of agents converge to the rational expectations equilibrium the two structural parameters may become what is commonly called 'strictly asymptotically unidentified'. Our results, however, show that convergence to rationality and consistent estimation of structural parameters are not mutually exclusive.

From a statistical perspective, the EEP presents itself as a simple regression model with a predetermined stochastic regressor. The regressor sequence is such that it apparently does not fit into any of the well established scenarios for consistency of the OLS estimator in regression models with stochastic regressors (cf. Lai and Wei (1982b), Lai and Wei (1982a)). Therefore, we shall take up the approach in Lai and Robbins (1977), where, for simple regression models with a deterministic regressor, a minimal (i.e. also necessary) condition for consistency

is derived. We show that this condition remains to be sufficient for a certain class of models with stochastic regressors and that a sharpened version of it is even necessary. These results are of interest since they complement the existing statistical literature in a useful way. Still, due to the complexity of the regressors, the application to our actual model of interest requires some preliminary steps to tailor the estimator to a form amenable to the methods just mentioned.

Several other properties of the OLS estimator, which are of interest in their own right, are only touched on. One of these is the asymptotic normality at rate $\sqrt{\ln T}$, which, to our knowledge, does not have any counterpart among commonly used econometric models. Also, as a byproduct of the preliminary considerations, consistency turns out to obtain for constant gain learning.

The outline of the paper is as follows: Following this Introduction, Section 2 provides a brief exposition of models with forecast feedback as an alternative to models with rational expectations. The internal forecasting problem (IFP) and the external estimation problem (EEP) arising in models with forecast feedback are discussed in Section 3. This will have set the scene for the main part of our analysis in Section 4 where we first give a synopsis of the statistical literature on consistent estimation in linear models and present a new sufficiency results for strongly consistent OLS estimation in linear regression models with predetermined stochastic regressors (Section 4.1). Then, in Section 4.2, we establish consistency of the OLS estimator in our learning model. Section 5 concludes. The proofs of the two main theorems are relegated to Appendices A and B.

2 An economic model with forecast feedback

We consider the economic model

$$y_t = \beta y^e_{t|t-1} + \delta' x_t + \varepsilon_t, \quad t = 1, 2, \dots . \tag{2.1}$$

Concerning the noise ε_t, we make the following

Maintained assumption 1. The ε_t are iid with mean 0 finite variance σ^2.

The exogenous variables x_t are K - dimensional random vectors, whose interaction with the noise ε_t will be governed by assumptions specific for the various approaches discussed below. $y^e_{t|t-1}$ models agents' expectations about y_t based on the information available at time $t-1$.

Maintained assumption 2. $\beta \neq 1$.

Models of the type as in (2.1) have a long tradition in economics. For instance, the classical cobweb model fits into this form, see for instance Bray and Savin (1986), as does the Lucas (1973) aggregate supply model. The 'classical' way of modelling the expectational term $y^e_{t|t-1}$ in (2.1) is via rational expectations: for completeness, this will be briefly summarised in Section 2.1 before the approach we concentrate on in this paper, viz. modelling $y^e_{t|t-1}$ by adaptive learning, is outlined in Section 2.2.

2.1 Rational expectations

The information available at any instant of time t is modelled by the filtration $(\mathcal{F}_t^0)$, where $\mathcal{F}_t^0 = \sigma\left(y_s, s \leq t; x_s, s \leq t+1\right)$,

$$y_{t|t-1}^e = \mathbf{E}\left(y_t | \mathcal{F}_{t-1}^0\right). \tag{2.2}$$

Inserting this into (2.1) and taking conditional expectations yields

$$\mathbf{E}\left(y_t | \mathcal{F}_{t-1}^0\right) = \delta' x_t + \beta \mathbf{E}\left(y_t | \mathcal{F}_{t-1}^0\right); \tag{2.3}$$

note that x_t is $\mathcal{F}_{t-1}^0$ - measurable. Combining (2.2) and (2.3) we obtain

$$y_{t|t-1}^e = \alpha' x_t \tag{2.4a}$$

with

$$\alpha = \frac{\delta}{1-\beta}. \tag{2.4b}$$

The so-called rational expectations equilibrium (REE) then is

$$y_t = \alpha' x_t + \varepsilon_t. \tag{2.5}$$

Note that only α is identified, not, however, δ and β separately.

2.2 Adaptive learning

The basic idea underlying all adaptive learning procedures to be discussed is that agents employ an *auxiliary model,* or *perceived law of motion (PLM)* (a potentially ficticious model which is generally misspecified)

$$y_t = \alpha' z_t + \varepsilon_t \tag{2.6}$$

to explain the quantity they want to predict. The z_t are auxiliary variables which the agents consider as relevant for the evolution of y_t. The information available at time t is modelled by $\mathcal{F}_t = \sigma\left(y_s, s \leq t; x_s, z_s, s \leq t+1\right)$. Note that the true regressors x_t are *predictable* w.r.t. the filtration $(\mathcal{F}_t)$. An important special case is obtained when $z_t = x_t$ and will be referred to as *proper auxiliary model.* If α is unknown (as will generally be the case), agents are assumed to replace α by some estimate a_{t-1} based on the information $\mathcal{F}_{t-1}$. Several choices of a_{t-1} are conceivable.

Common choices

1. Ordinary least squares (OLS)

The OLS estimator based on model (2.6) is given by

$$a_t = P_t^{-1} \sum_{s=1}^{t} z_s y_s \tag{2.7}$$

with $P_t = \sum_{s=1}^{t} z_s z_s'$. The prediction of y_t is then

$$y_{t|t-1}^e = a_{t-1}' z_t, \tag{2.8}$$

and the resulting *actual law of motion (ALM)*, or *data generating process (DGP)* is

$$y_t = \beta a'_{t-1} z_t + \delta' x_t + \varepsilon_t. \tag{2.9}$$

The ALM for the corresponding proper auxiliary model is

$$y_t = (\delta + \beta a_{t-1})' x_t + \varepsilon_t. \tag{2.10}$$

2. Recursive least squares (RLS) and its generalizations
By the matrix inversion lemma, the OLS estimator may be recursively calculated from

$$a_t = a_{t-1} + \frac{1}{t} R_t^{-1} z_t \left(y_t - a'_{t-1} z_t \right), \quad R_t = \frac{1}{t} P_t, \tag{2.11a}$$

with initial values $a_0 = 0, R_0 = 0$. Note that R_t may itself be calculated from the recursion

$$R_t = R_{t-1} + \frac{1}{t} \left(z_t z'_t - R_{t-1} \right). \tag{2.11b}$$

For arbitrary initial values a_0, R_0 the estimator computed from (2.11) will be called recursive least squares estimator.

(2.11) is a very special case of general stochastic approximation algorithms. In particular, it may be made less specific by substituting a general weighting, or gain, sequence γ_t for $1/t$:

$$a_t = a_{t-1} + \gamma_t R_t^{-1} z_t \left(y_t - a'_{t-1} z_t \right), \tag{2.12a}$$

$$R_t = R_{t-1} + \gamma_t \left(z_t z'_t - R_{t-1} \right). \tag{2.12b}$$

(2.12) describes a rather general class of *adaptive* or *error-correction* learning schemes. The special case (2.11) will henceforth be referred to as "recursive least squares learning". For certain purposes, the recursive law for the weighting matrices R_t^{-1} will not be of interest, in which cases we will consider the adaptive scheme

$$a_t = a_{t-1} + \gamma_t M_t z_t \left(y_t - a'_{t-1} z_t \right), \tag{2.13}$$

for a given sequence M_t satisfying certain convergence properties.

3. Stochastic gradient algorithm
This variant is obtained from (2.11a) by replacing R_t by its trace $r_t = \frac{1}{t} \sum_{s=1}^{t} z'_s z_s$:

$$a_t = a_{t-1} + \frac{1}{t} \frac{z_t}{r_t} \left(y_t - a'_{t-1} z_t \right). \tag{2.14}$$

With learning scheme like (2.12), (2.13) or (2.14), agents' expectation will still be given by (2.8), and the corresponding ALM is again (2.9) or (2.10). It is plain that, in models with adaptive learning, the expectational term $y^e_{t|t-1} = a'_{t-1} z_t$ thus creates a forecast feedback, resulting in a self-referential, and thus highly complex, DGP.

3 Forecasting and estimation

Our main focus in the context of model (2.9) is the consistent estimation of the structural parameters β and δ. In order to investigate this question it will be advantageous to proceed in two steps:
1. *Internal forecast problem (IFP)*: Will adaptive learning finally lead to the REE, i.e., is it true that $a_{t-1}z_t \sim \alpha x_t$ in some probabilistic sense as $t \to \infty$?
2. *External estimation problem (EEP)*: Will the ALM resulting from adaptive learning allow for the consistent estimation of the structural parameters δ, β?

Answering the second question will be tantamount to investigating the asymptotic behaviour of OLS estimators in linear regression models with predetermined stochastic regressors. The difficulty with our setup is that these regressors are functions of lagged endogenous variables determined via a recursive algorithms and not given in closed form. It turns out that the asymptotic behaviour of the regressors differs drastically from that of stationary or integrated regressors in traditional linear autoregressive models.

3.1 Internal forecasting problem

Various approaches to dealing with the internal forecasting problem have been suggested in the literature.

A. The martingale approach

The approach is based on a convergence theorem for almost supermartingales (cf. Robbins and Sigmund (1971)) and is mainly applied to OLS-learning. Usual assumptions made in this approach are the following; see for instance Bray and Savin (1986).
(A1) $z_t = x_t$, and the x_t are iid with finite fourth moments and $\overline{M}_{xx} = \mathbf{E}x_t x_t'$ is positive definite.
(A2) The ε_t are iid with mean zero and finite positive variance.
(A3) (x_t) and (ε_t) are independent processes.

B. The stochastic approximation (SA) approach

This approach is based on a convergence theorem for linear approximation schemes in a Banach space (cf. Walk (1985)). The application to our adaptive learning schemes above has been considered in Kottmann (1990). See also Walk and Zsidó (1989).

C. The ODE approach

In this approach, promoted by Benveniste, Métivier, and Priouret (1990), Ljung (1977) and Marcet and Sargent (1989a), an ordinary differential equation (ODE) is associated with the discrete difference equations (2.12). Under suitable regularity conditions, generally referred to as 'expectational stability' conditions, see Marcet and Sargent (1989b) and Evans (1989), the OLS estimates a_t converge to a stable equilibrium point of this ODE. These conditions are, however, hard to verify and will in general require a modification of the OLS estimator, e.g. the so-called projection facility, Evans and Honkapohja (1998). A rigorous treatment of the ODE approach has been given in Benveniste et al. (1990). Since it is designed

to cover a much broader class of approximation schemes, including nonlinear procedures, the verification of the assumptions for our simple model (2.12) turns out rather cumbersome.

In Subsections 3.1.1 and 3.1.2 we will consider the question whether two popular instances of the stochastic approximation algorithm (2.12) converge, namely so-called constant gain and decreasing gain learning, respectively. The analysis of the latter will be based on the SA approach and draw mainly on Kottmann (1990). The results discussed in this section will be used in the examination of the EEP in Section 4.

3.1.1 Constant gain learning

The simplest special case of the stochastic approximation algorithm in (2.12) is obtained by setting $\gamma_t = \gamma \in (0,1)$ for all t. Then the recursion (2.12) takes the form

$$\begin{aligned} a_t &= a_{t-1} + \gamma R_t^{-1} z_t \left(y_t - a_{t-1}' z_t\right), && \text{(3.1a)} \\ R_t &= R_{t-1} + \gamma \left(z_t z_t' - R_{t-1}\right). && \text{(3.1b)} \end{aligned}$$

This is a prototype for a model in which agents' forecasts do not converge to the rational expectations equilibrium. Put differently, neither a_t nor R_t will converge in probability in general.

To see this, consider R_t first and assume z_t to be iid for simplicity. Then (3.1b) is a stable (matrix valued) AR(1)-model possessing a stationary ergodic solution

$$R_t^0 = \gamma \sum_{i=0}^{\infty} (1-\gamma)^i z_{t-i} z_{t-i}'.$$

Whatever the starting value R_0 may be, R_t will converge in distribution to the invariant distribution of R_t^0. In general, no stronger kind of convergence will obtain. The sequence R_t need not even be a.s. bounded: cf. Lai and Wei (1985) for an example where $\mathbf{P}\left(\sup_t |R_t| = \infty\right) = 1$.

No matter whether R_t converges, a_t will generally not. This may be demonstrated by examining a simple example: let x_t be a scalar and set $z_t = 1$. Then $R_t = 1$ is the stationary solution to (3.1b), and (3.1a) becomes

$$a_t = a_{t-1} + \gamma \left(y_t - a_{t-1}\right). \tag{3.2}$$

The ALM (2.9) takes the form

$$y_t = \delta x_t + \beta a_{t-1} + \varepsilon_t. \tag{3.3}$$

Inserting this into (3.2) and introducing the new parameter $c = 1 - \gamma(1-\beta)$, yields

$$a_t = c a_{t-1} + \gamma \left(\delta x_t + \varepsilon_t\right). \tag{3.4}$$

If $|c| < 1$, which will be the case for γ sufficiently small and $\beta < 1$, (3.4) is a stable AR(1)-model. Assuming that (x_t, ε_t) is a strictly stationary ergodic sequence with finite second moments, (3.4) possesses a stationary ergodic solution

$$a_t^0 = \gamma \sum_{i=0}^{\infty} c^i \left(\delta x_{t-i} + \varepsilon_{t-i}\right).$$

Denoting $\mu_x = \mathbf{E}x_t$, its mean and variance are given by

$$\mu_0 = \frac{\gamma\delta}{1-c}\mu_x = \alpha\mu_x \quad \text{and} \quad \sigma_0^2 = \int_{-\pi}^{\pi} \frac{f(\lambda)}{1 + c^2 - 2\cos\lambda} d\lambda,$$

respectively, where $f(\lambda)$ is the spectral density of the process $(\delta x_t + \varepsilon_t)$. Apparently, $\sigma_0^2 > 0$. Since a_t converges to a_1^0 in distribution, this means that convergence $a_t \to \alpha$ cannot hold in any probabilistic sense. Unless $\mu_x = 1$, not even $\mathbf{E}a_t \to \alpha$ (a sort of convergence to the REE in the mean) will hold. On the other hand, a_t possesses all the nice asymptotic properties of a stationary ergodic process. In particular,

$$\frac{1}{T}\sum_{t=1}^{T} a_{t-1}^2 \xrightarrow{P} \mathbf{E}\left[a_{t-1}^0\right]^2, \quad \frac{1}{T}\sum_{t=1}^{T} a_{t-1}x_t \xrightarrow{P} \mathbf{E}\left[a_{t-1}^0 x_t\right], \quad \frac{1}{T}\sum_{t=1}^{T} a_{t-1}\varepsilon_t \xrightarrow{P} 0. \tag{3.5}$$

This follows from the ergodicity of the process $(a_t^0, x_t, \varepsilon_t)$, together with the fact that $\mathbf{E}\left|a_t - a_t^0\right|^2 = O(c^{2t})$ such that the empirical moments in (3.5) differ from the corresponding terms with a_{t-1} replaced by a_{t-1}^0 by an expression which tends to zero in L^1.[1]

To be more specific about (3.5), consider the special case where the x_t are iid and independent of the ε_t, with variance σ_x^2. Then

$$\sigma_0^2 = \frac{\gamma^2\left(\delta^2\sigma_x^2 + \sigma^2\right)}{1-c^2}$$

and the limits in (3.5) take the form

$$\mathbf{E}\left[a_{t-1}^0\right]^2 = \sigma_0^2 + \alpha^2\mu_x^2, \quad \mathbf{E}\left[a_{t-1}^0 x_t\right] = \alpha\mu_x^2.$$

Therefore, the matrix

$$M_T = \frac{1}{T}\begin{pmatrix} \sum_{t=1}^{T} a_{t-1}^2 & \sum_{t=1}^{T} a_{t-1}x_t \\ \sum_{t=1}^{T} a_{t-1}x_t & \sum_{t=1}^{T} x_t^2 \end{pmatrix} \tag{3.6a}$$

appearing in the formula for the OLS estimator of (β, δ) in (3.3) tends in probability to the limit

$$\overline{M} = \begin{pmatrix} \sigma_0^2 + \alpha^2\mu_x^2 & \alpha\mu_x^2 \\ \alpha\mu_x^2 & \sigma_x^2 + \mu_x^2 \end{pmatrix}. \tag{3.6b}$$

Unless $x_t \equiv 0$ for all t, $\overline{M} = \overline{M}(\gamma)$ is regular as long as $\gamma > 0$. We will come back to this in subsection 3.2 below.

[1] Note that the gain parameter γ in our analysis is fixed. This is in sharp contrast to the setup used by Benveniste et al. (1990) in Chapter 4 of their Part II who, for a triangular scheme $t_n \to \infty$ and $\gamma(t_n) \to 0$ as $n \to \infty$, derive an asymptotic Normal distribution for a suitably scaled a_t.

3.1.2 Decreasing gain learning

We will now examine the recursive algorithm (2.12) with a gain sequence γ_t decreasing to zero and present an extract of results obtained by the SA approach, based on Kottmann (1990). We consider only the *stationary static case,* i.e. the case where the vector x_t of explanatory variables does not contain any lagged endogenous variable y_{t-k}. For the dynamic scenario we refer to Zenner (1996). We make the following assumptions:
(B1) $R_t \to R$ a.s. with R regular.
(B2) $\frac{1}{t}\sum_{s=1}^{t} z_s x_s' \to W$ a.s.
(B3) $\frac{1}{t}\sum_{s=1}^{t} z_s \varepsilon_s \to v$ a.s.
If we assume
(B4) $V_t = (x_t, z_t, \varepsilon_t)$ is a strictly stationary ergodic square integrable process with $\overline{M}_{zz} = \mathbf{E} z_t z_t'$ positive definite,
then (B1)-(B3) will be satisfied with W and v equal to the corresponding theoretical moments. We will refer to this as the *stationary ergodic* scenario further below.

In the following, we report the main results from Kottmann (1990). The first concerns OLS-learning.

Theorem 3.1 *Under assumptions (B1)-(B3), the limit a of the sequence a_t of OLS estimators (if it exists) must satisfy*

$$a = \frac{1}{1-\beta} R^{-1} \left[W\delta + v \right]. \tag{3.7}$$

In particular, if $v = 0$ and the proper auxiliary model $z_t = x_t$ is applied, a coincides with the REE $\alpha = \delta/(1-\beta)$. If $\beta < 1$, the limit does exist almost surely and is given by (3.7).

Cf. Theorems 4.2 and 4.3 in Kottmann (1990). The result remains true if instead of the proper OLS estimator the recursive least squares estimator (2.11) is used. The sequence R_t may be replaced by any sequence M_t satisfying $M_t \to M$ a.s. for some regular matrix M.

For the general adaptive scheme (2.12) or (2.13), somewhat stronger assumptions than (B4) are needed, which ensure that the correlation between lagged values of the process V_t decline to zero fast enough. We quote the following interesting special case from Kottmann (1990) (cf. Theorem 7.4 and Lemma 7.6 loc.cit.).

Theorem 3.2 *Suppose that (V_t) is a linear process, i.e.*

$$V_t = V_0 + \sum_{i=0}^{\infty} A_i u_{t-i},$$

where (u_t) are iid random vectors with mean zero and finite fourth moments, and the nonrandom matrices A_i satisfy

$$\sum_{i=0}^{\infty} \|A_i\| < \infty.$$

Let the sequence a_t be obtained from the recursive scheme

$$a_t = a_{t-1} + \gamma_t M_t z_t \left(y_t - a'_{t-1} z_t\right),$$

where the M_t are (possibly random) matrices such that
(i) $M_t \to M$ a.s. for some regular matrix M,
and the γ_t are nonnegative numbers satisfying
(ii) $\gamma_t \to 0$, $\sum_t \gamma_t = \infty$ and $\sum_t \gamma_t^2 (\log t)^2 < \infty$.
Then $a_t \to a$ a.s. with a given by

$$a = \frac{1}{1-\beta} M \left[W\delta + v\right].$$

Note that, under assumptions (i) and (ii) and for appropriate initial value V_0, V_t is a strictly stationary ergodic process with finite fourth moments. Convergence of a_t to the REE obtains as a special case, as shown in the following corollary.

Corollary 3.3 *In addition to the assumptions of Theorem 3.2, assume that $v = 0$ and $M_t = R_t^{-1}$, with the limit M deterministic. Then, in the proper auxiliary model (i.e. $z_t = x_t$), $a_t \to a$ a.s. with a given by*

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

Proof. In this case, $W = \overline{M}_{xx}$. By Lemma 3.4 below, $M = R^{-1} = \overline{M}_{xx}^{-1}$. ■

Actually, in many cases of interest, the summability condition in (ii) of Theorem 3.2 may be replaced by the weaker condition $\sum_t \gamma_t^2 < \infty$, cf., e.g. Christopeit and Massmann (2009), who also verify the conditions of Benveniste et al. (1990) for scalar x_t forming a strictly stationary homogeneous Markov chain possessing an invariant measure.

The following two lemmas examine the limiting behavior of R_t.

Lemma 3.4 *For the recursive scheme (2.12b), the limit R, provided it exists, is given by*

$$R = \lim_{t\to\infty} \frac{\sum_{s=1}^t \gamma_s z_s z'_s}{\sum_{s=1}^t \gamma_s}. \tag{3.8}$$

Moreover, $\mathbf{E}R = \overline{M}_{zz}$.

Proof. Summing both sides of (2.12b) over t, dividing by $s_T = \sum_{t=1}^T \gamma_t$ and passing to the limit we obtain

$$\lim_{T\to\infty} \frac{1}{s_T} \sum_{t=1}^T \gamma_t z_t z'_t = \lim_{T\to\infty} \frac{1}{s_T} \sum_{t=1}^T \gamma_t R_{t-1}.$$

The limit on the rhs is R. This shows (3.8). Consider the sequence $\widetilde{R}_T = s_T^{-1} \sum_{t=1}^T \gamma_t z_t z'_t$. Obviously, $\sup_T \mathbf{E}||\widetilde{R}_T||^2 < \infty$, so that the $\widetilde{R}_T$ are uniformly integrable. Therefore, since $R = \lim_{T\to\infty} \widetilde{R}_T$ by (3.8),

$$\mathbf{E}R = \lim_{T\to\infty} \mathbf{E}\widetilde{R}_T = s_T^{-1} \sum_{t=1}^T \gamma_t \mathbf{E} z_t z'_t = \overline{M}_{zz}.$$

■

Lemma 3.4 leaves open the question of existence of the limit R. The next lemma gives a partial answer.

Lemma 3.5 *Assume that* $\sum_t \gamma_t = \infty$ *and that the limit*

$$R = \lim_{t\to\infty} \sum_{i=1}^{t} (1-\gamma_t)\cdots\left(1-\gamma_{i+1}\right)\gamma_i z_i z_i' \tag{3.9}$$

exist as a finite matrix. Then, for the recursive scheme (2.12b), $\lim_{t\to\infty} R_t = R$.

Proof. Denote $\alpha_t = (1-\gamma_t)$. Then

$$R_t = R_0 \prod_{i=1}^{t} \alpha_i + \sum_{i=1}^{t} \phi_{ti} z_i z_i',$$

with the ϕ_{ti} given by

$$\phi_{ti} = \begin{cases} \gamma_t, & i = t, \\ \alpha_t \cdots \alpha_{i+1}\gamma_i, & i = t-1,\ldots,1. \end{cases}.$$

Since $\sum_t \gamma_t = \infty$, $\lim_{t\to\infty} \prod_{i=1}^{t} \alpha_i = 0$. ■

If applied to the scheme (2.11b), i.e. $\gamma_t = 1/t$, (3.9) yields

$$R = \lim_{t\to\infty} \sum_{i=1}^{t} \frac{t-1}{t}\cdots\frac{i}{i+1}\frac{1}{i} z_i z_i' = \lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} z_i z_i'.$$

Hence, for stationary ergodic z_t, the recursive scheme (2.11b) converges to $R = \overline{M}_{zz}$.

3.2 The external estimation problem

The EEP is concerned with the problem of *consistent estimation* of the structural parameters of the ALM. We approach this question by first considering, in Subsection 3.2.1, the case in which agents' PLM is not well-specified, i.e. the case in which the ALM is given by (2.9); it is reproduced here for convenience:

$$y_t = \beta a_{t-1}' z_t + \delta' x_t + \varepsilon_t. \tag{3.10a}$$

For decreasing gain learning in the stationary ergodic scenario, we present some simple positive consistency results. Subsequently, in Subsection 3.2.2, we turn to the case in which agents use the proper auxiliary model for forming their expectations, i.e. the case in which their PLM is well-specified, and thus examine the question of consistent estimation of β and δ in

$$y_t = \beta a_{t-1}' x_t + \delta' x_t + \varepsilon_t, \tag{3.10b}$$

see also (2.10). In doing so, the issue that lies at the core of the present investigation will become apparent such that the scene will be set for our detailed analysis in Section 4.

3.2.1 Some positive consistency results

In the following, we use the notation $M_T^{uv} = \frac{1}{T}\sum_{t=1}^T u_t v_t'$ for the empirical moments of two (vector-valued) processes u and v.

Proposition 3.6 *Assume that*
(i) the empirical second moments $M_T^{a'_{-1}za'_{-1}z}, M_T^{xa'_{-1}z}$ *and* M_T^{xx} *have finite limits in probability* $\overline{M}^{a'_{-1}za'_{-1}z}, \overline{M}^{xa'_{-1}z}$ *and* $\overline{M}^{xx}$*, respectively, such that the matrix*

$$\overline{M} = \begin{pmatrix} \overline{M}^{a'_{-1}za'_{-1}z} & \overline{M}^{a'_{-1}zx} \\ \overline{M}^{xa'_{-1}z} & \overline{M}^{xx} \end{pmatrix}$$

is a.s. nonsingular;
(ii) $\operatorname{plim}_{T\to\infty} M_T^{a'_{-1}z\varepsilon} = 0$ *and* $\operatorname{plim}_{T\to\infty} M_T^{x\varepsilon} = 0$.
Then the OLS estimator of (β, δ') *is weakly consistent. If all empirical moments in (i) converge with probability one, then strong consistency obtains.*

The proof follows trivially from the formula

$$\begin{pmatrix} \widehat{\beta}_T \\ \widehat{\delta}_T \end{pmatrix} - \begin{pmatrix} \beta_T \\ \delta_T \end{pmatrix} = \begin{bmatrix} M_T^{a'_{-1}za'_{-1}z} & M_T^{a'_{-1}zx} \\ M_T^{xa'_{-1}z} & M_T^{xx} \end{bmatrix}^{-1} \begin{bmatrix} M_T^{a'_{-1}z\varepsilon} \\ M_T^{x\varepsilon} \end{bmatrix}. \tag{3.11}$$

for the OLS estimator.

Immaterial as it may seem, Proposition 3.6 does allow to infer strong consistency if the process $(x_t, a_{t-1}, z_t, \varepsilon_t)$ mimics the behavior of a stationary ergodic sequence. As an example, consider constant gain learning as discussed in Subsection 3.1.1. In this case, the limit matrix $\overline{M}$ is given by (3.6b) and is positive definite for $\gamma > 0$. Also, condition (ii) is satisfied by virtue of (3.5) and since $\lim_{T\to\infty} M_T^{x\varepsilon} = \mathbf{E}x_t\varepsilon_t = 0$ a.s. Hence, constant gain learning allows for consistent estimation of the parameters in the ALM. This result is in fact independent on whether agents use the proper auxiliary model or some misspecified PLM as forecasting device.

If the adaptive sequence a_t converges (not necessarily to α), then the matrix $\overline{M}$ takes a more specific form.

Proposition 3.7 *Suppose that assumption (B4) is satisfied and that* $a_{t-1} \to a$ *a.s. Furthermore, assume that* $\mathbf{E}z_t\varepsilon_t = 0$ *and* $\mathbf{E}x_t\varepsilon_t = 0$*. Then the matrix* $\overline{M}$ *in Proposition 3.6 takes the form*

$$\overline{M} = \begin{pmatrix} a'\mathbf{E}\left(z_t z_t'\right) a & a'\mathbf{E}\left(z_t x_t'\right) \\ \mathbf{E}\left(x_t z_t'\right) a & \mathbf{E}\left(x_t x_t'\right) \end{pmatrix}. \tag{3.12}$$

In particular, then, the OLS estimator of (β, δ') *is strongly consistent provided* $\overline{M}$ *is regular with probability one.*

Proof. This is an immediate consequence of the fact that for any two sequences a_t, A_t s.t.

$$a_t \to a,\ \frac{1}{T}\sum_{t=1}^T A_t \to A \ \text{ and } \sup_T \frac{1}{T}\sum_{t=1}^T |A_t| < \infty$$

it follows that

$$\frac{1}{T}\sum_{t=1}^{T} a_t' A_t \to a' A.$$

■

Consequently, as long as the PLM is misspecified, strongly consistent estimation of the structural parameters is also feasible when agents' forecasts converge to the REE, as is the case when the learning scheme is subject to a decreasing gain sequence; cf. Subsection 3.1.2.

3.2.2 EEP for proper auxiliary models

A completely different picture presents itself when agents use the proper auxiliary model as forecasting device. We will now look at constant and decreasing gain learning scenarios in turn.

Regarding constant gain, we have seen in subsection 3.1.1 that, for the special model considered there, the sequence a_t converges in distribution to some nondegenerate random variable, so that, in particular, convergence $a_t \to \alpha$ cannot hold in any probabilistic sense. On the other hand, as to the EEP, we found in subsection 3.2.1 (cf. the discussion following Proposition 3.6) that the OLS estimator for both coefficients is (weakly) consistent as long as $\gamma > 0$.

In the context of decreasing gain, however, we can expect a.s. convergence of the a_t to some (possibly random) limit a (cf. subsection 3.1.2). Therefore, reconsider the setup of Proposition 3.7 and recall its assumption that there is no asymptotic collinearity between the regressors $a_{t-1}' z_t$ and x_t. For the proper auxiliary model, this would boil down to requiring the regularity of the matrix

$$\overline{M} = \begin{pmatrix} a' \\ I \end{pmatrix} \mathbf{E}\left(x_t x_t'\right) \begin{pmatrix} a & I \end{pmatrix}. \tag{3.13}$$

This is obviously impossible and the conclusion of Proposition 3.7 consequently no longer holds.

It hence appears as if there is a *tradeoff* between (internal) convergence to rational expectations and (external) consistent estimation: We have just seen that, on the one hand, convergence of a_t to a, as in the case of decreasing gain learning, implies that $\overline{M}$ is singular. Put differently, as a_t converges to some a, the coefficients in (3.10b) are no longer separately identified asymptotically: $y_t \sim \beta a' x_t + \delta' x_t + \varepsilon_t$. On the other hand, when a_t does not converge to the REE, as in the case of constant gain learning, the structural parameters may well be consistently estimable. This supposed tradeoff, however, is specious: In the context of decreasing gain learning, the singularity of $\overline{M}$ does not automatically imply that the OLS estimators $\widehat{\beta}_T$ and $\widehat{\delta}_T$ are not consistent. Specifically, it will be shown in Section 4 below that, for the special case where $x_t = 1$, the structural parameters β and δ in model (3.10b) may be consistently estimated by OLS although they are asymptotically non-identified.

4 Estimating the parameters of the ALM

Let us return to our model of interest (2.10), where a_t is obtained from the recursive scheme (2.12). Henceforth, we consider the scalar case $x_t = z_t = 1$. Then (2.10) and (2.12a) take the form

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t, \tag{4.1a}$$
$$a_t = a_{t-1} + \gamma_t (y_t - a_{t-1}), \tag{4.1b}$$

respectively, since $R_t = 1$ is the stationary solution to (2.12b). We are interested in the problem of consistent OLS estimation of the parameters δ, β. The OLS estimator is given by

$$\begin{pmatrix} \widehat{\beta}_T - \beta \\ \widehat{\delta}_T - \delta \end{pmatrix} = M_T^{-1} \begin{pmatrix} \frac{1}{T}\sum_{t=1}^T a_{t-1}\varepsilon_t \\ \frac{1}{T}\sum_{t=1}^T \varepsilon_t \end{pmatrix}, \tag{4.2}$$

with M_T as in (3.6a) (that's the way econometricians like to write it). If the limit $\lim_{t\to\infty} a_t = a$ exists a.s. (where a need not be the REE α, but may be any finite random variable, see, for instance, Theorem 3.2), then the matrix M_T tends a.s. to the singular matrix

$$\overline{M} = \begin{pmatrix} a^2 & a \\ a & 1 \end{pmatrix}. \tag{4.3}$$

This condition is generally referred to as absence of strict, or strong, asymptotic identification, see e.g. Davidson and MacKinnon (1993) or Newey and McFadden (1994). Strict asymptotic identification is, however, only sufficient for the asymptotic identification of the parameters and not necessary, with the consequence that the non-singularity of $\overline{M}$ need not preclude the OLS estimator from being consistent. This is indeed a point that seems to be sometimes overlooked in the literature. What we will show in this section is that the structural parameters in (4.1a) remain strongly consistently estimable, and hence asymptotically identified, even when $\lim_{t\to\infty} = a$ a.s..

In particular, in Section 4.1, we shall first provide an overview of results regarding consistent estimation in linear regression models obtained in the statistical literature over the past three decades, mainly by Lai, Wei and Robbins. For the case of deterministic regressors, these include a sufficient condition for strong consistency of the slope estimator in the simple regression model which is minimal in the sense of also being necessary for weak consistency, see Lai and Robbins (1977). For stochastic regressors, even the weakest sufficient conditions seem not to be met in our scenario (cf. Appendix B.6 in Christopeit and Massmann (2010b)). As an intermediate step, we therefore proceed to proving that the minimal condition in Lai and Robbins (1977) remains sufficient in the presence of stochastic regressors for a certain class of models. In addition, we show that a sharpened version of this condition is also necessary.

Subsequently, in Section 4.2, we turn to our model of interest. Still, due to the complexity of the regressor a_{t-1} in (4.1) the newly-derived result cannot be applied directly. We therefore use a rather special splitting technique tailored to

our specific model to show at first weak consistency of our OLS estimator. For a certain subclass, finally, strong consistency can be obtained following the lines of resoning in the proof of Theorem 4.1.

4.1 The simple linear regression model revisited

Since the problem is of some interest in itself, we shall formulate it in standard textbook notation. We will revert to the notation used so far in this paper in Section 4.2 below.

The so called simple linear regression model is of the form

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, 2, \ldots \tag{4.4}$$

The ordinary least squares (OLS) estimators a_n, b_n for α, β on the basis of the first n observations are given by

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix} = (X_n' X_n)^{-1} X_n' y_{(n)} \tag{4.5}$$

(provided $X_n' X_n$ is regular), where

$$X_n' = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix}, \; y_{(n)}' = (y_1, \ldots, y_n)'.$$

Henceforth, denote

$$M_n = X_n' X_n = n \begin{pmatrix} 1 & \overline{x}_n \\ \overline{x}_n & \overline{x_n^2} \end{pmatrix}.$$

As usual, we use the notation $\overline{x}_n = n^{-1} \sum_{i=1}^n x_i$. If the focus lies on the estimation of the slope β, then the formula

$$b_n = \frac{\sum_{i=1}^n (x_i - \overline{x}_n) y_i}{\sum_{i=1}^n (x_i - \overline{x}_n)^2} \tag{4.6}$$

will turn out more useful for theoretical considerations.

In the sequel, we shall review the basic results concerning consistency of the OLS estimator. Unless stated otherwise, maintained assumption 1 will assumed to be valid.

4.1.1 Deterministic regressors

The following results concerning *joint estimation* of the parameters actually hold true for multivariate regression models with an $n \times p$ deterministic regressor matrix X_n. A *sufficient* condition for L^2-convergence of (a_n, b_n) to (α, β) (L^2 - *consistency*) and hence convergence in probability (*weak consistency*) is

$$M_n^{-1} \to 0, \tag{4.7}$$

or, equivalently, the minimal eigenvalue $\lambda_{\min}(M_n) \to \infty$; see, for instance, Amemiya (1985). Sufficiency follows trivially from the fact that $\sigma^2 M_n^{-1}$ is the covariance matrix of $(a_n, b_n)'$ and, incidentally, remains true if the ε_i are merely uncorrelated. In the Gauss-Markov model, condition (4.7) is also necessary for weak convergence (cf. Eicker (1963) and Drygas (1976)). Needless to say that (4.7) is much weaker than the classical textbook condition

$$\frac{1}{n} M_n \to M, \tag{4.8}$$

where M is some positive definite matrix.

The question whether (4.7) implies *almost sure* convergence of (a_n, b_n) to (α, β) (*strong consistency*) is much more delicate. For normally distributed ε_i, Anderson and Taylor (1976) have shown that (4.7) is indeed necessary and sufficient. Without the normality assumption, additional restrictions like

$$M_n^{-1} \mathrm{diag}\left(d_{n1}^2, \ldots, d_{np}^2\right) = O(1),$$

where $d_{nk}^2 = \sum_{i=1}^n x_{ik}^2$ is the sum of squares of the k - th regressor, have been imposed (cf. Drygas (1976)). This assumption is much stronger than (4.7). For the simple regression model (4.4) (where $x_{i1} = 1, x_{i2} = x_i$), it amounts to requiring, e.g., that

$$\frac{\overline{x}_n \overline{x_n^2}}{\overline{x_n^2} - \left(\overline{x}_n\right)^2} = O(1),$$

which is not even fulfilled for polynomial regressors $x_i = i^m$, which do nevertheless satisfy (4.7). The ultimate affirmative answer was given in Lai, Robbins, and Wei (1978), who prove that (4.7) is indeed sufficient for strong consistency. Actually, all that is required of the error terms ε_i is that they are independent with $\mathbf{E}\varepsilon_i = 0$ and $\sup_i \mathbf{E}\varepsilon_i^2 < \infty$. In Lai, Robbins, and Wei (1979), this result was extended to a much more general class of ε_i.

If one is only interested in the estimation of the slope β, then the complete answer is given in Lai and Robbins (1977). To state it, introduce

$$A_n = \sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2. \tag{4.9}$$

Then

$$b_n - \beta = \frac{\sum_{i=1}^n \left(x_i - \overline{x}_n\right) \varepsilon_i}{A_n}. \tag{4.10}$$

According to Lai and Robbins (1977), a sufficient condition for almost sure convergence $b_n \to \beta$ is

$$\lim_{n \to \infty} A_n = \infty. \tag{4.11a}$$

This condition is also necessary for convergence in probability. Actually, the necessity part remains valid if the error terms are wide sense white noise. Observe that $A_n = n\left[\overline{x_n^2} - \left(\overline{x}_n\right)^2\right]$ and therefore

$$M_n^{-1} = \frac{1}{A_n} \begin{pmatrix} \overline{x_n^2} & -\overline{x}_n \\ -\overline{x}_n & 1 \end{pmatrix}. \tag{4.12}$$

Hence (4.7) implies (4.11a). The converse will be true if, in addition to (4.11a),

$$\lim_{n\to\infty} \frac{\overline{x_n^2}}{A_n} = 0 \tag{4.11b}$$

holds. In this case, i.e. if both conditions (4.11) are satisfied, both coefficients may be consistently estimated by least squares, with the slope estimator b_n being *strongly consistent.* Condition (4.11b) is easily verified to be equivalent to

$$\lim_{n\to\infty} \frac{(\overline{x}_n)^2}{A_n} = 0, \tag{4.13}$$

so that the equivalence of (4.11) and (4.7) will hold whenever $\sup_n |\overline{x}_n| < \infty$.

Since even in some serious econometric textbooks the implications of these results do not seem to be fully appreciated, we shall consider a simple example for illustrative purposes.

Example. Consider model (4.4) with $x_i = i^{-1/2}$. Then, to use a frequently encountered phrase, each observation x_i provides less and less information about β, with the consequence that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} X_n' X_n &= \lim_{n\to\infty} \frac{1}{n} \begin{pmatrix} n & \sum_{i=1}^n \frac{1}{\sqrt{i}} \\ \sum_{i=1}^n \frac{1}{\sqrt{i}} & \sum_{i=1}^n \frac{1}{i} \end{pmatrix} \\
&= \lim_{n\to\infty} \frac{1}{n} \begin{pmatrix} n & 2\,(\sqrt{n}-1) \\ 2\,(\sqrt{n}-1) & \log n \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}$$

As a result, the sufficient condition for consistency (4.7) is not met. Nevertheless,

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^n \frac{1}{\sqrt{i}} \sim \frac{2}{\sqrt{n}} + o(1/n), \quad \overline{x_n^2} = \frac{1}{n}\sum_{i=1}^n \frac{1}{i} \sim \frac{\log n}{n} + o(1/n),$$

and therefore

$$A_n = n\left[\overline{x_n^2} - (\overline{x}_n)^2\right] = \log n + o(1).$$

Hence, the sufficient condition (4.11a) is well satisfied and the OLS estimator b_n is a strongly consistent estimator of β. Note that condition (4.11b) is also satisfied, so that the OLS estimator for the intercept α is also consistent. With consistency of the OLS estimators established, it is plain that α and β are also asymptotically identified.

4.1.2 Stochastic regressors

For stochastic regressors, the situation presents itself less clearcut. There seems to be no "minimal" condition like (4.7). The best available result at present is the one obtained in Lai and Wei (1982b). It states that for predetermined regressors and martingale difference errors, both with respect to some filtration $(\mathcal{F}_n)$ and

the latter satisfying in addition $\sup_n \mathbf{E}\left(\varepsilon_n^2|\mathcal{F}_{n-1}\right) < \infty$ a.s., a sufficient condition for strong convergence of the least squares estimator of α and β is the following:

$$\lambda_{\min}\left(M_n\right) \to \infty \quad \text{and} \quad \left[\log \lambda_{\max}\left(M_n\right)\right]^{1+\delta} = o(\lambda_{\min}\left(M_n\right)) \quad \text{a.s.} \tag{4.14}$$

for some $\delta > 0$. If $\sup_n \mathbf{E}\left(|\varepsilon_n|^\alpha |\mathcal{F}_{n-1}\right) < \infty$ for some $\alpha > 2$, then is suffices to require (4.14) for $\delta = 0$.

Lai and Wei give an example which shows that even a marginal violation of (4.14) (replacing the $o\left(\cdot\right)$ by an $O\left(\cdot\right)$) is destructive to consistency. With a good deal of more effort, one can obtain some refinements of the condition (4.14) (cf. Lai and Wei (1982a)). Applied to the simple regression model (4.4) with $\sup_n \mathbf{E}\left(|\varepsilon_n|^\alpha |\mathcal{F}_{n-1}\right) < \infty$ for some $\alpha > 2$, the following sharpened version of condition (4.11a) turns out to be sufficient for strong consistency of the *slope estimator* b_n :

$$\frac{A_n}{\log n} \to \infty \text{ a.s.} \tag{4.15}$$

If $\overline{\lim}_n \overline{x^2} \leq 1$, (4.15) is also sufficient for consistent OLS estimation of α. Otherwise, additional conditions have to be imposed. As is shown in Lai and Wei (1982a), (4.14) (with $\delta = 0$) implies the conditions of their Theorem 2.

As one main contribution of this paper, we shall extend the approach in Lai and Robbins (1977) mentioned in the previous Subsection to a certain class of models with stochastic regressors. In particular, we show that (4.11a) remains a minimal sufficient condition for this class. Introduce the processes

$$u_n = \sum_{i=1}^{n} \left(x_i - \overline{x}_n\right) \varepsilon_i, \ n \geq 1, \tag{4.16}$$

and

$$v_n = \varepsilon_n - \overline{\varepsilon}_{n-1}, \ n \geq 2, \tag{4.17}$$

as well as the filtration $\mathcal{F}_n = \mathcal{F}_n^v = \sigma\left(v_2, \ldots, v_n\right), n \geq 2$.

The following results are proved in the Appendix A.

Theorem 4.1 *Suppose that the ε_i are independent $N(0, \sigma^2)$-distributed random variables, and that the centered regressors $x_n - \overline{x}_{n-1}$ are F_{n-1}^v-measurable for $n \geq 3$. Assume that*

$$\lim_{n\to\infty} A_n = \infty \quad \text{a.s.}. \tag{4.18}$$

Then, for every $\delta > 0$,

$$\lim_{n\to\infty} \frac{u_n}{A_n^{1/2} \left[\log A_n\right]^{(1+\delta)/2}} = 0 \quad \text{a.s.} \tag{4.19}$$

In view of (4.10) and (4.16), an immediate consequence is the following result on the strong consistency of the OLS estimator of β.

Corollary 4.2 *Under the assumptions of Theorem 4.1,*

$$\frac{A_n^{1/2}}{\left[\log A_n\right]^{(1+\delta)/2}} \left(b_n - \beta\right) \to 0 \quad \text{a.s.}$$

for every $\delta > 0$. In particular, the OLS estimator for β is strongly consistent, i.e.

$$b_n \to \beta \quad \text{a.s..}$$

In contrast to the setting with deterministic regressors, condition (4.18) will generally not be necessary in the presence of stochastic regressors. However, a restricted version of the negation of (4.18) rules out consistency:

Proposition 4.3 *Under the assumptions of Theorem 4.1, the condition*

$$\mathbf{E}A_\infty < \infty$$

implies inconsistency of the OLS estimator for the slope coefficient.

4.2 Application to adaptive learning

We come back to the estimation of β in the ALM (4.1), reproduced here for convenience:

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t, \tag{4.20}$$
$$a_t = a_{t-1} + \gamma_t (y_t - a_{t-1}). \tag{4.21}$$

To embed the model in the framework of section 4.1, note that (4.20) is a special case of the regression model (4.4), with $x_t = a_{t-1}$ (returning to the notation t instead of i). Introduce $c_t = (1-\beta)\gamma_t$. Insert (4.20) into (4.21) to obtain

$$a_t = (1 - c_t) a_{t-1} + \gamma_t (\delta + \varepsilon_t). \tag{4.22}$$

By the results of section 3 (cf. Corollary 3.3), for gain sequences γ_t satisfying condition (ii) of Theorem 3.2, $a_t \to \alpha = \delta/(1-\beta)$, so that a.s. convergence to the REE obtains. As a consequence, the asymptotic moment matrix

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

is singular, so that standard textbook asymptotic theory for the OLS estimator in (4.20) does not work. One may try to apply the result quoted in section 4.1.2, i.e. try to verify the condition (4.14): The eigenvalues of M_T are $\lambda_{\min} = To(1)$ and $\lambda_{\max} = T(1 + \alpha^2 + o(1))$, so that

$$\frac{\log \lambda_{\max}}{\lambda_{\min}} = \frac{\log T + \log(1+\alpha^2) + o(1)}{To(1)}.$$

But the $o(1)$ - terms depend in a complicated way on the arithmetic means $\overline{a}_T$ and $\overline{a_T^2}$, so that is seems a rather hopeless task to decide whether $\log T/To(1)$ tends to zero a.s. or not.

Therefore, our approach to the EEP will be along the lines of the lines of Theorem 4.1, concentrating first on the estimation of the slope β. We will, however, not apply Corollary 4.2 directly. Rather, we will split the process u_T (cf. (4.16))

into various components and treat each of these separately. This way we avoid the rather special measurability condition imposed on the centered regressors, which will generally not be met. The same decomposition will be made for A_T (cf. (4.9)). It will finally turn out that, indeed, $\lim_{T\to\infty} A_T = \infty$ a.s. and that $\operatorname{plim}_{T\to\infty} u_T/A_T = 0$, so that, in view of (4.10) and (4.16), strong consistency of the OLS estimator holds.

Start with the general solution to (4.22), which is given by

$$a_t = a_0 \prod_{i=1}^{t} (1 - c_i) + \sum_{i=1}^{t} \phi_{ti} (\delta + \varepsilon_i),$$

with

$$\phi_{ti} = \begin{cases} \gamma_t, & i = t, \\ \gamma_i \prod_{j=i+1}^{t} (1 - c_j), & i = 1, \ldots, t-1. \end{cases}$$

Note that $\lim_{t\to\infty} \phi_{ti} = 0$ since the series $\sum_j c_j$ diverges. Henceforth, we shall assume that $a_0 = 0$, so that

$$a_t = \sum_{i=1}^{t} \phi_{ti} (\delta + \varepsilon_i). \tag{4.23}$$

Then

$$\overline{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t = \frac{1}{T} \sum_{t=1}^{T} a_{t-1} = \frac{1}{T} \sum_{t=1}^{T-1} a_i = \frac{T-1}{T} \overline{a}_{T-1},$$

and the OLS estimator is given by

$$b_T - \beta = \frac{u_T}{A_T}, \tag{4.24}$$

where

$$u_T = \sum_{t=1}^{T} (x_t - \overline{x}_T) \varepsilon_t = \sum_{t=2}^{T} \left(a_{t-1} - \frac{T-1}{T} \overline{a}_{T-1} \right) \varepsilon_t. \tag{4.25}$$

and

$$A_T = \sum_{t=1}^{T} (x_t - \overline{x}_T)^2 = \sum_{t=2}^{T} \left(a_{t-1} - \frac{T-1}{T} \overline{a}_{T-1} \right)^2. \tag{4.26}$$

We shall confine our analysis to the *special case* where the gain sequence is given by

$$\gamma_t = \frac{\gamma}{t} \tag{4.27}$$

with some positive constant γ. Then the c_t introduced above become $c_t = c/t$, with $c = (1 - \beta)\gamma$, and $1 - c_t = 1 - c/t$. Hence

$$\phi_{ti} = \frac{\gamma}{i} \left(1 - \frac{c}{i+1} \right) \cdots \left(1 - \frac{c}{t} \right).$$

Using a straightforward recursive representation of the ϕ_{ti} (in i for fixed t) one can show that, at least for noninteger c, the ϕ_{ti} have the following properties.

(i) For $c < 1$, they are strictly decreasing.

(ii) For $c > 1$, they are *alternating* for $i \leq [c-1]$ and constant in sign afterwards. Moreover, they are *decreasing in absolute value* up to $i = \left[\frac{c-1}{2}\right]$ and are *increasing* afterwards.
(iii) For $c = 1$, they are constant:

$$\phi_{ti} = \frac{\gamma}{i}\frac{i}{i+1}\frac{i+1}{i+2}\cdots\frac{t-1}{t} = \frac{\gamma}{t} = \frac{1}{(1-\beta)\,t}.$$

Denote $i_0 = \max\{i : i \leq c\}$. For $i \geq i_0$, take logarithms to obtain

$$\ln\phi_{ti} = \ln\gamma - \ln i + \sum_{j=i+1}^{t} \ln\left(1 - \frac{c}{j}\right) = \ln\gamma - \ln i + \sum_{j=i+1}^{t}\left[-\frac{c}{j} + \frac{d_j}{j^2}\right]$$

with

$$d_j = O(1).$$

Making use of the integral comparison test,

$$\begin{aligned}
\sum_{j=i+1}^{t}\frac{1}{j} &= \ln t - \ln i + O_{ti}(1)\frac{1}{i}, \\
\sum_{j=i+1}^{t}\frac{d_j}{j^2} &= O_{ti}(1)\sum_{j=i+1}^{t}\frac{1}{j^2} = O_{ti}(1)\left[\frac{1}{t} - \frac{1}{i} + \frac{1}{i^2}\right].
\end{aligned}$$

Here and in the sequel we shall denote by $O_{ti}(1)$ numbers R_{ti} for which $\sup_t \sup_{i\leq t} |R_{ti}| \leq K < \infty$ for some constant K. $O_i(1)$ and $O(1) = O_n(1)$ are to be understood in a similar way. As a consequence, for $i \geq i_0$,

$$\begin{aligned}
\ln\phi_{ti} &= \ln\gamma - \ln i + \sum_{j=i+1}^{t}\left[-\frac{c}{j} + \frac{d_j}{j^2}\right] \\
&= \ln\gamma - \ln i - c\left[\ln t - \ln i\right] + O_{ti}(1)\frac{1}{i} \\
&= \ln\gamma - (1-c)\ln i - c\ln t + O_{ti}(1)\frac{1}{i}.
\end{aligned}$$

Therefore, denoting the $O_{ti}(1)$-terms by a_{ti},

$$\phi_{ti} = \gamma\frac{1}{t^c}\frac{1}{i^{1-c}}e^{O_{ti}(1)/i} = \gamma\frac{1}{t^c}\frac{1}{i^{1-c}}\left[1 + \frac{a_{ti}}{i}\right]. \tag{4.28a}$$

By the integral comparison test, for $c \neq 1$,

$$\begin{aligned}
\sum\nolimits_{i=i_0}^{t}\phi_{ti} &= \gamma\frac{1}{t^c}\sum\nolimits_{i=i_0}^{t}\frac{1}{i^{1-c}} + \gamma\frac{1}{t^c}\sum\nolimits_{i=i_0}^{t}\frac{a_{ti}}{i^{2-c}} \\
&= \gamma\frac{1}{t^c}\left[\frac{1}{c}\left(t^c - i_0^c\right) + O(1)\left(t^{c-1}\vee 1\right)\right] \\
&\quad + \gamma\frac{O(1)}{t^c}\left[\frac{1}{c-1}\left(t^{c-1} - i_0^{c-1}\right) + t^{c-2}\vee 1\right] \\
&= \frac{1}{1-\beta} + O(1/t^c) + O(1/t)
\end{aligned}$$

since $\gamma/c = 1/(1-\beta)$. For $c = 1$,

$$\sum_{i=1}^{t} \phi_{ti} = \frac{1}{1-\beta}.$$

For $i < i_0$,

$$\phi_{ti} = \gamma_i \prod_{j=i+1}^{i_0} (1 - c_j) \prod_{j=i_0+1}^{t} (1 - c_j) = \lambda_i \phi_{ti_0} \tag{4.28b}$$

with

$$\lambda_i = \frac{\gamma_i}{\gamma_{i_0}} \prod_{j=i+1}^{i_0} (1 - c_j).$$

Hence, since $\max_{i<i_0} |\lambda_i| \leq K, \phi_{ti} = O(t^{-c})$ for $i < i_0$ and therefore

$$\sum_{i=1}^{t} \phi_{ti} = \frac{1}{1-\beta} + O(1/t^c) + O(1/t).$$

Coming back to (4.23) and remembering that $\alpha = \delta/(1-\beta)$, it follows from (4.28) that

$$a_t = \alpha + O(1/t^c) + O(1/t) + \sum_{i=1}^{i_0-1} \lambda_i \phi_{ti_0} \varepsilon_i + \gamma \frac{1}{t^c} \sum_{i=i_0}^{t} \frac{1}{i^{1-c}} \left[1 + \frac{a_{ti}}{i}\right] \varepsilon_i.$$

But

$$\begin{aligned} \sum_{i=1}^{i_0-1} \lambda_i \phi_{ti_0} \varepsilon_i &= O\left(t^{-c}\right), \\ \frac{1}{t^c} \sum_{i=i_0}^{t} \frac{1}{i^{1-c}} \left[1 + \frac{a_{ti}}{i}\right] \varepsilon_i &= \frac{1}{t^c} \sum_{i=1}^{t} \frac{1}{i^{1-c}} \left[1 + \frac{a_{ti}}{i}\right] \varepsilon_i + O\left(t^{-c}\right) \end{aligned}$$

with probability one. Hence

$$\begin{aligned} a_t &= \alpha + \frac{\gamma}{t^c} \sum_{i=1}^{t} \frac{1}{i^{1-c}} \varepsilon_i + \frac{\gamma}{t^c} \sum_{i=1}^{t} \frac{a_{ti}}{i^{2-c}} \varepsilon_i + O(1) \left(\frac{1}{t^c} + \frac{1}{t}\right) \\ &= \alpha + \frac{\gamma}{t^c} s_t + \frac{\gamma}{t^c} r_t + O(1) \left(\frac{1}{t^c} + \frac{1}{t}\right) \\ &= \alpha + \xi_t + \eta_t + \zeta_t \end{aligned} \tag{4.29}$$

with

$$s_t = \sum_{i=1}^{t} \frac{1}{i^{1-c}} \varepsilon_i, \quad r_t = \sum_{i=1}^{t} \frac{a_{ti}}{i^{2-c}} \varepsilon_i \tag{4.30}$$

and

$$\begin{aligned} \xi_t &= \gamma \frac{1}{t^c} s_t, \\ \eta_t &= \gamma \frac{1}{t^c} r_t, \\ \zeta_t &= O(1) \left(\frac{1}{t^c} + \frac{1}{t}\right). \end{aligned} \tag{4.31}$$

Accordingly,

$$\begin{aligned} x_t - \overline{x}_T &= a_{t-1} - \frac{T-1}{T}\overline{a}_{T-1} \\ &= \frac{\alpha}{T} + \left[\xi_{t-1} - \frac{T-1}{T}\overline{\xi}_{T-1}\right] + \left[\eta_{t-1} - \frac{T-1}{T}\overline{\eta}_{T-1}\right] \\ &\qquad + \left[\zeta_{t-1} - \frac{T-1}{T}\overline{\zeta}_{T-1}\right] \end{aligned} \tag{4.32}$$

Based on this decomposition, the following theorem is proved in Appendices B.1 to B.4.

Theorem 4.4 *Assume model (4.20) with adaptive learning scheme (4.21) and decreasing gain sequence given by (4.27). Then, for $c > 1/2$, the least squares estimator for β is weakly consistent.*

Actually, then, also the OLS estimator for δ is consistent. This follows immediately from the representation

$$\widehat{\delta}_T = \delta - (b_T - \beta)\,\overline{a}_{T-1} + \overline{\varepsilon}_T.$$

For *Gaussian* error terms, one may even obtain strong consistency.

Theorem 4.5 *In addition to the assumptions made in Theorem 4.4, assume that the ε_t are Gaussian. Then, for $c > 1$, the least squares estimator for β is strongly consistent.*

The proof is given in Appendix B.5.

5 Summary and outlook

The main point we make in this paper is that the OLS estimators of α and β in the model (4.20) with the regressor a_{t-1} being formed via the adaptive learning algorithm (4.21) and the gain sequence specified as in (4.27) are consistent as long as $\gamma > 1/[2(1-\beta)] > 0$, see Theorem 4.4. Put differently, even though agents' expectations are formed via a decreasing gain learning recursion, and thus $a_t \to a$ a.s., see Theorem 3.2, the structural parameters in the model remain consistently estimable. The result may appear counterintuitive at first sight since the regressors 1 and a_{t-1} are asymptotically collinear. Yet note that this so-called strict asymptotic identification condition is not necessary for consistency.

Consistency being settled, the next step would be to investigate *asymptotic normality*. This problem is currently under investigation in a companion paper, cf. Christopeit and Massmann (2010a). Appendix B.6 in Christopeit and Massmann (2010b) provides a brief sketch of the procedure and of the results that are to be expected. The bottom line is that, in the setting of Theorem 4.4 and under the additional assumption that the ε_t possess finite fourth moments, the least squares estimator of β will be asymptotically normal at rate $\sqrt{\ln T}$ (cf. (B.41)). Moreover, the asymptotic variance of $\sqrt{\ln T}\,(b_T - \beta)$ tends to zero when γ becomes large or when $\gamma \searrow 1/[2(1-\beta)]$, indicating that the convergence of the OLS estimator

to the true parameter value β gets faster, see the discussion following (B.41) in Appendix B.6 of Christopeit and Massmann (2010b).

The question of consistency and asymptotic normality in a more general class of models is left for future research.

A Proof of Theorem 4.1

A.1 Some basic conditional expectations

Maintained assumption: The ε_i are *independent Gaussian* with $\mathbf{E}\varepsilon_i = 0$ and $\mathbf{E}\varepsilon_i^2 = \sigma^2$.

Define

$$\begin{aligned} v_n &= \varepsilon_n - \overline{\varepsilon}_{n-1}, \\ e_{ni} &= \varepsilon_n - \varepsilon_i \end{aligned}$$

for $n \geq 2, i = 1, \ldots, n-1$. As usual, let

$$\overline{z}_n = \frac{1}{n}\sum_{i=1}^{n} z_i.$$

Consider the filtrations $(\mathcal{F}_n^v)_{n\geq 2}$ and $(\mathcal{F}_n^e)_{n\geq 2}$, where $\mathcal{F}_n^v = \sigma(v_2, \ldots, v_n)$ and $\mathcal{F}_n^e = \sigma(e_{n1}, \ldots, e_{n,n-1})$. Note that $\mathcal{F}_n^v \subset \mathcal{F}_n^\varepsilon = \sigma(\varepsilon_1, \ldots, \varepsilon_n)$ and $\mathcal{F}_n^e \subset \mathcal{F}_n^\varepsilon$, the inclusions being proper.

Lemma A.1 *(i)* $\mathcal{F}_n^v = \mathcal{F}_n^e$.
(ii) The $v_n, n \geq 2$, *are independent Gaussian with variance* $\mathbf{E}v_n^2 = \frac{n}{n-1}\sigma^2$.

Proof. *(i)* Denote, for $n \geq 2$,

$$\underset{(n-1)\times 1}{v_{(n)}} = \begin{pmatrix} v_2 \\ v_3 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \varepsilon_2 - \overline{\varepsilon}_1 \\ \varepsilon_3 - \overline{\varepsilon}_2 \\ \vdots \\ \varepsilon_n - \overline{\varepsilon}_{n-1} \end{pmatrix},$$

$$\underset{(n-1)\times 1}{e_{(n)}} = \begin{pmatrix} e_{n1} \\ e_{n2} \\ \vdots \\ e_{n,n-1} \end{pmatrix} = \begin{pmatrix} \varepsilon_n - \varepsilon_1 \\ \varepsilon_n - \varepsilon_2 \\ \vdots \\ \varepsilon_n - \varepsilon_{n-1} \end{pmatrix}.$$

By simple calculation,

$$v_{(n)} = C_n e_{(n)}$$

with

$$\underset{(n-1)\times(n-1)}{C_n} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 & 0 & \cdots & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -1 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ \frac{1}{n-2} & \frac{1}{n-2} & \frac{1}{n-2} & \cdots & \frac{1}{n-2} & -1 \\ \frac{1}{n-1} & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & \frac{1}{n-1} & \frac{1}{n-1} \end{pmatrix}.$$

Since C_n is regular,

$$\mathcal{F}_n^v = \sigma\left(v_{(n)}\right) = \sigma\left(e_{(n)}\right) = \mathcal{F}_n^e.$$

(ii) By direct calculation,

$$v_{(n)} = (C_n\, e_{n-1})\, \varepsilon_{(n)},$$

where $\varepsilon_{(n)} = (\varepsilon_1, \ldots, \varepsilon_n)'$ and e_{n-1} denotes the $(n-1)-$ th unit vector in dimension $n-1$. Since

$$(C_n\, e_{n-1})\,(C_n\, e_{n-1})' = \text{ diag}\left[2, \frac{3}{2}, \ldots, \frac{n}{n-1}\right],$$

the assertion follows. ■

Henceforth, denote $\mathcal{F}_n = \mathcal{F}_n^v = \mathcal{F}_n^e$.

We calculate the *joint distribution* of the random $(2n-1)$ - vector $u = u_n = (\varepsilon_1, \ldots, \varepsilon_n, \varepsilon_n - \varepsilon_1, \ldots, \varepsilon_n - \varepsilon_{n-1})'$. Clearly u is Gaussian with mean zero and covariance

$$\Sigma = \sigma^2 \begin{pmatrix} I_n & \Lambda \\ \Lambda' & \Gamma \end{pmatrix}.$$

I_n is the $n-$ dimensional unit matrix, and

$$\begin{aligned} \Gamma &= I_{n-1} + \iota\iota', \\ \underset{n\times(n-1)}{\Lambda} &= \begin{pmatrix} -I_{n-1} \\ \iota' \end{pmatrix} \end{aligned}$$

(with $\iota' = \iota'_{n-1} = (1, \ldots, 1)$ in dimension $n-1$). Below we shall need the inverse

$$\Gamma^{-1} = I_{n-1} - \frac{1}{n}\iota\iota'.$$

By the theorem on normal correlation, the *conditional distribution* of $\varepsilon_{(n)} = (\varepsilon_1, \ldots, \varepsilon_n)'$ given $e_{(n)} = (\varepsilon_n - \varepsilon_1, \ldots, \varepsilon_n - \varepsilon_{n-1})'$, is normal with mean

$$\begin{aligned} m_n &= \Lambda\Gamma^{-1} e_{(n)} = \left[\begin{pmatrix} -I_{n-1} \\ \iota' \end{pmatrix} - \frac{1}{n}\begin{pmatrix} -I_{n-1} \\ \iota' \end{pmatrix}\iota\iota'\right] e_{(n)} \\ &= \begin{pmatrix} -\left(I_{n-1} - \frac{1}{n}\iota\iota'\right) \\ \frac{1}{n}\iota' \end{pmatrix} e_{(n)} \end{aligned} \tag{A.1}$$

and covariance $\sigma^2 V_n$, where

$$\begin{aligned} V_n &= I_n - \Lambda\Gamma^{-1}\Lambda' \\ &= I_n - \left[\begin{pmatrix} -I_{n-1} \\ \iota' \end{pmatrix} - \frac{1}{n}\begin{pmatrix} -I_{n-1} \\ \iota' \end{pmatrix}\iota\iota'\right]\begin{pmatrix} -I_{n-1} & \iota \end{pmatrix} \\ &= I_n - \left[\begin{pmatrix} I_{n-1} & -\iota \\ -\iota' & n-1 \end{pmatrix} - \frac{1}{n}\begin{pmatrix} \iota\iota' & -(n-1)\,\iota \\ -(n-1)\,\iota' & (n-1)^2 \end{pmatrix}\right] \\ &= \begin{pmatrix} 0_{n-1} & \iota \\ \iota' & -(n-2) \end{pmatrix} + \frac{1}{n}\begin{pmatrix} \iota\iota' & -(n-1)\,\iota \\ -(n-1)\,\iota' & (n-1)^2 \end{pmatrix} \\ &= \frac{1}{n}\begin{pmatrix} \iota\iota' & \iota \\ \iota' & 1 \end{pmatrix}. \end{aligned}$$

Lemma A.2 *The conditional expectation of ε_i given $\mathcal{F}_{n-1}$ is*

$$\begin{aligned}\mathbf{E}\left(\varepsilon_i|\mathcal{F}_{n-1}\right) &= \begin{cases} \frac{1}{n-1}\sum_{k=1}^{n-2}\left(\varepsilon_{n-1}-\varepsilon_k\right)-\left(\varepsilon_{n-1}-\varepsilon_i\right), & 1\leq i\leq n-2, \\ \\ \frac{1}{n-1}\sum_{k=1}^{n-2}\left(\varepsilon_{n-1}-\varepsilon_k\right), & i=n-1. \end{cases} \\ &= \begin{cases} \frac{n-2}{n-1}\left(\varepsilon_{n-1}-\overline{\varepsilon}_{n-2}\right)-\left(\varepsilon_{n-1}-\varepsilon_i\right), & 1\leq i\leq n-2, \\ \\ \frac{n-2}{n-1}\left(\varepsilon_{n-1}-\overline{\varepsilon}_{n-2}\right), & i=n-1. \end{cases} \end{aligned} \tag{A.2}$$

Proof. (A.2) follows by direct calculation of $m_{n-1}=\mathbf{E}\left(\varepsilon_{(n-1)}|\mathcal{F}_{n-2}\right)$ from (A.1). ■

A.2 A sufficient condition

Remark. As a consequence of Lemma A.1**,** $(v_n)_{n\geq 2}$ is a *martingale difference sequence (MDS)* $(\mathcal{F}_n)_{n\geq 2}$. However**,** (v_n) is *not* a MDS w.r.t. the filtration $(\mathcal{F}_n^{\varepsilon})$. This is immediate from

$$\mathbf{E}\left(v_n|\mathcal{F}_{n-1}^{\varepsilon}\right)=\mathbf{E}\left\{\varepsilon_n-\frac{1}{n-1}S_{n-1}|\mathcal{F}_{n-1}^{\varepsilon}\right\}=-\frac{1}{n-1}S_{n-1}.$$

As in section 4, denote

$$u_n=\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)\varepsilon_i, \tag{A.3}$$

and introduce

$$w_n=u_n-u_{n-1}.$$

Lemma A.3 *The following identity is true:*

$$w_n=\frac{n-1}{n}\left(x_n-\overline{x}_{n-1}\right)\left(\varepsilon_n-\overline{\varepsilon}_{n-1}\right). \tag{A.4}$$

If the sequence $(x_n-\overline{x}_{n-1})$ is predictable w.r.t. $(\mathcal{F}_n)$, then $(u_n)_{n\geq 1}$ is a martingale transform (local martingale) w.r.t. $(\mathcal{F}_n)$.

Proof. (A.4) follows from

$$\begin{aligned} u_n-u_{n-1} &= \sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)\varepsilon_i-\sum_{i=1}^{n-1}\left(x_i-\overline{x}_{n-1}\right)\varepsilon_i \\ &= x_n\varepsilon_n-\sum_{i=1}^{n-1}\left(\overline{x}_n-\overline{x}_{n-1}\right)\varepsilon_i-\overline{x}_n\varepsilon_n \\ &= x_n\varepsilon_n-\frac{n-1}{n}\left(x_n-\overline{x}_{n-1}\right)\overline{\varepsilon}_{n-1}-\overline{x}_n\varepsilon_n \\ &= -\frac{n-1}{n}\left(x_n-\overline{x}_{n-1}\right)\overline{\varepsilon}_{n-1}-\left(\overline{x}_n-x_n\right)\varepsilon_n \\ &= -\frac{n-1}{n}\left(x_n-\overline{x}_{n-1}\right)\overline{\varepsilon}_{n-1}-\frac{n-1}{n}\left(\overline{x}_{n-1}-x_n\right)\varepsilon_n \\ &= w_n. \end{aligned}$$

Here we have made use of the facts that

$$\begin{aligned} \overline{x}_n - \overline{x}_{n-1} &= \frac{x_n - \overline{x}_{n-1}}{n}, \\ \overline{x}_n - x_n &= \frac{n-1}{n}\left(\overline{x}_{n-1} - x_n\right). \end{aligned}$$

Noting that $u_1 = 0$, we have the representation

$$u_n = \sum_{i=2}^{n} w_i = \sum_{i=2}^{n} \frac{i-1}{i}\left(x_i - \overline{x}_{i-1}\right) v_i. \tag{A.5}$$

Hence, if $x_n - \overline{x}_{n-1}$ is $\mathcal{F}_{n-1}-$ measurable, (u_n) is a martingale transform by virtue of Lemma A.1. ■

Introduce

$$\widetilde{v}_n = \sqrt{\frac{n-1}{n}} v_n$$

and

$$z_n = x_n - \overline{x}_{n-1}, \quad \widetilde{z}_n = \sqrt{\frac{n-1}{n}} z_n.$$

Then the $\widetilde{v}_n, n \geq 2$, are iid $\mathcal{N}(0, \sigma^2)$, and (u_n) may be written as martingale transform

$$u_n = \sum_{i=2}^{n} \widetilde{z}_i \widetilde{v}_i.$$

Making use of the algebraic identity

$$\sum_{i=1}^{n} \left(x_i - \overline{x}_n\right)^2 = \sum_{i=2}^{n} \frac{i-1}{i}\left(x_i - \overline{x}_{i-1}\right)^2$$

(cf. Lai and Robbins (1977), (11)), we find that

$$A_n = \sum_{i=2}^{n} \widetilde{z}_i^2 = \sum_{i=2}^{n} \frac{i-1}{i}\left(x_i - \overline{x}_{i-1}\right)^2. \tag{A.6}$$

In particular, this shows that A_n is increasing to some limit A_∞. By well known convergence theorems for martingale transforms (cf. Lai and Wei (1982b), Lemma 2), we have the following result.

Lemma A.4 *Under the assumptions of Lemma A.3,*

$$\frac{u_n}{A_n^{1/2}\left[\log A_n\right]^{(1+\delta)/2}} \to 0 \text{ a.s. on } \{A_\infty = \infty\}$$

for any $\delta > 0$.

In particular, since $b_n - \beta = u_n / A_n$, the condition

$$\mathbf{P}\left(A_\infty = \infty\right) = 1 \tag{A.7}$$

is *sufficient* for strong consistency of the OLS estimator.

A.3 A necessary condition

Consider now the case where A_n converges to some finite random variable A_∞ on a set Ω_0 with $\mathbf{P}(\Omega_0) > 0$. By a standard convergence theorem for local martingales (cf. Lai and Wei (1982b), Lemma 2 (iii)), u_n converges on Ω_0 to some finite random variable. Denote this limit by u_∞ (putting $u_\infty = 0$ outside Ω_0). We would like to show that

$$\mathbf{P}(u_\infty \neq 0) > 0. \tag{A.8}$$

In the case of *deterministic regressors,* the proof of (A.8) runs as follows (cf. Lai and Robbins (1977)). $\Omega_0 = \Omega$, and $\mathbf{E}u_n^2 = \sigma^2 A_n$. Therefore (cf. Doob (1953), Theorem 7.2, p. 165) convergence $u_n \to u_\infty$ takes place in L^2. Consequently, $\mathbf{E}u_\infty = 0$ and $\mathbf{E}u_\infty^2 = \lim_{n\to\infty} \mathbf{E}u_n^2 = \sigma^2 A_\infty > 0$ (unless all x_i are identical). As a consequence, $b_n - \beta = u_n/A_n$ converges in probability to some random variable $z = u_\infty/A_\infty$ with $\mathbf{E}z = 0$ and $\mathbf{E}z^2 > 0$. Hence the OLS estimator b cannot be consistent. In other words, for deterministic regressors condition (A.7) is also *necessary* for strong consistency of the OLS estimator.

For stochastic regressors, we shall consider only a very special case of violation of (A.7), namely the case where

$$\mathbf{E}A_\infty < \infty. \tag{A.9}$$

Obviously, in this case, $\Omega_0 = \Omega$ modulo a nullset.

Lemma A.5 *Suppose that (A.9) holds. Then u is a martingale bounded in L^2 (i.e. $\sup_n Eu_n^2 < \infty$), and $M = u - \langle u \rangle$ is a uniformly integrable (u.i.) martingale.*

Proof. Assume $\sigma^2 = 1$ w.r.o.g. For $N = 1, 2, \ldots$, consider the stopping times

$$\tau_N = \inf\{n \geq 1 : A_{n+1} > N\}.$$

Since A is predictable, the sets $\{\tau_N < i\} = \{A_1 > N\} \cup \cdots \cup \{A_i > N\}$ and hence also $\{\tau_N \geq i\}$ are in $\mathcal{F}_{i-1}$. For each N, the stopped process $u^N = (u_{n\wedge\tau_N})$, given by

$$u_{n\wedge\tau_N} = \sum_{i=1}^{n\wedge\tau_N} \widetilde{z}_i \widetilde{v}_i$$

is then a square integrable martingale with predictable quadratic variation

$$\langle u^N \rangle_n = \sum_{i=1}^{n} \widetilde{z}_i^2 1_{\{i\leq\tau_N\}} \mathbf{E}\left(\widetilde{v}_i^2 | \mathcal{F}_{i-1}\right) = \sum_{i=1}^{n} \frac{i-1}{i} z_i^2 1_{\{i\leq\tau_N\}} \leq A_n \tag{A.10}$$

(cf. (A.6)). Hence

$$\mathbf{E}\left(u_n^N\right)^2 = \mathbf{E}\langle u^N \rangle_n \leq \mathbf{E}A_\infty.$$

Since $\tau_N \nearrow \infty$ a.s., it follows from Fatou's lemma that

$$\mathbf{E}u_n^2 = \mathbf{E}\left(\lim_{N\to\infty}\left(u_n^N\right)^2\right) \leq \underline{\lim}_{N\to\infty} \mathbf{E}\left(u_n^N\right)^2 \leq \mathbf{E}A_\infty.$$

This shows that (u_n) is bounded in L^2. By Doob's inequality, denoting $u_n^* = \sup_{k\leq n} |u_k|$,

$$\mathbf{E}\,|u_n^*|^2 \leq 4\mathbf{E}u_n^2 \leq 4\mathbf{E}A_\infty.$$

Again by Fatou's lemma, $u^* = \sup_n |u_n|$ is also in L^2. As a consequence, the u_n are uniformly integrable, so that we may pass to the limit $N \to \infty$ in the martingale equality

$$\mathbf{E}\left\{u_{n\wedge\tau_N}|\mathcal{F}_{n-1}\right\} = u_{(n-1)\wedge\tau_N}$$

to obtain

$$\mathbf{E}\left\{u_n|\mathcal{F}_{n-1}\right\} = u_{n-1}.$$

I.e., u is a square integrable martingale. Finally, observe that

$$\sup_n \left|u_n^2 - \langle u\rangle_n\right| \leq (u^*)^2 + A_\infty.$$

Since the rhs is integrable, this shows that $M = u^2 - \langle u\rangle$ is u.i. and hence (as u.i. local martingale) a martingale. ■

Corollary A.6 *If (A.13) holds, u_n converges a.s. and in L^1 to some random variable u_∞ for which $\mathbf{E}u_\infty = 0$ and*

$$\mathbf{E}u_\infty^2 = \sigma^2\mathbf{E}A_\infty. \tag{A.11}$$

Proof. Since u is bounded in L^2, it is u.i.. It then follows from the fundamental convergence theorem for u.i. martingales that u_n converges a.s. and in L^1 to to some random variable u_∞. By the same theorem, the u.i. martingale M converges to some limit M_∞ a.s. and in L^1, and $\overline{M} = (M_n)_{1\leq n\leq\infty}$ is also a martingale. In particular, this means that

$$\mathbf{E}M_\infty = \lim_{n\to\infty}\mathbf{E}M_n = 0.$$

The predictable quadratic variation of u is

$$\langle u\rangle_n = \sum_{i=1}^n \widetilde{z}_i^2\mathbf{E}\left(\widetilde{v}_i^2|\mathcal{F}_{i-1}\right) = \sigma^2\sum_{i=1}^n \frac{i-1}{i}z_i^2 = \sigma^2 A_n.$$

Since $u_n^2 = M_n + \langle u\rangle_n$, it follows that

$$u_\infty^2 = M_\infty + \sigma^2 A_\infty,$$

from which (A.11) follows. ■

Consequence: Since $\mathbf{E}A_\infty > 0$ (unless all x_i are identical), this shows that (A.8) holds. A forteriori, the OLS estimator cannot be consistent.

B Proof of Theorems 4.4 and 4.5

In view of (4.24), for Theorem 4.4 we have to show that

$$\frac{u_T}{A_T} \xrightarrow{P} 0. \tag{B.0}$$

According to (4.25) and (4.30)-(4.31), we have the following decomposition of u_t :

$$u_t = \frac{\alpha}{t}\sum_{i=2}^{t}\varepsilon_t + u_{1t} + u_{2t} + u_{3t}, \tag{B.1}$$

where

$$u_{1t} = \sum_{i=2}^{t}\left(\xi_{i-1} - \frac{t-1}{t}\overline{\xi}_{t-1}\right)\varepsilon_i, \tag{B.2a}$$

$$u_{2t} = \sum_{i=2}^{t}\left(\eta_{i-1} - \frac{t-1}{t}\overline{\eta}_{t-1}\right)\varepsilon_i, \tag{B.2b}$$

$$u_{3t} = \sum_{i=2}^{t}\left(\zeta_{i-1} - \frac{t-1}{t}\overline{\zeta}_{t-1}\right)\varepsilon_i. \tag{B.2c}$$

Accordingly, we introduce the processes

$$A_T^1 = \sum_{t=1}^{T}\left(\xi_{t-1} - \frac{T-1}{T}\overline{\xi}_{T-1}\right)^2 \tag{B.3}$$

and similarly A_T^2 and A_T^3 (where ξ is replaced by η or ζ). In view of the alternative representation (A.6), the processes $A_T, A_T^\nu, \nu = 1, 2, 3$, are monotone increasing.

We will show the following:

Appendix B.1: u_{3t} is a.s. bounded, and u_{2t} is bounded in probability.

Appendix B.2: u_{1t} is of the form

$$u_{1t} = \sum_{i=2}^{t}\xi_{i-1}\varepsilon_i + O_P(1). \tag{B.4}$$

Appendix B.3: The limits A_∞^2 and A_∞^3 are a.s. finite.

Appendix B.4: A_T^1 is of the form

$$A_T^1 = [1 + o_P(1)]\sum_{t=1}^{T}\xi_t^2, \tag{B.5a}$$

and

$$\sum_{t=1}^{\infty}\xi_t^2 = \infty \quad \text{a.s.} \tag{B.5b}$$

Hence A_T^1 tends to ∞ in probability. By monotonicity, this is equivalent to a.s. convergence to ∞, i.e. $A_\infty^1 = \infty$ a.s..

As a consequence of the local martingale convergence theorem cited in Appendix A, it will then follow from (B.5b) that

$$U_T = \frac{\sum_{i=2}^{T} \xi_{i-1}\varepsilon_i}{\sum_{i=1}^{T} \xi_i^2} \to 0 \quad \text{a.s.}$$

and further, taking account of (B.4) and (B.5a), that

$$\frac{u_{1T}}{A_T^1} = \left[U_T + \frac{O_P(1)}{\sum_{i=1}^{T} \xi_i^2} \right] [1 + o_P(1)] \xrightarrow{P} 0. \tag{B.6}$$

Moreover, in view of the decomposition (4.32), A_T is of the form

$$A_T = \sum_{t=1}^{T} (y_{Tt} + z_{Tt})^2 , \tag{B.7a}$$

where

$$\sum_{t=1}^{T} y_{Tt}^2 = A_T^1 \quad \text{and} \quad B_T = \sum_{t=1}^{T} z_{Tt}^2 \leq C \left[\frac{1}{T} + A_T^2 + A_T^3 \right]. \tag{B.7b}$$

It is then easily shown that $A_\infty^1 = \infty$ a.s. together with the result of Appendix B.3 implies that

$$A_T = [1 + o(1)]\, A_T^1 \quad \text{a.s.}. \tag{B.8}$$

Therefore, synthesizing Appendix B.1 as well as (B.6) and (B.8),

$$\frac{u_{1T} + u_{2T} + u_{3T}}{A_T} \xrightarrow{P} 0.$$

Finally, taking account of (B.1) we obtain the desired conclusion (B.0).

In **Appendix B.5,** we will show strong consistency. The proof is based on the results obtained in Appendices B.1 - B.4 as well as on Appendix A.

B.1 The terms u_2 and u_3

Ad u_3
By (4.31),

$$\zeta_t = O(1) \left(\frac{1}{t^c} + \frac{1}{t} \right).$$

Therefore

$$\overline{\zeta}_t \;=\; O(1) \frac{1}{t} \sum_{i=1}^{t} \left(\frac{1}{i^c} + \frac{1}{i} \right) = O(1) \frac{1}{t} \left(1 + \frac{1}{t^{c-1}} + \ln t \right) = O(1) \left(\frac{1}{t^c} + \frac{\ln t}{t} \right), \tag{B.9}$$

so that

$$\overline{\zeta}_{t-1} \sum_{t=2}^{t} \varepsilon_i = o(1) \text{ a.s.}$$

for $c > 1/2$ (by Kolmogorov's strong LLN). Also, $\sum_{t=2}^{t} \zeta_{i-1}\varepsilon_i$ converges a.s. to some finite random variable since it is a martingale difference sequence with

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

Hence we have the following

Intermediate result 1. For $c > 1/2$, u_{3t} converges a.s. to some finite random variable u_3.

Ad u_2

Consider the decomposition

$$u_{2t} = \sum_{k=2}^{t} \eta_{k-1}\varepsilon_k - \frac{t-1}{t}\overline{\eta}_{t-1} \sum_{k=2}^{t} \varepsilon_k = v_{2t} - w_{2t}. \tag{B.10}$$

By definition of η_t (cf. 4.31), $\eta_t = \gamma\sigma_t$ with

$$\sigma_k = \frac{1}{k^c} r_k = \frac{1}{k^c} \sum\nolimits_{i=1}^{k} \frac{a_{ki}}{i^{2-c}} \varepsilon_i.$$

Note that the coefficients under the sum depend on i and k, so that actually we have a martingale array, to which standard martingale convergence theorems are not directly applicable. Note, however, that – resulting from an expansion of the ϕ_{ki} (cf. (4.28a)) – the a_{ki} are deterministic and uniformly bounded in i, k. By the integral comparison test,

$$\sum\nolimits_{i=1}^{k} \frac{1}{i^{2(2-c)}} = \begin{cases} O(1) & \text{for} \quad c < 3/2, \\ O(\ln k) & \text{for} \quad c = 3/2, \\ O\left(k^{2c-3}\right) & \text{for} \quad c > 3/2. \end{cases}$$

Hence

$$\mathbf{E}\sigma_k^2 = \frac{O(1)}{k^{2c}} \sum\nolimits_{i=1}^{k} \frac{1}{i^{2(2-c)}} = \begin{cases} O(k^{-2c}) & \text{for} \quad c < 3/2, \\ O(k^{-2c} \ln k) & \text{for} \quad c = 3/2, \\ O\left(k^{-3}\right) & \text{for} \quad c > 3/2. \end{cases} \tag{B.11}$$

In any case, for $c > 1/2$,

$$\sum_{k=1}^{\infty} \mathbf{E}\sigma_k^2 < \infty.$$

As a consequence, by monotone convergence,

$$\sum_{k=1}^{\infty} \sigma_k^2 < \infty \quad \text{a.s..} \tag{B.12}$$

Since $\eta_t = \gamma\sigma_t$, it follows from a standard martingale convergence theorem (cf. (Lai and Wei, 1982b, Lemma 2)) that

$$v_{2t} = \sum_{k=2}^{t} \eta_{k-1}\varepsilon_k \to v_2 \quad \text{a.s.} \tag{B.13}$$

for some finite random variable v_2.

As to w_2, note that (B.11) also implies that

$$k^{2p}\mathbf{E}\sigma_k^2 \to 0 \tag{B.14a}$$

for any $p < \min\{c, 3/2\}$, and even

$$\sum_{k=1}^{\infty} k^{2p}\mathbf{E}\sigma_k^2 < \infty \tag{B.14b}$$

for any $p < \min\{c, 3/2\} - 1/2$. Consider now the mean $\overline{\sigma}_t = (1/t)\sum_{k=1}^t \sigma_k$ together with the sequences $z_k = k^p\sigma_k$ and

$$Z_t = \frac{1}{\sum_{k=1}^t k^{-p}} \sum_{k=1}^t k^{-p} z_k = \frac{1}{\sum_{k=1}^t k^{-p}} \sum_{k=1}^t \sigma_k.$$

From (B.14a) it follows that

$$\mathbf{E}\,|Z_t| \leq \frac{\sum_{k=1}^t k^{-p}\mathbf{E}\,|z_k|}{\sum_{k=1}^t k^{-p}} \leq \frac{\sum_{k=1}^t k^{-p}\sqrt{\mathbf{E}\,[z_k^2]}}{\sum_{k=1}^t k^{-p}} = O(1).$$

Since, for $0 < p < 1$, $\sum_{k=1}^t k^{-p} = O(t^{1-p})$, this implies that

$$\mathbf{E}\,|t^p\overline{\sigma}_t| = \frac{1}{t^{1-p}}\mathbf{E}\left|\sum_{k=1}^t \sigma_k\right| = \frac{\sum_{k=1}^t k^{-p}}{t^{1-p}}\mathbf{E}\,|Z_t| = O(1).$$

As a consequence,

$$t^p\overline{\sigma}_t = O_P(1). \tag{B.15}$$

Finally, remembering that $\eta_t = \gamma\sigma_t$ and by definition of w_{2t}, we obtain from Kolmogorov's strong LLN that, for $1/2 < p < \min\{c, 3/2\}$,

$$w_{2t} = O(1)t^p\overline{\eta}_t\frac{1}{t^p}\sum_{k=2}^t \varepsilon_k = o_P(1). \tag{B.16}$$

(B.10), (B.13) and (B.16) together yield

Intermediate result 2. For $c > 1/2$, $u_{2t} \xrightarrow{P} 0$.

Remark B.1a. For $c > 1$, we may find $1/2 < p < 1$ s.t. (B.14b) holds. By Chebychev's inequality, then,

$$\sum_{k=1}^{\infty}\mathbf{P}\,(|z_k| > \varepsilon) \leq \varepsilon^{-2}\sum_{k=1}^{\infty} k^{2p}\mathbf{E}\sigma_k^2 < \infty$$

for every $\varepsilon > 0$. This means that $z_k = k^p\sigma_k$ converges *completely* to zero (cf. Stout (1974)). By the Borel-Cantelli Lemma, it will then follows that

$$\mathbf{P}\,(|z_k| > \varepsilon \text{ i.o.}) = 0,$$

which is equivalent to almost sure convergence of z_k to zero. As a further consequence, by the Toeplitz lemma,

$$Z_t = \frac{\sum_{k=1}^t k^{-p} z_k}{\sum_{k=1}^t k^{-p}} \to 0 \quad \text{a.s..}$$

Hence

$$t^p \overline{\sigma}_t = \frac{1}{t^{1-p}} \sum_{k=1}^t \sigma_k = \frac{\sum_{k=1}^t k^{-p}}{t^{1-p}} Z_t \to 0 \quad \text{a.s..}$$

As a consequence, convergence $w_{2t} \to 0$ will hold with probability one, and so will $u_{2t} \to 0$.

Remark B.1b. If, in addition, $\mathbf{E}\,|\varepsilon_i|^{2/\alpha} < \infty$ holds for some $0 < \alpha < c - 1/2$, Theorem 4.1.3 in Stout (1974) may be used to show that this can also be achieved for $1/2 < c \leq 1$.

B.2 The term u_1

Ad u_1
Consider the decomposition

$$u_{1t} = \sum_{k=2}^t \xi_{k-1}\varepsilon_k - \frac{t-1}{t}\overline{\xi}_{t-1} \sum_{k=2}^t \varepsilon_k = v_{1t} - w_{1t}, \tag{B.17}$$

which is analogous to (B.10) in Appendix B.1. Also in analogy, we have that $\xi_t = \gamma\rho_t$, where

$$\rho_k = \frac{1}{k^c} s_k = \frac{1}{k^c} \sum\nolimits_{i=1}^k \frac{1}{i^{1-c}} \varepsilon_i.$$

Ad w_1
Since

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

we have

$$k\mathbf{E}\rho_k^2 = O(1). \tag{B.18}$$

Hence, arguing as in Appendix B.1 (but now with $p = 1/2$), we find that

$$t^{1/2}\overline{\rho}_t = O_P(1).$$

Hence, by the central limit theorem,

$$\overline{\xi}_{t-1} \sum_{k=2}^t \varepsilon_k = O_P(1) \frac{1}{\sqrt{t}} \sum_{k=2}^t \varepsilon_k = O_P(1).$$

Intermediate result 3. For $c > 1/2$, w_{1t} is bounded in probability.

Intermediate result 3 together with (B.17) shows (B.4).

Remark B.2. If w_{1t} were bounded a.s., then the $O_P(1)$ - term in (B.4) would actually be $O(1)$ a.s.. If, in addition, the situation in Remark B.1 is assumed, then also u_{21} would be bounded a.s. and, a forteriori, *strong consistency* would hold for the OLS estimator. Actually, what one needs to show is that in (B.6) convergence of u_{1T}/A_T^1 takes place with probability one instead of merely in probability.

B.3 The terms A^2 and A^3

Ad A_T^3

$$A_T^3 = \sum_{t=1}^{T} \left(\zeta_{t-1} - \frac{T-1}{T} \overline{\zeta}_{T-1} \right)^2 \leq 2 \left[\sum_{t=1}^{T} \zeta_{t-1}^2 + \sum_{t=1}^{T} \overline{\zeta}_{T-1}^2 \right].$$

Taking account of the definition of ζ_t (cf. (4.31)) and (B.9),

$$\begin{aligned} \sum_{t=1}^{T} \zeta_t^2 &\leq O(1) \sum_{t=1}^{T} \left[\frac{1}{t^{2c}} + \frac{1}{t^2} \right] = O(1), \\ \sum_{t=1}^{T} \overline{\zeta}_T^2 &\leq O(1) T \left[\frac{1}{T^{2c}} + \frac{\ln^2 T}{T^2} \right] = O(1). \end{aligned}$$

This shows that $\lim_{T\to\infty} A_T^3 < \infty$ a.s..

Ad A_T^2

$$A_T^2 = \sum_{t=1}^{T} \left(\eta_{t-1} - \frac{T-1}{T} \overline{\eta}_{T-1} \right)^2 \leq 2 \left[\sum_{t=1}^{T} \eta_{t-1}^2 + \sum_{t=1}^{T} \overline{\eta}_{T-1}^2 \right].$$

By (B.15), $\overline{\eta}_T = O_P(1) T^{-p}$ for some $p > 1/2$, so that

$$\sum_{t=1}^{T} \overline{\eta}_{T-1}^2 = O_P(1) \frac{1}{T^{2p-1}}.$$

Also, since $\eta_t = \gamma \sigma_t$ (cf. Appendix B.1),

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

is just (B.12). Hence $A_T^2 = O_P(1)$. Since A_T^2 is monotone increasing, this implies that $\lim_{T\to\infty} A_T^2 < \infty$ a.s..

B.4 The term A^1

Ad A_T^1

$$A_T^1 = \sum_{t=1}^{T} \left(\xi_{t-1} - \frac{T-1}{T} \overline{\xi}_{T-1} \right)^2 \leq 2 \left[\sum_{t=1}^{T} \xi_{t-1}^2 + \sum_{t=1}^{T} \xi_{T-1}^2 \right]. \tag{B.19}$$

By (B.18), $\mathbf{E}\overline{\rho}_t^2 = O\left(1/t\right)$ .Therefore, remembering that $\overline{\xi} = \gamma\overline{\rho}_t$, it follows that $\mathbf{E}\sum_{t=1}^{T}\overline{\xi}_{T-1}^2 = O(1)$ and hence

$$\sum_{t=1}^{T}\overline{\xi}_{T-1}^2 = O_P(1).$$

If it can be shown that

$$\sum_{t=1}^{\infty}\xi_t^2 = \infty \quad \text{a.s.}, \tag{B.20}$$

then (B.5a) will follow by a similar argument (an "in probability" version) as for (B.8).

As to the first term in (B.19), it can be shown that $\mathbf{E}\sum_{t=1}^{T}\xi_t^2 = O(\ln T)$. Therefore, in order to prove (B.5b), we will have to take a different approach.

Ad (B.5b)

Remark B.3. Actually, the following calculations could be made exclusively in discrete time, by referring to the discrete time Ito formula (cf., e.g., Rogers and Williams (1987)). However, to our taste, things look nicer and seem easier to handle if passing to continuous time.

Define processes in continuous time $t \geq 1$ by putting

$$Y_t = \sum_{k=1}^{[t]}\frac{1}{k^{1-c}}\varepsilon_k = s_{[t]},\ X_t = \sum_{k=1}^{[t]}\frac{Y_{k-1}}{k^c}\varepsilon_k,\ Z_t = \frac{1}{[t]^{2c-1}}$$

($[t]$ denoting the largest integer $\leq t, Y_0 = 0$). Y and X are L^2 - martingales w.r.t. the filtration $(\mathcal{F}_t)_{t\geq 1}$, with $\mathcal{F}_t = \mathcal{F}_{[t]}^{\varepsilon}$. Note that, for martingales of the form

$$M_t = \sum_{k=1}^{[t]} v_k\varepsilon_k$$

(with (v_k) predictable w.r.t. the filtration $(\mathcal{F}_n^{\varepsilon})$), the predictable quadratic variation and the quadratic variation processes are given by

$$\langle M\rangle_t = \sigma^2\sum_{k=1}^{[t]} v_k^2 \quad \text{and} \quad [M]_t = \sum_{s\leq t}\left(\Delta M_s\right)^2 = \sum_{k=1}^{[t]} v_k^2\varepsilon_k^2,$$

respectively. Here ΔM_s are the increments $\Delta M_s = M_s - M_{s-}$. If referring to the corresponding martingale in discrete time, we will use the time index n instead: $M_n, n = 1, 2, \ldots$. The corresponding quadratic variation processes are, of course, $\langle M\rangle_n$ and $[M]_n$. Note that, in our notation so far, $Y_n = s_n$.

We have the following properties.

(i)

$$\int_1^t Z_{s-}Y_{s-}dY_s = \sum_{k=1}^{[t]}\frac{Y_{k-1}}{k^{2c-1}}\frac{1}{k^{1-c}}\varepsilon_k = X_t.$$

(ii)

$$\langle Y\rangle_n = \sigma^2 \sum_{i=1}^{n} \frac{1}{i^{2(1-c)}} = O(1)\left[1 + n^{2c-1}\right]. \tag{B.21}$$

Since $\langle Y\rangle_\infty = \infty$, it follows from the LIL that

$$\overline{\lim}_{n\to\infty} \frac{|Y_n|}{\sqrt{2\langle Y\rangle_n \log_2\langle Y\rangle_n}} = \sigma^2 \quad \text{a.s..}$$

For an appropriate LIL, cf. Chow and Teicher (1973), Theorem 1. Hence,

$$\frac{|Y_n|}{n^{2c-1}} = O\left(\log_2 n\right) \quad \text{a.s.,} \tag{B.22a}$$

or, in terms of the continuous time processes introduced above,

$$Y_t^2 Z_t = O\left(\log_2 t\right) \text{ a.s.} \tag{B.22b}$$

(iii)

$$[Y]_t = \sum_{s\le t} (\Delta Y_s)^2 = \sum_{k=1}^{[t]} \frac{1}{k^{2(1-c)}} \varepsilon_k^2$$

(iv) By Ito's partial integration formula (cf. Mtivier (1982)),

$$Y_t^2 = Y_1^2 + 2\int_1^t Y_{s-} dY_s + [Y]_t.$$

(v) By the same formula,

$$\begin{aligned} Y_t^2 Z_t &= Y_1^2 Z_1 + \int_1^t Y_{s-}^2 dZ_s + \int_1^t Z_{s-} dY_s^2 + \left[Y^2, Z\right]_t \\ &= O(1) + \int_1^t Y_{s-}^2 dZ_s + 2\int_1^t Z_{s-} Y_{s-} dY_s \\ &\qquad + \int_1^t Z_{s-} d[Y]_s + \left[Y^2, Z\right]_t. \end{aligned} \tag{B.23}$$

(vi) Since

$$\Delta Z_k = \frac{1}{k^{2c-1}} - \frac{1}{(k-1)^{2c-1}} = O(1)\frac{1}{k^{2c}},$$

we find that

$$\int_1^t Y_{s-}^2 dZ_s = \sum_{k=1}^{[t]} Y_{k-1}^2 \Delta Z_k = O(1)\sum_{i=1}^{[t]} Y_{k-1}^2 \frac{1}{k^{2c}} = O(1)\langle X\rangle_t. \tag{B.24}$$

(vii)

$$\int_1^t Z_{s-} d[Y]_s = \sum_{k=1}^{[t]} \frac{1}{k^{2c-1}} \frac{1}{k^{2(1-c)}} \varepsilon_k^2 = \sum_{k=1}^{[t]} \frac{1}{k} \varepsilon_k^2.$$

By the martingale convergence theorem (discrete time),

$$\sum_{k=1}^{n} \frac{1}{k}\varepsilon_k^2 = \sum_{k=1}^{n} \frac{1}{k}\left(\varepsilon_k^2 - \sigma^2\right) + \sum_{k=1}^{n} \frac{1}{k} = O(1)\left[1 + \log n\right],$$

whence

$$\int_1^t Z_{s-} d\left[Y\right]_s = O(1)\left[1 + \log t\right]. \tag{B.25}$$

(vii)

$$\left[Y^2, Z\right]_t = \sum_{s \le t} \Delta\left(Y_s^2\right) \Delta Z_s = O(1) \sum_{k=1}^{[t]} \Delta\left(Y_k^2\right) \frac{1}{k^{2c}}.$$

But

$$\Delta\left(Y_k^2\right) = 2\frac{\varepsilon_k}{k^{1-c}} \sum_{i=1}^{k-1} \frac{1}{i^{1-c}}\varepsilon_i + \frac{\varepsilon_k^2}{k^{2(1-c)}} = 2\frac{\varepsilon_k}{k^{1-c}} Y_{k-1} + \frac{\varepsilon_k^2}{k^{2(1-c)}},$$

so that, taking account of (B.22a),

$$\begin{aligned} \left|\Delta\left(Y_k^2\right)\right| &\le O(1)\frac{|\varepsilon_k|}{k^{1-c}}\sqrt{k^{2c-1}\log_2 k} + \frac{\varepsilon_k^2}{k^{2(1-c)}} \\ &= O(1)\frac{|\varepsilon_k|}{k^{3/2-2c}}\sqrt{\log_2 k} + \frac{\varepsilon_k^2}{k^{2(1-c)}}. \end{aligned}$$

Hence

$$\left[Y^2, Z\right]_t = O(1) \sum_{k=1}^{[t]} \frac{|\varepsilon_k|}{k^{3/2}}\sqrt{\log_2 k} + O(1) \sum_{k=1}^{[t]} \frac{\varepsilon_k^2}{k^2}.$$

The second term converges to some finite limit (cf. Lai and Wei (1982b), Lemma 2 (iii)). As to the first term,

$$\sum_{k=1}^{[t]} \frac{|\varepsilon_k|}{k^{3/2}}\sqrt{\log_2 k} = \sum_{k=1}^{[t]} \frac{|\varepsilon_k| - \mathbf{E}\,|\varepsilon_k|}{k^{3/2}}\sqrt{\log_2 k} + \mathbf{E}\,|\varepsilon_k| \sum_{k=1}^{[t]} \frac{1}{k^{3/2}}\sqrt{\log_2 k}.$$

The first term (a martingale) converges to some finite limit by the standard martingale convergence theorem, the second trivially. Hence

$$\left[Y^2, Z\right]_\infty < \infty \quad \text{a.s..} \tag{B.26}$$

Now consider the identity (B.23) again, making use of (i) and (B.22b) as well as (B.24)-(B.26):

$$O(1)\log_2 t = O(1) + O(1)\langle X\rangle_t + 2X_t + \log t. \tag{B.27}$$

X_t converges a.s. to some finite limit on the set $\{\langle X\rangle_\infty < \infty\}$. This is compatible with (B.27) only if $\mathbf{P}\left(\langle X\rangle_\infty < \infty\right) = 0$. Therefore, finally, we conclude that, with probability one,

$$\lim_{n\to\infty} \langle X\rangle_n = \sigma^2 \sum_{k=1}^{n} \frac{Y_{k-1}^2}{k^{2c}} = \infty. \tag{B.28}$$

Coming back to (B.5b), we have that $\xi_n = \gamma n^{-c} Y_n$ and hence

$$\sum_{k=1}^{n} \xi_k^2 = \gamma \sum_{k=1}^{n} \frac{1}{k^{2c}} Y_k^2 = \gamma \sum_{k=2}^{n} \left(\frac{k}{k-1} \right)^{2c} \frac{1}{k^{2c}} Y_{k-1}^2 \geq \gamma \langle X \rangle_n.$$

This proves the assertion.

B.5 Strong consistency: proof of Theorem 4.5

According to Remark B.2 in Appendix B.2, strong consistency will hold in the setting of Remark 4 provided it can be shown that

$$\frac{u_{1T}}{A_T^1} \to 0 \quad \text{a.s..} \tag{B.29}$$

We will show (B.29) making use of the results of Appendix A. In particular, we make the following

Maintained assumption 1: The ε_i are Gaussian.

In view of the identity (A.5) (here with $x_i = \xi_{i-1}$), we have that

$$\begin{aligned} u_{1t} &= \sum_{i=2}^{t} \left(\xi_{i-1} - \frac{t-1}{t} \overline{\xi}_{t-1} \right) \varepsilon_i = \sum_{i=2}^{t} \frac{i-1}{i} \left(\xi_{i-1} - \frac{i-2}{i-1} \overline{\xi}_{i-2} \right) v_i \\ &= \sum_{i=2}^{t} \frac{i-1}{i} \phi_{i-1} v_i, \end{aligned} \tag{B.30}$$

where $v_i = \varepsilon_i - \overline{\varepsilon}_{i-1}$. The coefficients ϕ_t may be decomposed as follows:

$$\begin{aligned} \phi_t &= \xi_t - \frac{t-1}{t} \overline{\xi}_{t-1} \\ &= \frac{1}{t^c} \sum_{i=1}^{t} \frac{1}{i^{1-c}} \varepsilon_i - \frac{1}{t} \sum_{j=1}^{t-1} \frac{1}{j^c} \sum_{i=1}^{j} \frac{1}{i^{1-c}} \varepsilon_i \\ &= t^{1-c} \sum_{i=1}^{t} \frac{1}{i^{1-c}} \varepsilon_i - \frac{1}{t} \sum_{i=1}^{t-1} \frac{1}{i^{1-c}} \left(\sum_{j=i}^{t-1} \frac{1}{j^c} \right) \varepsilon_i \\ &= \frac{1}{t} \left[\varepsilon_t + \sum_{i=1}^{t-1} h_{ti} \varepsilon_i \right] \end{aligned} \tag{B.31}$$

with

$$h_{ti} = \frac{1}{i^{1-c}} \left[t^{1-c} - \sum_{j=i}^{t-1} \frac{1}{j^c} \right].$$

From now on, we make the following

Maintained assumption 3: $c > 1$.

Then

$$\begin{aligned}
\sum_{i=1}^{t-1} h_{ti} &= t^{1-c}\sum_{i=1}^{t-1}\frac{1}{i^{1-c}} - \sum_{i=1}^{t-1}\frac{1}{i^{1-c}}\sum_{j=i}^{t-1}\frac{1}{j^c} \\
&= t^{1-c}\sum_{i=1}^{t-1}\frac{1}{i^{1-c}} - \sum_{j=1}^{t-1}\frac{1}{j^c}\sum_{i=1}^{j}\frac{1}{i^{1-c}} \\
&= \frac{t^{1-c}}{c}\left[(t-1)^c - 1 + O\left(\frac{1}{t^{1-c}}\right)\right] \\
&\qquad - \frac{1}{c}\sum_{j=1}^{t-1}\frac{1}{j^c}\left[j^c - 1 + O\left(\frac{1}{j^{1-c}}\right)\right] \\
&= \frac{t-1}{c}\left[1 + O\left(\frac{1}{t}\right)\right] - \frac{1}{c}(t-1) + \frac{1}{c}\sum_{j=1}^{t-1}\frac{1}{j^c} + O(1)\sum_{j=1}^{t-1}\frac{1}{j} \\
&= R_t
\end{aligned}$$

with $R_t = O(\ln t)$. Hence (B.31) may be written

$$\begin{aligned}
\phi_t &= \frac{1}{t}\left[\left(1 + \sum_{i=1}^{t-1} h_{ti}\right)\varepsilon_t - \sum_{i=1}^{t-1} h_{ti}\left(\varepsilon_t - \varepsilon_i\right)\right] \\
&= c_t\varepsilon_t - \frac{1}{t}\sum_{i=1}^{t-1} h_{ti}\left(\varepsilon_t - \varepsilon_i\right) \\
&= c_t\varepsilon_t - z_t, \qquad \text{(B.32)}
\end{aligned}$$

with $c_t = O(R_t/t)$ and

$$z_t = \frac{1}{t}\sum_{i=1}^{t-1} h_{ti}\left(\varepsilon_t - \varepsilon_i\right).$$

Inserting into (B.30), we obtain

$$u_{1t} = \sum_{i=2}^{t}\frac{i-1}{i}c_{i-1}\varepsilon_{i-1}v_i - \sum_{i=2}^{t}\frac{i-1}{i}z_{i-1}v_i = U_t - \widetilde{u}_{1t}. \qquad \text{(B.33)}$$

The first term U_t may be decomposed as

$$\begin{aligned}
U_t &= \sum_{i=2}^{t}\frac{i-1}{i}c_{i-1}\varepsilon_{i-1}v_i \\
&= \sum_{i=2}^{t}\frac{i-1}{i}c_{i-1}\varepsilon_{i-1}\varepsilon_i - \sum_{i=2}^{t}\frac{i-1}{i}c_{i-1}\overline{\varepsilon}_{i-1}\varepsilon_{i-1} \\
&= V_t - W_t. \qquad \text{(B.34)}
\end{aligned}$$

Since

$$\mathbf{E}\sum_{i=2}^{T} c_{i-1}^2\varepsilon_{i-1}^2 = O(1)\sum_{i=2}^{T}\frac{\ln^2 i}{i^2} < \infty,$$

it follows that $\sum_{i=2}^{\infty} c_{i-1}^2 \varepsilon_{i-1}^2 < \infty$ a.s. and therefore, by the standard martingale convergence theorem,

$$V_t = O(1) \quad \text{a.s.} \tag{B.35}$$

As to the second term,

$$\begin{aligned} W_t &= \sum_{i=2}^{t} \frac{i-1}{i} c_{i-1} \overline{\varepsilon}_{i-1} \varepsilon_{i-1} \\ &= \sum_{i=2}^{t} c_{i-1} \frac{1}{i} \left[\sum_{j=1}^{i-2} \varepsilon_j + \varepsilon_{i-1} \right] \varepsilon_{i-1} \\ &= \sum_{i=2}^{t} \underbrace{\frac{1}{i} c_{i-1} \left[\sum_{j=1}^{i-2} \varepsilon_j \right]}_{u_i} \varepsilon_{i-1} + \sum_{i=2}^{t} \frac{1}{i} c_{i-1} \varepsilon_{i-1}^2 \\ &= W_t^1 + W_t^2. \end{aligned} \tag{B.36}$$

Since $\mathbf{E} u_i^2 = O\left(i^{-3} \ln^2 i\right)$, it holds that $\sum_{i=2}^{\infty} u_i^2 < \infty$ a.s. and hence

$$W_t^1 = O(1) \quad \text{a.s..} \tag{B.37a}$$

Also, $\sum_{i=2}^{\infty} i^{-1} |c_{i-1}| < \infty$ and therefore

$$W_t^2 = O(1) \quad \text{a.s.} \tag{B.37b}$$

(cf. (2.12) in Lai and Wei (1982a)). Summarizing (B.33)-(B.37), we find that

$$u_{1t} = \widetilde{u}_{1t} + O(1) \quad \text{a.s.} \tag{B.38}$$

By definition (cf. (B.32)) together with Lemma A.1 in Appendix A,

$$z_t = \frac{1}{t} \sum_{i=1}^{t-1} h_{ti} \left(\varepsilon_t - \varepsilon_i \right)$$

is measurable with respect to $\mathcal{F}_t^v$. In other words, the sequence (z_{t-1}) is predictable with respect to the filtration $(\mathcal{F}_t^v)$. Hence the sequence

$$\widetilde{u}_{1t} = \sum_{i=2}^{t} \frac{i-1}{i} z_{i-1} v_i$$

is a martingale with respect to the filtration $(\mathcal{F}_t^v)$. Arguing as in Appendix A.2,

$$\frac{\widetilde{u}_{1t}}{\widetilde{A}_t} \to 0 \quad \text{a.s. on the set } \left\{ \widetilde{A}_\infty = \infty \right\},$$

where

$$\widetilde{A}_t = \sum_{i=2}^{t} z_{i-1}^2.$$

It remains to show that $\widetilde{A}_\infty = \infty$ a.s. But, as in (A.6),

$$\begin{aligned} A_t^1 &= \sum_{i=1}^{t} \left(\xi_{i-1} - \frac{t-1}{t} \overline{\xi}_{t-1} \right)^2 = \sum_{i=2}^{t} \frac{i-1}{i} \left(\xi_{i-1} - \frac{i-2}{i-1} \overline{\xi}_{i-2} \right)^2 \\ &= \sum_{i=2}^{t} \frac{i-1}{i} \phi_{i-1}^2. \end{aligned}$$

By virtue of (B.32), this may be written as

$$A_t^1 = \sum_{i=2}^{t} \frac{i-1}{i} \left(c_{i-1}\varepsilon_{i-1} - z_{i-1} \right)^2 \leq 2 \left[\sum_{i=2}^{t} c_{i-1}^2 \varepsilon_{i-1}^2 + \widetilde{A}_t \right].$$

Since $\sum_{i=2}^{\infty} c_{i-1}^2 \varepsilon_{i-1}^2 < \infty$ a.s. and $\lim_{t\to\infty} A_t^1 = \infty$ a.s., it follows that $\widetilde{A}_\infty = \infty$ a.s.. Therefore, finally,

$$\lim_{T\to\infty} \frac{u_{1T}}{A_T^1} = \lim_{T\to\infty} \frac{\widetilde{u}_{1T}}{A_T^1} = \lim_{T\to\infty} \frac{u_{1T}}{\widetilde{A}_T} \frac{\widetilde{A}_T}{A_T^1} = 0 \quad \text{a.s.},$$

thus proving (B.29).

For $c = 1$, this approach does not work any longer since then $R_t = (t-1)(1-t/2) = O(t^2)$.

References

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Anderson, T. W. and J. B. Taylor (1976). Strong consistency of least squares estimates in normal linear regression. *Annals of Statistics 4*, 788–790.

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

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

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

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

Christopeit, N. and M. Massmann (2009). A stochastic approximation approach to models with forecast feedback. Mimeo.

Christopeit, N. and M. Massmann (2010a). Asymptotic normality of ordinary least squares estimators in regression models with adaptive learning. Mimeo.

Christopeit, N. and M. Massmann (2010b). Supplementary material to "Consistent estimation of structural parameters in regression models with adaptive learning". Mimeo.

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

Doob, J. L. (1953). *Stochastic Processes*. New York: John Wiley.

Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 34*, 119–127.

Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of Mathematical Statistics 34*, 447–456.

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

Evans, G. W. and S. Honkapohja (1998). Economic dynamics with learning: new stability results. *Review of Economic Studies 65*, 23–44.

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

Evans, G. W. and S. Honkapohja (2008). Expectations, learning and monetary policy: an overview of recent research. Paper prepared for the Bank of Chile conference 'Monetary Policy under Uncertainty and Learning' in November 2007.

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

Lai, T. L. and H. Robbins (1977). Strong consistency of least squares estimates in regression models. *Proceedings of the National Academy of Sciences in the USA 74*, 2667–2669.

Lai, T. L., H. Robbins, and C. Z. Wei (1978). Strong consistency of least squares estimates in multiple regression. *Proceedings of the National Academy of Sciences in the USA 75*, 3034–3036.

Lai, T. L., H. Robbins, and C. Z. Wei (1979). Strong consistency of least squares estimates in multiple regression II. *Journal of Multivariate Analysis 9*, 343–361.

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

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

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

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

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

Marcet, A. and T. J. Sargent (1989a). Convergence of least squares learning mechanisms in environments with hidden state variables and private information. *Journal of Political Economy 97*, 1306–1322.

Marcet, A. and T. J. Sargent (1989b). Convergence of least squares learning mechanisms in self-referential linear stochastic models. *Journal of Economic Theory 48*, 337–368.

Milani, F. (2006). A Bayesian DSGE model with infinite-horizon learning: Do "mechanical" sources of persistence become superfluous? *International Journal of Central Banking 2*(3), 87–106.

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

Mtivier, M. (1982). *Semimartingales: A Course on Stochastic Processes*. Berlin: De Gruyter.

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

Robbins, H. and D. Sigmund (1971). A convergence theorem for non-negative almost-supermartingales and some applications. In J. S. Rustagi (Ed.), *Optimizing Methods in Statistics*, pp. 233–257. New York: Academic Press.

Rogers, L. C. G. and D. Williams (1987). *Diffusions, Markov Processes and Martingales*, Volume 2. Chichester: Wiley.

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

Stout, W. F. (1974). *Almost Sure Convergence*. New York: Academic Press.

Walk, H. (1985). Almost sure convergence of stochastic approximation processes. *Statistics and Decisions 2*, 137–141.

Walk, H. and L. Zsidó (1989). Convergence of the Robbins-Munro method for linear problems in a Banach space. *Journal of Mathematical Analysis and Applications 139*, 152–177.

Zenner, M. (1996). *Learning to Become Rational*, Volume 439 of *Lecture Notes in Economics and Mathematical Systems*. Berlin: Springer.