

TI 2012-109/III  
Tinbergen Institute Discussion Paper



# Strong Consistency of the Least-Squares Estimator in Simple Regression Models with Stochastic Regressors

*Norbert Christopeit*<sup>1</sup>

*Michael Massmann*<sup>2</sup>

<sup>1</sup> *University of Bonn;*

<sup>2</sup> *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute.*

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

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

Tinbergen Institute has two locations:

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

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

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

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

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

# Strong Consistency of the Least-Squares Estimator in Simple Regression Models with Stochastic Regressors

Norbert Christopeit  
University of Bonn

Michael Massmann  
Vrije Universiteit Amsterdam,  
Tinbergen Institute, and  
EBS University, Wiesbaden

7 October 2012

## Abstract

Strong consistency of least squares estimators of the slope parameter in simple linear regression models is established for predetermined stochastic regressors. The main result covers a class of models which falls outside the applicability of what is presently available in the literature. An application to the identification of economic models with adaptive learning is discussed.

## 1 Introduction

In this paper, we revisit the simple linear regression model

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, 2, \dots, \quad (1.1)$$

where the  $\varepsilon_i$  are unobservable random errors with zero means,  $\alpha, \beta$  are unknown parameters and  $y_i$  is the observed response to the deterministic or stochastic scalar regressors  $x_i$ . Our interest is in the consistency of the ordinary least squares estimator (OLSE) for the parameters based on the first  $n$  pairs of observations  $x_1, y_1, \dots, x_n, y_n$ , and in particular of the OLSE for the slope parameter  $\beta$ . The latter is given by

$$\hat{\beta}_n = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}, \quad (1.2)$$

where  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ . Inserting (1.1) into (1.2) yields the representation

$$\hat{\beta}_n - \beta = \frac{\sum_{i=1}^n (x_i - \bar{x}_n) \varepsilon_i}{A_n}, \quad (1.3)$$

where we have introduced

$$A_n = \sum_{i=1}^n (x_i - \bar{x}_n)^2 = n \left[ \overline{x_n^2} - (\bar{x}_n)^2 \right]. \quad (1.4)$$

In section 2, we shall give a brief review of the main results on the strong consistency of the OLSE available in the literature. Actually, most of the work done is related to the general multivariate regression model

$$y_i = \theta' x_i + \varepsilon_i, \quad (1.5)$$

where now  $x_i = (x_{1i}, \dots, x_{pi})'$  are  $p$ -dimensional regressors and  $\theta = (\theta_1, \dots, \theta_p)'$  is a vector of unknown parameters. In this case, the OLSE  $\widehat{\theta}_n$  is given by

$$\widehat{\theta}_n - \theta = M_n^{-1} \sum_{i=1}^n x_i \varepsilon_i, \quad (1.6)$$

where

$$M_n = \sum_{i=1}^n x_i x_i'.$$

The conditions to be imposed on the regressors to ensure consistency depend heavily on whether the latter are deterministic or stochastic. Our main results to be presented in section 3 concern model (1.1) with predetermined stochastic regressors and i.i.d. Gaussian error terms. The proofs will be given in section 4. In section 5, we apply these results to an economic model with adaptive learning, for which the conditions provided by the available literature are not met. The associated proof follows in section 6.

## 2 Review of existing results

In this section, we will give a brief overview of known results on consistency of the OLSE. As pointed out in the Introduction, it will be necessary to distinguish the cases of deterministic and stochastic regressors. For a recent account of this subject in the context of stochastic approximation see Lai (2003).

### 2.1 Deterministic regressors

When the regressors  $x_i$  are deterministic, the OLSE  $\widehat{\theta}_n$  is unbiased. For uncorrelated error terms with common variance  $\mathbf{E}\varepsilon_i^2 = \sigma^2$ , i.e. for white noise errors (the so-called Gauß-Markov model),

$$\text{cov}(\widehat{\theta}_n, \widehat{\theta}_n) = \sigma^2 M_n^{-1}.$$

It is thus plain that the condition

$$\lim_{n \rightarrow \infty} M_n^{-1} = 0 \quad (2.1)$$

is both necessary and sufficient for  $L^2$ -consistency. Equivalently,

$$\lambda_{\min}(M_n) \rightarrow \infty. \quad (2.2)$$

Hence (2.1) is also sufficient for *weak consistency*, i.e. convergence  $\widehat{\theta}_n \rightarrow \theta$  in probability. Actually, as shown by Eicker (1963) (for normal disturbances) and Drygas (1976), it is also necessary for weak consistency. Needless to say that (2.1) is much weaker than the classical textbook condition

$$\frac{1}{n} M_n \rightarrow M,$$

where  $M$  is some positive definite matrix.

The analysis is more involved when it comes to *strong consistency*, i.e. almost sure convergence  $\widehat{\theta}_n \rightarrow \theta$ . For independent normally distributed  $\varepsilon_i$ , Anderson and Taylor (1976) show that (2.1) is indeed necessary and sufficient. Without the normality assumption, but uniformly bounded variances, additional restrictions like

$$M_n^{-1} \text{diag}(d_{n1}^2, \dots, d_{np}^2) = O(1), \quad (2.3)$$

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, however, much stronger than (2.1). A sufficient condition for (2.3) is

$$\lambda_{\max}(M_n) = O[\lambda_{\min}(M_n)],$$

see Christopheit (1986).

The ultimate affirmative answer is given in Lai, Robbins, and Wei (1978), who prove that (2.1) is indeed sufficient for strong consistency. Actually, all that is required of the error terms  $\varepsilon_i$  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 is extended to a more general class of  $\varepsilon_i$ .

Returning to the simple regression model (1.1), if one is only interested in the estimation of the slope  $\beta$ , then the complete answer is given in Lai and Robbins (1977). For white noise errors, they show that the condition

$$\lim_{n \rightarrow \infty} A_n = \infty \tag{2.4}$$

is *necessary and sufficient for weak consistency*. If one assumes that the  $\varepsilon_i$  are i.i.d. with  $\mathbf{E}[\varepsilon_i^2 (\log(1 + |\varepsilon_i|))^r] < \infty$  for some  $r > 1$ , then (2.4) is even sufficient for *strong consistency*. Indeed, Lai and Robbins (1977) show that

$$\lim_{n \rightarrow \infty} \sqrt{\frac{A_n}{(\log A_n)^{1+\delta}}} (\hat{\beta}_n - \beta) = 0 \quad \text{a.s.} \tag{2.5}$$

for every  $\delta > 0$ .

## 2.2 Stochastic regressors

For stochastic regressors, the situation presents itself less clearcut. There seems to be no ‘minimal’ condition like (2.1). After some intermediate results obtained by Anderson and Taylor (1979) and Christopheit and Helmes (1980), involving the condition

$$\lambda_{\max}^r(M_n) = O[\lambda_{\min}(M_n)]$$

for some  $r > 0$ , the best available result at present seems to be the one obtained in Lai and Wei (1982a). 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  $\sup_n \mathbf{E}(\varepsilon_n^2 | \mathcal{F}_{n-1}) < \infty$  a.s., a sufficient condition for strong convergence of the least squares estimator of  $\alpha$  and  $\beta$  is the following:

$$\lambda_{\min}(M_n) \rightarrow \infty, \tag{2.6a}$$

$$[\log \lambda_{\max}(M_n)]^{1+\delta} = o(\lambda_{\min}(M_n)) \quad \text{a.s.} \tag{2.6b}$$

for some  $\delta > 0$ . If, in addition,  $\sup_n \mathbf{E}(|\varepsilon_n|^\alpha | \mathcal{F}_{n-1}) < \infty$  for some  $\alpha > 2$ , then it suffices to require (2.6b) for  $\delta = 0$ . Lai and Wei (1982a) give an example which shows that even a marginal violation of (2.6b), viz. replacing the  $o(\cdot)$  by an  $O(\cdot)$ , is destructive to consistency. With a good deal of more effort, one can obtain some refinements of the condition (2.6b), see Lai and Wei (1982b). Applied to the simple regression model (1.1) with  $\sup_n \mathbf{E}(|\varepsilon_n|^\alpha | \mathcal{F}_{n-1}) < \infty$  for some  $\alpha > 2$ , the following sharpened version of condition (2.4) turns out to be sufficient for strong consistency of the *slope estimator*  $b_n$ :

$$\frac{A_n}{\log n} \rightarrow \infty \quad \text{a.s.} \tag{2.7}$$

If  $\overline{\lim}_{n \rightarrow \infty} \overline{x_n^2} \leq 1$ , then (2.7) is also sufficient for consistent OLS estimation of  $\alpha$ . Otherwise, additional conditions have to be imposed. As is shown in Lai and Wei (1982b), (2.6b) (with  $\delta = 0$ ) implies the conditions of their Theorem 2.

### 3 Main results

The main result in this paper is an extension of the minimal condition (2.4) for almost sure convergence of the OLSE of the slope parameter to predetermined stochastic regressors. In particular, it will turn out to be applicable to situations in which conditions (2.6b) or (2.7) are not satisfied. The price to be paid is a somewhat restrictive measurability condition to be imposed on the regressors as well as the restriction to Gaussian noise.

In order to formulate the result, introduce the processes

$$u_n = \sum_{i=1}^n (x_i - \bar{x}_n) \varepsilon_i, \quad n \geq 1, \quad (3.1)$$

and

$$v_n = \varepsilon_n - \bar{\varepsilon}_{n-1}, \quad n \geq 2, \quad (3.2)$$

as well as the filtration  $\mathcal{F}_n = \mathcal{F}_n^v = \sigma(v_2, \dots, v_n)$ ,  $n \geq 2$ . In section 4 an alternative description of  $\mathcal{F}_n$  will be given, namely  $\mathcal{F}_n = \sigma(\varepsilon_n - \varepsilon_{n-1}, \dots, \varepsilon_n - \varepsilon_1)$ , which is often easier to handle. The OLSE (1.3) may then be written in the form

$$\widehat{\beta}_n - \beta = \frac{u_n}{A_n}. \quad (3.3)$$

The basis of our approach will be the observation that, for i.i.d. Gaussian  $\varepsilon_n$ , the  $v_n$  are *independent* Gaussian with variance  $\frac{n}{n-1}\sigma^2$ . In particular,  $(v_n)_{n \geq 2}$  is a *martingale difference sequence* (MDS) with respect to the filtration  $(\mathcal{F}_n)_{n \geq 2}$ . Gaussianness seems to be essential for this property (cf. Remark 2 below). Making use of the algebraic identity

$$u_n = \sum_{i=1}^n \frac{i-1}{i} (x_i - \bar{x}_{i-1}) v_i \quad (3.4)$$

(cf. section 4), it turns out that  $(u_n)_{n \geq 3}$  is a *martingale transform*, provided the centered regressors  $x_i - \bar{x}_{i-1}$  are measurable with respect to  $\mathcal{F}_{i-1}$ , i.e. the process  $(x_n - \bar{x}_{n-1})_{n \geq 3}$  is predictable with respect to the filtration  $(\mathcal{F}_n)_{n \geq 2}$ . This observation makes it possible to approach the question of consistency by martingale convergence methods.

Our main result is the following.

**Theorem 1.** Suppose that the  $\varepsilon_i$  are independent  $N(0, \sigma^2)$ -distributed random variables and that the centered regressors  $x_n - \bar{x}_{n-1}$  are  $\mathcal{F}_{n-1}$ -measurable for  $n \geq 3$ . Assume further that

$$\lim_{n \rightarrow \infty} A_n = \infty \quad \text{a.s.} \quad (3.5)$$

Then,

$$\overline{\lim}_{n \rightarrow \infty} \sqrt{\frac{A_n}{2 \log_2 A_n}} |\widehat{\beta}_n - \beta| = \sigma. \quad (3.6)$$

**Corollary 1.** Under the assumptions of Theorem 1,

$$\sqrt{\frac{A_n}{\log A_n}}(\widehat{\beta}_n - \beta) \rightarrow 0 \text{ a.s.}$$

In particular, the OLSE for  $\beta$  is strongly consistent, i.e.

$$\widehat{\beta}_n \rightarrow \beta \text{ a.s.} \quad (3.7)$$

In contrast to the setting with deterministic regressors, condition (3.5) will generally not be necessary in the presence of stochastic regressors. However, a restricted version of the negation of (3.5) rules out consistency:

**Proposition 1.** Under the assumptions of Theorem 1, the condition

$$\mathbf{E}A_\infty < \infty \quad (3.8)$$

implies (weak and hence strong) inconsistency of the OLS estimator for the slope coefficient.

**Remark 1.** In situations in which the regressors do not satisfy the measurability condition of Theorem 1, it is often possible to split the centered regressors  $x_i - \bar{x}_{i-1}$  into two parts:

$$x_i - \bar{x}_{i-1} = w'_i + w''_i, \quad (3.9)$$

such that the sequence  $(w'_n)$  is predictable with respect to the filtration  $(\mathcal{F}_n)$ , whereas the  $w''_n$  do not affect the asymptotic behavior. More details will be provided in section 4.2, Remark 3.

## 4 Proof of main results

### 4.1 Some preliminaries

Define

$$e_{ni} = \varepsilon_n - \varepsilon_i$$

for  $n \geq 2$ ,  $i = 1, \dots, n-1$ . As usual, let  $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n z_i$ . Consider the filtrations  $(\mathcal{F}_n^v)_{n \geq 2}$  and  $(\mathcal{F}_n^e)_{n \geq 2}$ , where  $\mathcal{F}_n^v = \sigma(v_2, \dots, v_n)$  and  $\mathcal{F}_n^e = \sigma(e_{n1}, \dots, e_{n,n-1})$ . Note that  $\mathcal{F}_n^v \subset \mathcal{F}_n^\varepsilon = \sigma(\varepsilon_1, \dots, \varepsilon_n)$  and  $\mathcal{F}_n^e \subset \mathcal{F}_n^\varepsilon$ , the inclusions being proper.

**Lemma 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  $n \geq 2$ ,

$$\begin{aligned} v_{(n)} &= \begin{pmatrix} v_2 \\ v_3 \\ \vdots \\ v_n \end{pmatrix}_{(n-1) \times 1} = \begin{pmatrix} \varepsilon_2 - \bar{\varepsilon}_1 \\ \varepsilon_3 - \bar{\varepsilon}_2 \\ \vdots \\ \varepsilon_n - \bar{\varepsilon}_{n-1} \end{pmatrix}, \\ e_{(n)} &= \begin{pmatrix} e_{n1} \\ e_{n2} \\ \vdots \\ e_{n,n-1} \end{pmatrix}_{(n-1) \times 1} = \begin{pmatrix} \varepsilon_n - \varepsilon_1 \\ \varepsilon_n - \varepsilon_2 \\ \vdots \\ \varepsilon_n - \varepsilon_{n-1} \end{pmatrix}. \end{aligned}$$

By simple calculation,

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

with

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

Since  $C_n$  is regular,

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

(ii) By direct calculation,

$$v_{(n)} = -(C_n, -\mathbf{e}_{n-1}) \varepsilon_{(n)},$$

where  $\varepsilon_{(n)} = (\varepsilon_1, \dots, \varepsilon_n)'$  and  $\mathbf{e}_{n-1}$  denotes the  $(n-1)$ -th unit vector of dimension  $n-1$ . Since

$$(C_n, -\mathbf{e}_{n-1})(C_n, -\mathbf{e}_{n-1})' = \text{diag} \left[ 2, \frac{3}{2}, \dots, \frac{n}{n-1} \right],$$

the assertion follows. □

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

## 4.2 Proof of Theorem 1

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

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

where  $S_n = \sum_{i=1}^n \varepsilon_i$ . Also, Gaussianness is essential for the  $v_n$  to be a MDS with respect to  $(\mathcal{F}_n)$ , as can be seen by considering  $\varepsilon_n$  taking values  $\pm 1$  with probability  $1/2$ .

Remember the definition of  $u_n$  in (3.1):

$$u_n = \sum_{i=1}^n (x_i - \bar{x}_n) \varepsilon_i = \sum_{i=1}^n (x_i - \bar{x}_n) (\varepsilon_i - \bar{\varepsilon}_n). \quad (4.1)$$

We shall make use of the following elementary algebraic identity.

**Lemma 2.** For  $n \geq 2$ ,

$$\sum_{i=1}^n (a_i - \bar{a}_n) (b_i - \bar{b}_n) = \sum_{i=2}^n \frac{i-1}{i} (a_i - \bar{a}_{i-1}) (b_i - \bar{b}_{i-1}). \quad (4.2)$$



Applying this to (4.1), we may write

$$\begin{aligned} u_n &= \sum_{i=2}^n \frac{i-1}{i} (x_i - \bar{x}_{i-1}) (\varepsilon_i - \bar{\varepsilon}_{i-1}) \\ &= \sum_{i=2}^n \frac{i-1}{i} (x_i - \bar{x}_{i-1}) v_i. \end{aligned} \quad (4.3)$$

Or, introducing

$$\xi_i = \sqrt{\frac{i-1}{i}} (x_i - \bar{x}_{i-1}), \quad w_i = \sqrt{\frac{i-1}{i}} v_i,$$

we have the representation

$$u_n = \sum_{i=2}^n \xi_i w_i. \quad (4.4)$$

Note that the  $w_i$  are independent  $\mathcal{N}(0, \sigma^2)$ -distributed and generate the same filtration  $(\mathcal{F}_n)$  as the  $v_i$ . By assumption,  $\xi_i$  is measurable with respect to  $\mathcal{F}_{i-1}$  for  $i \geq 3$ .  $\xi_2 = (x_2 - x_1)/\sqrt{2}$  will generally not be  $\mathcal{F}_1$ -measurable ( $\mathcal{F}_1$  being the trivial  $\sigma$ -algebra). Since the first value  $x_1$  affects neither the asymptotic behavior of  $u_n$  nor that of  $A_n$ , we may as well put  $x_1 = x_2$  and thus achieve full predictability of the process  $(\xi_n)$  with respect to  $(\mathcal{F}_n)$ . Then  $u_n$  is a martingale transform (or local martingale) with predictable quadratic variation

$$\langle u \rangle_n = \sigma^2 \sum_{i=2}^n \xi_i^2 = \sigma^2 A_n. \quad (4.5)$$

For the last step we have again made use of (4.2), now in the form

$$\sum_{i=2}^n \xi_i^2 = \sum_{i=2}^n \frac{i-1}{i} (x_i - \bar{x}_{i-1})^2 = \sum_{i=2}^n (x_i - \bar{x}_n)^2,$$

so that

$$A_n = \sum_{i=2}^n \frac{i-1}{i} (x_i - \bar{x}_{i-1})^2. \quad (4.6)$$

This representation also shows that  $A_n$  is monotone increasing.

From now on let us suppose that the  $w_i$  are of the form

$$w_i = \sigma [W(i) - W(i-1)],$$

where  $W$  is a standard Brownian motion (enlarging the probability space if necessary). Define a continuous time process

$$\xi(s) = \sum_{i=1}^{\infty} \xi_i 1_{(i-1, i]}(s), \quad s \geq 0,$$

where we put  $\xi_1 = \xi_2 = 0$ . Then the stochastic integral

$$M_t = \int_0^t \xi(s) dW(s)$$

is well defined for all  $t \geq 0$ , and is a continuous local martingale with quadratic variation

$$\langle M \rangle_t = \int_0^t \xi(s)^2 ds.$$

Also, it is plain to see that

$$u_n = \sigma M_n$$

and

$$\langle M \rangle_n = \sum_{i=2}^n \xi_i^2 = A_n.$$

If  $A_\infty = \infty$  a.s., we may represent  $M_t$  as

$$M_t = \widetilde{W}(\langle M \rangle_t),$$

where  $\widetilde{W}$  a Brownian motion obtained from  $W$  by the time transformation  $\langle M \rangle_t^{-1}$ . It then follows from the law of the iterated logarithm for Brownian motion that

$$\overline{\lim}_{n \rightarrow \infty} \frac{|M_t|}{\sqrt{2 \langle M \rangle_t \log_2 \langle M \rangle_t}} = 1.$$

For  $u_n$ , this means that

$$\overline{\lim}_{n \rightarrow \infty} \frac{|u_n|}{\sqrt{2 A_n \log_2 A_n}} = \sigma.$$

Or, put differently,

$$\overline{\lim}_{n \rightarrow \infty} \sqrt{\frac{A_n}{2 \log_2 A_n}} \left| \frac{u_n}{A_n} \right| = \sigma. \quad (4.7)$$

Since

$$\widehat{\beta}_n - \beta = \frac{u_n}{A_n}$$

by (1.3) and (4.1), (3.6) is an immediate consequence.  $\square$

### 4.3 Proof of Proposition 1

We shall mimic the proof of Lai and Robbins (1977) for deterministic regressors. If (3.8) holds, then, in particular,  $A_\infty < \infty$  with probability one, such that it follows from (4.5) that

$$u_n \rightarrow u \quad \text{a.s.} \quad (4.8)$$

for some finite random variable  $u$ . Since

$$\sup_{m \geq 0} \mathbf{E} (u_{n+m} - u_n)^2 = \sigma^2 \sup_{m \geq 0} \sum_{i=n+1}^{n+m} \mathbf{E} \xi_i^2 = \sigma^2 (\mathbf{E} A_\infty - \mathbf{E} A_n) \rightarrow 0$$

by monotone convergence, convergence in (4.8) also takes place in  $L^2$ . Hence

$$\mathbf{E} u^2 = \lim_{n \rightarrow \infty} \mathbf{E} u_n^2 = \sigma^2 \lim_{n \rightarrow \infty} \mathbf{E} A_n = \sigma^2 \mathbf{E} A_\infty.$$

Therefore  $u$  must be different from 0 on a set of positive measure, and, a fortiori, this is also true for

$$\lim_{n \rightarrow \infty} (\widehat{\beta}_n - \beta) = \frac{u}{A_\infty}.$$

This completes the proof.  $\square$

**Remark 3.** Coming back to Remark 1 in section 3, if we can find  $w'_i, w''_i$  such that (3.9) holds, then, making use of the representation (4.3), we can write

$$u_n = \sum_{i=2}^n \frac{i-1}{i} w'_i v_i + \sum_{i=2}^n \frac{i-1}{i} w''_i v_i = u'_n + u''_n. \quad (4.9)$$

Hence, denoting the corresponding parts of  $A_n$  by  $A'_n$  and  $A''_n$ , i.e.

$$A'_n = \sum_{i=1}^n \frac{i-1}{i} w_i'^2, \quad A''_n = \sum_{i=1}^n \frac{i-1}{i} w_i''^2 \quad (4.10)$$

(cf. (4.6)), it follows that

$$\frac{u_n}{A_n} = \frac{u'_n}{A_n} + \frac{u''_n}{A_n} = \frac{u'_n}{A'_n} \frac{A'_n}{A_n} + \frac{u''_n}{A_n}.$$

If the  $w'_i, w''_i$  can be chosen in such a way that

$$\frac{u''_n}{A_n} \rightarrow 0 \quad \text{a.s.} \quad (4.11)$$

together with

$$A'_\infty = \infty \quad \text{and} \quad \overline{\lim}_{n \rightarrow \infty} \frac{A'_n}{A_n} < \infty \quad \text{a.s.}, \quad (4.12)$$

then we can apply (3.7) to the  $u'$ -part of  $u$  to obtain

$$\widehat{\beta}_n - \beta \rightarrow 0 \quad \text{a.s.}$$

Note that (4.12) implies that  $A_\infty = \infty$  a.s..

**Remark 4.** The procedure described in Remark 2 is especially promising if the  $x_n$  are linear combinations of the past error terms, i.e.

$$x_n = \sum_{i=1}^{n-1} h_{ni} \varepsilon_i.$$

In this case, they are themselves Gaussian, and condition (3.8) is equivalent to

$$A_\infty < \infty \quad \text{a.s.}$$

cf. Shiryaev (1996, Chapter VII, §6, Lemma p. 533).

## 5 An economic model with forecast feedback

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

$$y_t = \beta y_{t|t-1}^e + \delta' x_t + \varepsilon_t, \quad t = 1, 2, \dots \quad (5.1)$$

where  $y_{t|t-1}^e$  denotes agents' expectations about  $y_t$  based on the information available at time  $t-1$  and the driving variables  $x_t$  are  $p$ -dimensional exogenous random vectors. Models of this type have a long tradition in economics. For instance, the classical cobweb model fits into this form, see e.g. Bray and Savin (1986), as does the Lucas (1973)

aggregate supply model. The basic issue is the way in which the expectations  $y_{t|t-1}^e$  are modelled. The traditional approach to modelling  $y_{t|t-1}^e$  is via *rational expectations*, cf. Muth (1961) or Sargent (2008). There it is assumed that the agents, in building up their expectations at time  $t$ , have complete knowledge of the model and the past  $\mathcal{F}_{t-1} = \sigma(y_s, s \leq t-1; x_s, s \leq t)$  and make best use of it, i.e. form the conditional expectations

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

Taking conditional expectations in (5.1) yields

$$\mathbf{E}(y_t | \mathcal{F}_{t-1}) = \delta' x_t + \beta y_{t|t-1}^e,$$

so that

$$y_{t|t-1}^e = \frac{\delta' x_t}{1 - \beta} = \alpha' x_t,$$

with

$$\alpha = \frac{\delta}{1 - \beta}. \quad (5.2)$$

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

$$y_t = \alpha' x_t + \varepsilon_t. \quad (5.3)$$

Consequently, under the assumption of rational expectations, only  $\alpha$  is identified, not, however,  $\delta$  and  $\beta$  separately.

More recently, however, economic agents are frequently assumed to be boundedly rational and to form their expectations via *adaptive learning*. The basic idea underlying all adaptive learning procedures is that agents employ an auxiliary model, or so-called perceived law of motion (PLM), to form their expectations  $y_{t|t-1}^e$ . One way to specify the PLM is to assume that its functional form corresponds to that of the REE in (5.3). Generally, the agents will not know the parameter  $\alpha$  and therefore replace it by some estimate  $a_{t-1}$ , based on the information  $\mathcal{F}_{t-1}$ . Typically, the parameter  $\alpha$  will be estimated by some recursive procedure. In general, this will have the form of a stochastic approximation algorithm:

$$a_t = a_{t-1} + \gamma_t R_t^{-1} x_t (y_t - a_{t-1}' x_t), \quad (5.4a)$$

$$R_t = R_{t-1} + \gamma_t (x_t x_t' - R_{t-1}), \quad (5.4b)$$

where  $(\gamma_t)$  is a general weighting, or gain, sequence. With this learning scheme, agents' expectation will be given by  $y_{t|t-1}^e = a_{t-1}' x_t$ , and the resulting so-called actual law of motion (ALM), or data generating process (DGP), is

$$y_t = \beta a_{t-1}' x_t + \delta' x_t + \varepsilon_t. \quad (5.4c)$$

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

In the context of (5.4), two issues arise:

- (i) *The internal forecasting problem:* Will the estimated  $a_t$  converge to the REE  $\alpha$ ?

- (ii) *The external estimation problem:* Can the structural parameters  $\beta, \delta$  in (5.4c) be consistently estimated, e.g. by least squares?

As to (i), (5.4a)-(5.4b) is a special case of a general class of recursive algorithms, for which there is a rich literature. A good account of the state of the art can be found in Benveniste, Métivier, and Priouret (1990), Evans and Honkapohja (2001), and Kottmann (1990). The main results are that, for the case of *constant gain* learning, i.e. with  $\gamma_t = \gamma$ , convergence  $a_t \rightarrow \alpha$  will generally not hold. As opposed to that, for *decreasing gain* sequences, i.e. with  $\gamma_t \rightarrow 0$ , convergence  $a_t \rightarrow \alpha$  does hold with probability one under suitable summability assumptions on  $\gamma_t$  and provided that  $\beta < 1$ . For details, see Christopheit and Massmann (2010). If, however,  $\beta \geq 1$ , it can be shown that  $a_t$  diverges.

As to (ii), results are scarce, and pertain mainly to the case of constant gain learning, see Chevillon, Massmann, and Mavroeidis (2010). Since  $a_t \rightarrow \alpha$ , the regressors possess all nice properties of a stationary ergodic sequence so that consistent OLS estimation is possible. On the other hand, if  $a_t \rightarrow \alpha$ , say with probability one, as is the case in decreasing gain learning then the regressors in (5.4c) will be asymptotically collinear

$$y_t \sim \beta \alpha' x_t + \delta x_t + \varepsilon_t,$$

i.e. the asymptotic moment matrix

$$M = \begin{pmatrix} \alpha' \\ I \end{pmatrix} x_t x_t' \begin{pmatrix} \alpha & I \end{pmatrix}$$

will be singular. In the econometrics literature, singularity of  $M$  is generally referred to as absence of strict, or strong, asymptotic identification, see e.g. Davidson and MacKinnon (1993) or Newey and McFadden (1994). As strict asymptotic identification is only sufficient but not necessary for the asymptotic identification of the parameters, the non-singularity of  $M$  need of course not preclude the OLSE from being consistent, and, in fact, does not. There is, indeed, a trade-off between the convergence of agents' expectations to the REE and desirable properties of the OLSE.

The focus of this paper will be on problem (ii) and, in particular, on OLS estimation of the model with the decreasing gain coefficients specified as  $\gamma_t = \gamma/t$  for some constant  $\gamma > 0$ . In view of the discussion of problem (i) above, we restrict ourselves to the case of  $\beta < 1$ . For simplicity and in order not to obscure the main ideas, we will restrict ourselves to the one-dimensional case with  $x_t = 1$  (note that any other constant value of  $x_t$  results in an obvious linear transformation of the condition on  $\gamma$  to be imposed in Theorem 1). Then the dynamics in (5.4) simplify to

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t, \tag{5.5a}$$

$$a_t = a_{t-1} + \frac{\gamma}{t} (y_t - a_{t-1}). \tag{5.5b}$$

Inserting (5.5a) into (5.5b) and introducing

$$c = (1 - \beta) \gamma, \tag{5.6}$$

we obtain

$$a_t = \left(1 - \frac{c}{t}\right) a_{t-1} + \frac{\gamma}{t} (\delta + \varepsilon_t). \tag{5.7}$$

Note that the condition  $\beta < 1$  corresponds to  $c > 0$ .

**Remark 5.** For (5.5b), we have assumed that we start with the stationary value  $R_0 = 1$ . Actually, for any starting value  $R_0$  it will hold that  $\lim_{t \rightarrow \infty} R_t = 1$ .

(5.5a) is a simple linear regression model with predetermined stochastic regressors  $a_{t-1}$  as considered in section 3. Returning for a moment to the notation used there (with index  $t$  instead of  $i$ ), we have that  $x_t = a_{t-1}$  (not to be confused with the  $x$ -variables in (5.4c), which we have assumed to be equal to 1). In particular, since now

$$\begin{aligned}\bar{x}_T &= \frac{1}{T} \sum_{t=1}^T a_{t-1} = \frac{1}{T} \sum_{t=1}^{T-1} a_{t-1} = \frac{T-1}{T} \bar{a}_{T-1}, \\ x_t - \bar{x}_{t-1} &= a_{t-1} - \frac{t-2}{t-1} \bar{a}_{t-2}\end{aligned}$$

(for initial value  $a_0 = 0$ ), the OLSE is given by

$$\hat{\beta}_T - \beta = \frac{u_T}{A_T}, \quad (5.8)$$

with

$$u_T = \sum_{t=1}^T \left( a_{t-1} - \frac{T-1}{T} \bar{a}_{T-1} \right) \varepsilon_t \quad (5.9)$$

and

$$\begin{aligned}A_T &= \sum_{t=1}^T \left( a_{t-1} - \frac{T-1}{T} \bar{a}_{T-1} \right)^2 \\ &= \sum_{i=2}^n \frac{t-1}{t} \left( a_{t-1} - \frac{t-2}{t-1} \bar{a}_{t-2} \right)^2\end{aligned} \quad (5.10)$$

cf. (4.6). As mentioned in the introduction, none of the conditions provided in the literature for strong consistency is satisfied. Actually, considering the moment matrix

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

it is shown in Christopheit and Massmann (2012b) that

$$\overline{\lim}_{T \rightarrow \infty} \frac{\log \lambda_{\max}(M_T)}{\lambda_{\min}(M_T)} = \infty \quad \text{a.s.}$$

and in Christopheit and Massmann (2012a) that

$$\frac{A_T}{\log T} = O_P(1).$$

Hence both (2.6b) and (2.7) are violated.

**Proposition 2.** Consider the model (5.5) and assume that the  $\varepsilon_t$  are independent Gaussian with  $E\varepsilon_i = 0$  and  $E\varepsilon_i^2 = \sigma^2$ . Then, for  $c > 1/2$ , the OLS estimator for  $\beta$  is strongly consistent.

Note that the condition  $c > 1/2$  corresponds to  $\gamma > 1/2(1 - \beta)$ . Christopheit and Massmann (2012a) show that, under this condition, the OLSE for  $\beta$  is asymptotically normal at rate  $\sqrt{\log T}$  even if the  $\varepsilon_t$  are not Gaussian. In particular, this implies *weak* consistency. While the proof in Christopheit and Massmann (2012a) is based on examining second and third moments, the proof of *strong* consistency in section 6 below draws on martingale limit theory. It is, therefore, not clear a priori whether one can do without Gaussianness for strong consistency.

The analysis of the OLSE's behaviour when  $c \leq 1/2$  is left for future research. Preliminary calculations indicate that, if asymptotic normality holds, it should be at rate  $T^{1/2-c}$ , so that convergence would be much faster than when  $c > 1/2$ . The analysis is, however, complicated by the fact that all the terms  $\zeta$ ,  $\eta$  and  $\xi$  into which we decompose  $a$  in section 6.3 below are of the same order of magnitude and thus would have to be considered jointly in the analysis of the limit behavior of  $A_T$ . This seems to be challenging since the  $O(1)$ -terms appearing in the approximations can hardly be qualified any further. Therefore, an approach different from the one adopted here may have to be considered.

## 6 Proof of Proposition 2

### 6.1 Some preliminaries

It is intuitively clear and can, indeed, be easily shown that the choice of initial value  $a_0$  has no influence on the asymptotic behavior. For simplicity of exposition, we shall therefore suppose that  $a_0 = 0$ . It then follows from (5.7) that

$$a_t = \sum_{i=1}^t \phi_{ti} (\delta + \varepsilon_i), \quad (6.1)$$

where

$$\begin{aligned} \phi_{ti} &= \frac{\gamma}{i} \left(1 - \frac{c}{i+1}\right) \cdots \left(1 - \frac{c}{t}\right), \quad i = 1, \dots, t-1, \\ \phi_{tt} &= \frac{\gamma}{t}. \end{aligned} \quad (6.2)$$

In particular, for  $c = 1$ ,

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

for all  $i$ . Taking logarithms and using a second order Taylor expansion, it is shown in Christopheit and Massmann (2010) that

$$\phi_{ti} = \gamma \frac{1}{t^c} \frac{1}{i^{1-c}} \exp \left[ \frac{O_{ti}(1)}{i} \right] = \gamma \frac{1}{t^c} \frac{1}{i^{1-c}} \left[ 1 + \frac{O_{ti}(1)}{i} \right] \quad (6.3)$$

for  $c \neq 1$  and  $i_0 = \max \{i : i \leq c\} \leq i \leq t$ . The  $O_{ti}(1)$  are uniformly bounded in  $t$  and  $i \leq t$ . For  $i < i_0$ ,

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

with

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

Hence, since  $\max_{i < i_0} |\lambda_i| \leq K$ , we have that  $\phi_{ti} = O(t^{-c})$  for  $i < i_0$ . Therefore, denoting by  $\tilde{\phi}_{ti}$  the rightmost expression of (6.3) for a moment (so that  $\tilde{\phi}_{ti} = \phi_{ti}$  for  $i \geq i_0$ ),

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

the same holding for  $\tilde{\phi}_{ti}$ . As a consequence,

$$\sum_{i=1}^t \phi_{ti} = \sum_{i=1}^t \tilde{\phi}_{ti} + \sum_{i=1}^{i_0-1} (\phi_{ti} - \tilde{\phi}_{ti}) = \zeta_t + O(t^{-c}), \quad (6.5)$$

$$\sum_{i=1}^t \phi_{ti} \varepsilon_i = \sum_{i=1}^t \tilde{\phi}_{ti} \varepsilon_i + \sum_{i=1}^{i_0-1} (\phi_{ti} - \tilde{\phi}_{ti}) \varepsilon_i = z_t + O(t^{-c}), \quad (6.6)$$

where we have put

$$\zeta_t = \frac{1}{t^c} \sum_{i=1}^t \frac{1}{i^{1-c}} \left[ 1 + \frac{O_{ti}(1)}{i} \right], \quad (6.7)$$

$$z_t = \frac{1}{t^c} \sum_{i=1}^t \frac{1}{i^{1-c}} \left[ 1 + \frac{O_{ti}(1)}{i} \right] \varepsilon_i. \quad (6.8)$$

(6.1) may be then written in the form

$$a_t = \delta \gamma \zeta_t + \gamma z_t + O(t^{-c}). \quad (6.9)$$

As announced in Remark 3, we shall show in section 6.2 that the centered regressors can be decomposed into two parts:

$$w_t = a_t - \frac{t-1}{t} \bar{a}_{t-1} = w'_t + w''_t, \quad (6.10)$$

of which the first is measurable with respect to  $\mathcal{F}_t = \sigma(\varepsilon_t - \varepsilon_1, \dots, \varepsilon_t - \varepsilon_{t-1})$ , while the second satisfies

$$w''_T = \sum_{t=2}^T \frac{t-1}{t} w''_{t-1} (\varepsilon_t - \bar{\varepsilon}_{t-1}) = O(1) \quad \text{a.s.}, \quad (6.11)$$

see (4.9). Noting that

$$A_T = \sum_{t=2}^T \frac{t-1}{t} (w'_{t-1} + w''_{t-1})^2,$$

(cf. (4.6)) and defining  $A'_T, A''_T$  as in (4.10), we then show in section 6.3 that

$$A''_\infty < \infty \quad \text{and} \quad A_\infty = \infty \quad \text{a.s.} \quad (6.12)$$

It will then follow from the Cauchy-Schwarz inequality that

$$A'_\infty = \infty \quad \text{and} \quad \frac{A'_T}{A_T} = O(1) \quad \text{a.s.}$$

In accordance with Remark 2 in section 4, we can then apply Theorem 1 to the  $w'$ -part, showing that

$$\lim_{T \rightarrow \infty} \hat{\beta}_T = \beta \quad \text{a.s.}$$

as desired.



## 6.2 Proof of (6.10)

We go back to the basic representation (6.1). Then

$$\begin{aligned}
\frac{t-1}{t}\bar{a}_{t-1} &= \frac{1}{t} \sum_{s=1}^{t-1} a_s = \frac{1}{t} \sum_{s=1}^{t-1} \sum_{i=1}^s \phi_{si} (\delta + \varepsilon_i) \\
&= \delta \frac{1}{t} \sum_{i=1}^{t-1} \sum_{s=i}^{t-1} \phi_{si} + \frac{1}{t} \sum_{i=1}^{t-1} \varepsilon_i \sum_{s=i}^{t-1} \phi_{si} \\
&= \delta \frac{1}{t} \sum_{i=1}^{t-1} g_{ti} + \frac{1}{t} \sum_{i=1}^{t-1} g_{ti} \varepsilon_i,
\end{aligned}$$

where we have put

$$g_{ti} = \begin{cases} \sum_{s=i}^{t-1} \phi_{si}, & \text{for } 1 \leq i \leq t-1, \\ 0, & \text{for } i = t. \end{cases} \quad (6.13)$$

Hence

$$\begin{aligned}
w_t &= \delta \left[ \sum_{i=1}^t \phi_{ti} - \frac{1}{t} \sum_{i=1}^{t-1} g_{ti} \right] + \sum_{i=1}^t \phi_{ti} \varepsilon_i - \frac{1}{t} \sum_{i=1}^{t-1} g_{ti} \varepsilon_i \\
&= \delta \sum_{i=1}^t h_{ti} + \sum_{i=1}^t h_{ti} \varepsilon_i,
\end{aligned} \quad (6.14)$$

with

$$h_{ti} = \phi_{ti} - \frac{1}{t} g_{ti} = \left(1 + \frac{1}{t}\right) \phi_{ti} - \frac{1}{t} \sum_{s=i}^t \phi_{si}. \quad (6.15)$$

Write the second term in (6.14) in the form

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

and put

$$w'_t = - \sum_{i=1}^t h_{ti} (\varepsilon_t - \varepsilon_i), \quad (6.16)$$

$$w''_t = (\delta + \varepsilon_t) \mu_t \quad (6.17)$$

with

$$\mu_t = \sum_{i=1}^t h_{ti}.$$

Then (6.10) will hold, and  $(w'_t)$  is clearly adapted to  $(\mathcal{F}_t)$ .

We now have to show that (6.11) holds. To this end, write

$$\begin{aligned}
\mu_t &= \sum_{i=1}^t \phi_{ti} - \frac{1}{t} \sum_{i=1}^{t-1} g_{ti} = \sum_{i=1}^t \phi_{ti} - \frac{1}{t} \sum_{i=1}^{t-1} \sum_{s=i}^{t-1} \phi_{si} \\
&= \sum_{i=1}^t \phi_{ti} - \frac{1}{t} \sum_{s=1}^{t-1} \sum_{i=1}^s \phi_{si} \\
&= \zeta_t - \frac{1}{t} \sum_{s=1}^{t-1} [\zeta_s + O(s^{-c})] + O(t^{-c})
\end{aligned}$$

(cf. (6.5)). By (6.7), making use of the integral comparison theorem,

$$\begin{aligned}
\zeta_t &= \frac{1}{t^c} \sum_{i=1}^t \frac{1}{i^{1-c}} + O(1) \frac{1}{t^c} \sum_{i=1}^t \frac{1}{i^{2-c}} \\
&= \frac{1}{t^c} \left[ \frac{1}{c} t^c + O(1) t^{c-1} + O(1) \right] \\
&= \frac{1}{c} + O(1/t^c) + O(1/t).
\end{aligned} \tag{6.18}$$

Also,

$$\begin{aligned}
\frac{1}{t} \sum_{s=1}^{t-1} \zeta_s &= \frac{1}{c} + \frac{1}{t} \left[ \frac{t^{1-c}}{1-c} + \log t + O(1) \right] \\
&= \frac{1}{c} + O\left(\frac{1}{t^c}\right) + O\left(\frac{\log t}{t}\right),
\end{aligned} \tag{6.19}$$

so that

$$\mu_t = O\left(\frac{1}{t^c}\right) + O\left(\frac{\log t}{t}\right). \tag{6.20}$$

As a consequence,

$$\mathbf{E} [w_t'']^2 = (\delta^2 + \sigma^2) \mu_t^2 = O(1) \left[ \frac{1}{t^{2c}} + \frac{\log^2 t}{t^2} \right]. \tag{6.21}$$

Hence, for  $c > 1/2$ ,

$$\sum_{t=1}^{\infty} \mathbf{E} [w_t'']^2 < \infty.$$

By monotone convergence, the latter implies that

$$\sum_{t=1}^{\infty} (w_t'')^2 < \infty \text{ a.s.} \tag{6.22}$$

for  $c > 1/2$ . Going back to (6.11), write  $u_T''$  in the form

$$u_T'' = \sum_{t=2}^T \frac{t-1}{t} w_{t-1}'' \varepsilon_t - \sum_{t=2}^T \frac{t-1}{t} w_{t-1}'' \bar{\varepsilon}_{t-1}. \tag{6.23}$$

The first term on the right hand side is a martingale transform  $M_T$ , with quadratic variation

$$\langle M \rangle_T = \sigma^2 \sum_{t=2}^T \left( \frac{t-1}{t} \right)^2 (w_{t-1}'')^2. \tag{6.24}$$

By (6.22),  $\langle M \rangle_{\infty} < \infty$  a.s. for  $c > 1/2$ , so that  $M_T$  converges with probability one to some finite random variable. As to the second term in (6.23), call it  $N_T$  for a moment, choose any  $1/2 < \rho < c \wedge 1$ . Then, by (6.20),

$$t^{\rho} \mu_t = O\left(\frac{1}{t^{c-\rho}}\right) + O\left(\frac{\log t}{t^{1-\rho}}\right). \tag{6.25}$$

Also,

$$t^{\rho} \mu_t \varepsilon_t = O(1) \left[ \frac{1}{t^{c-\rho}} + \frac{\log t}{t^{1-\rho}} \right] \varepsilon_t = o(1) \text{ a.s.} \tag{6.26}$$

This follows from the fact that

$$\frac{\varepsilon_t}{t^p} = o(1) \quad \text{a.s.} \quad (6.27)$$

for every  $p > 0$ . (6.27), in turn, can be seen from Chebychev's inequality

$$\mathbf{P} \left( \frac{|\varepsilon_t|}{t^p} > \eta \right) \leq \frac{1}{\eta^m t^{mp}} \mathbf{E} |\varepsilon_t|^m$$

(for every  $\eta > 0$ ), observing that for  $mp > 1$  the right hand side is summable and therefore, by the Borel-Cantelli lemma,

$$\mathbf{P} \left( \frac{|\varepsilon_t|}{t^p} > \eta \text{ i.o.} \right) = 0.$$

But this is equivalent to (6.27). From (6.25) and (6.26), together with (6.17), it follows that

$$t^\rho w_t'' = o(1) \quad \text{a.s.}$$

On the other hand, by the law of the iterated logarithm,

$$\sqrt{\frac{t}{\log_2 t}} \bar{\varepsilon}_t = O(1) \quad \text{a.s.}$$

As a consequence,

$$t^\rho w_t'' \cdot \sqrt{\frac{t}{\log_2 t}} \bar{\varepsilon}_t = o(1) \quad \text{a.s.}$$

and therefore

$$N_T \sim \sum_{t=1}^T \left[ \sqrt{\frac{\log_2 t}{t^{1+2\rho}}} o(1) \right] = o(1) \quad \text{a.s.} \quad (6.28)$$

Summarizing, we find that, for  $c > 1/2$ ,  $u_T''$  is bounded with probability one, thus proving (6.11).

### 6.3 Proof of (6.12)

As before, we shall assume throughout that  $c > 1/2$ . From (6.22), we already know that

$$A_\infty'' < \infty \quad \text{a.s.}$$

In order to show that  $A_\infty = \infty$ , it turns out advantageous to go back to the original definition

$$A_T = \sum_{t=1}^T \left( a_{t-1} - \frac{T-1}{T} \bar{a}_{T-1} \right)^2$$

and use the representation

$$a_t = \delta \gamma \zeta_t + \gamma z_t + O(t^{-c}),$$

see (6.9), with  $\zeta$  and  $z$  given by (6.7) and (6.8). For the further analysis, we split the  $z$ -term into two parts:

$$z_t = \xi_t + \eta_t, \quad (6.29)$$

with

$$\xi_t = \frac{1}{t^c} \sum_{i=1}^t \frac{1}{i^{1-c}} \varepsilon_i, \quad (6.30)$$

and

$$\eta_t = \begin{cases} \frac{1}{t^c} \sum_{i=1}^t \frac{O_{ti}(1)}{i^{2-c}} \varepsilon_i, & \text{for } c \neq 1, \\ 0, & \text{for } c = 1. \end{cases} \quad (6.31)$$

Our procedure will be to consider the various components of  $A_T$  corresponding to the decomposition (6.9) and (6.29) separately. Denoting by  $A_T^{(\zeta)}$ ,  $A_T^{(\eta)}$  and  $A_T^{(\xi)}$  the corresponding parts of  $A_T$ , e.g.

$$A_T^{(\zeta)} = \sum_{t=1}^T \left( \zeta_{t-1} - \frac{T-1}{T} \bar{\zeta}_{T-1} \right)^2$$

and similarly for the other components, it will turn out that  $A_T^{(\zeta)}$  and  $A_T^{(\eta)}$  are  $O(1)$  with probability one, while  $\lim_{T \rightarrow \infty} A_T^{(\xi)} = \infty$  a.s.. From this it follows

$$A_\infty = \infty \quad \text{a.s.} \quad (6.32)$$

### 6.3.1 Ad $\zeta$

By (6.18) and (6.19),

$$\begin{aligned} A_T^{(\zeta)} &= O(1) \sum_{t=1}^T \left[ \frac{1}{t^{2c}} + \frac{1}{t^2} \right] + O(1)T \left[ \frac{1}{T^{2c}} + \frac{\log^2 T}{T^2} \right] \\ &= O(1). \end{aligned} \quad (6.33)$$

### 6.3.2 Ad $\eta$

This term vanishes for  $c = 1$ , so that we may assume  $c \neq 1$ . For  $c \neq 3/2$ , noting that the  $O_{ti}(1)$ -terms are deterministic,

$$\begin{aligned} \mathbf{E}\eta_t^2 &= O(1) \frac{1}{t^{2c}} \sum_{i=1}^t \frac{1}{i^{2(2-c)}} = O(1) \frac{1}{t^{2c}} \left[ \frac{1}{2c-3} t^{2c-3} + O(1) \right] \\ &= O(t^{-3}) + O(t^{-2c}), \end{aligned}$$

while, for  $c = 3/2$ ,

$$\mathbf{E}\eta_t^2 = O(1) \frac{1}{t^3} \sum_{i=1}^t \frac{1}{i} = O\left(\frac{\ln t}{t^3}\right).$$

Therefore, for  $c > 1/2$ ,

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

Since  $(\eta_t)$  is a Gaussian sequence, this is equivalent to

$$\sum_{t=1}^{\infty} \eta_t^2 < \infty \quad \text{a.s.} \quad (6.34)$$

cf. Shiryaev (1996, Chapter VII, §6, Lemma p.533). As to  $\bar{\eta}_T$ , some tedious but straightforward calculation, making repeated use of the integral comparison theorem, shows that

$$\mathbf{E}\bar{\eta}_T^2 = O\left(\frac{1}{T^2} + \frac{1}{T^{2c}} + \frac{1}{T^{1+c}}\right). \quad (6.35)$$

As a consequence,  $T\bar{\eta}_T^2 = o_P(1)$ , so that

$$A_T^{(\eta)} = \sum_{t=1}^T \left( \eta_{t-1} - \frac{T-1}{T} \bar{\eta}_{T-1} \right)^2 \leq 2 \left[ \sum_{t=1}^T \eta_t^2 + T\bar{\eta}_{T-1}^2 \right] = O(1) \text{ a.s.} \quad (6.36)$$

Actually, (6.34) and (6.35) imply only that  $A_T^{(\eta)} = O_P(1)$ . But since  $A_T^{(\eta)}$  is monotone increasing, this is equivalent to a.s. finiteness of the limit.

### 6.3.3 Ad $\xi$

We finally come to the dominating term

$$A_T^{(\xi)} = \sum_{t=1}^T \left( \xi_{t-1} - \frac{T-1}{T} \bar{\xi}_{T-1} \right)^2. \quad (6.37)$$

As to the  $\bar{\xi}$ -term, a similar calculation as for  $\bar{\eta}$  shows that

$$\mathbf{E}\bar{\xi}_T^2 = O(1) \left\{ \frac{\log T}{T^2} + \frac{1}{T^{2c}} + \frac{1}{T^{c+1}} \right\}. \quad (6.38)$$

Hence,  $T\bar{\xi}_T^2 = o_P(1)$ . Therefore, since

$$A_T^{(\xi)} = \sum_{t=1}^T \xi_{t-1}^2 - \frac{(T-1)^2}{T} \bar{\xi}_{T-1}^2, \quad (6.39)$$

it follows that

$$A_T^{(\xi)} = \sum_{t=1}^T \xi_{t-1}^2 + o_P(1). \quad (6.40)$$

If it can be shown that

$$\sum_{t=1}^{\infty} \xi_t^2 = \infty \text{ a.s.}, \quad (6.41)$$

then, since  $A_T^{(\xi)}$  is monotone increasing, it will follow that

$$A_{\infty}^{(\xi)} = \infty \text{ a.s.} \quad (6.42)$$

and hence  $A_{\infty} = \infty$  a.s..

The crucial condition (6.41) has been verified in Christopheit and Massmann (2010) by embedding the discrete time scenarion into a continuous time framework and applying methods from stochastic analysis. A different approach is taken in Christopheit and Massmann (2012a), where it is shown by direct calculation that

$$\alpha_T^{-2} \sum_{t=1}^T \xi_t^2 \xrightarrow{L^2} 1,$$

the sequence  $\alpha_T$  being given by

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

As a byproduct, Christopeit and Massmann (2012a) also prove the asymptotic normality of the OLSE at rate  $\sqrt{\log T}$ . Here, we give a new proof which makes explicit use of the Gaussianness of the error terms. It is based on the theorem in Shiryaev (1996) cited above and on a zero-one law.

### 6.3.4 Proof of (6.41).

Denote

$$X_T = \sum_{t=1}^T \xi_t^2 \quad \text{and} \quad X_{sT} = \sum_{t=s}^T \xi_t^2.$$

For fixed  $s \geq 2$ , decompose  $\xi_t, t \geq s$ , in the form

$$\xi_t = \frac{1}{t^c} [w_s + w_{st}],$$

with

$$w_s = \sum_{i=1}^{s-1} \frac{1}{i^{1-c}} \varepsilon_i \quad \text{and} \quad w_{st} = \sum_{i=s}^t \frac{1}{i^{1-c}} \varepsilon_i.$$

Then, for  $T \geq s$ ,

$$X_{sT} = w_s^2 \sum_{t=s}^T \frac{1}{t^{2c}} + 2w_s \sum_{t=s}^T \frac{1}{t^{2c}} w_{st} + \sum_{t=s}^T \frac{1}{t^{2c}} w_{st}^2. \quad (6.43)$$

Denote

$$\eta_{sT} = \sum_{t=s}^T \frac{1}{t^{2c}} w_{st}, \quad X'_{sT} = \sum_{t=s}^T \frac{1}{t^{2c}} w_{st}^2.$$

Since, for  $t \geq s+1$ ,

$$w_{st} = w_{s+1,t} + \frac{\varepsilon_s}{s^{1-c}},$$

it follows that, for  $\eta_{sT}$  for  $T \geq s+1$ ,

$$\eta_{sT} = \eta_{s+1,T} + \rho_{sT},$$

where

$$\rho_{sT} = \frac{\varepsilon_s}{s^{1-c}} \left[ 1 + \sum_{t=s+1}^T \frac{1}{t^{2c}} \right].$$

Since, for all  $s$ ,  $\sup_{T \geq s} |\rho_{sT}(\omega)| < \infty$  for all  $\omega$ , the set  $\Gamma_0 = \{\sup_{T \geq s} |\eta_{sT}| < \infty\}$  is the same for all  $s$ , and coincides with the set  $\{\sup_{T \geq s} |w_s \eta_{sT}| < \infty\}$ . Since the first term on the right hand side of (6.43) is uniformly bounded in  $T$  for all  $s$ , and apparently  $\Gamma_s = \{X_{s\infty} = \infty\} = \Gamma = \{X_\infty = \infty\}$  for all  $s$ , it follows a fortiori that the sets  $\Gamma'_s = \{X'_{s\infty} = \infty\}$  coincide for all  $s$ :  $\Gamma'_s = \Gamma'$ . Since  $X'_{sT}$  involves only  $\varepsilon_t, t \geq s$ , so does  $X'_{s\infty}$ . Hence  $\Gamma' \in \bigcap_{s=1}^{\infty} \sigma(\varepsilon_s, \varepsilon_{s+1}, \dots)$ . In other words,  $\Gamma'$  is a tail event. By

Kolmogorov's zero-one law, either  $\mathbf{P}(\Gamma') = 0$  or  $\mathbf{P}(\Gamma') = 1$ . Since  $\Gamma \cap \Gamma_0 \subset \Gamma'$ , it holds that

$$\mathbf{P}(\Gamma \cap \Gamma_0) \leq \mathbf{P}(\Gamma'). \quad (6.44)$$

We will show that

$$\mathbf{P}(\Gamma_0) = 1 \quad (6.45)$$

and

$$\mathbf{P}(\Gamma) > 0. \quad (6.46)$$

Then it will follow from (6.44) that  $\mathbf{P}(\Gamma') = 1$  and hence  $\mathbf{P}(\Gamma) = 1$ .

**Ad (6.45)** Consider

$$\begin{aligned} \eta_{sT} &= \sum_{t=s}^T \frac{1}{t^{2c}} \sum_{i=s}^t \frac{\varepsilon_i}{i^{1-c}} = \sum_{i=s}^T \frac{\varepsilon_i}{i^{1-c}} \sum_{t=i}^T \frac{1}{t^{2c}} \\ &= \frac{1}{1-2c} \sum_{i=s}^T \frac{\varepsilon_i}{i^{1-c}} \left[ \frac{1}{T^{2c-1}} - \frac{1}{i^{2c-1}} + O_{Ti}(1)i^{-2c} \right] \\ &= \frac{1}{1-2c} \left[ \frac{1}{T^{2c-1}} \sum_{i=s}^T \frac{\varepsilon_i}{i^{1-c}} - \sum_{i=s}^T \frac{\varepsilon_i}{i^c} + \sum_{i=s}^T \frac{O_{Ti}(1)\varepsilon_i}{i^{1+c}} \right]. \end{aligned}$$

The (deterministic)  $O_{Ti}(1)$ -terms are uniformly bounded in  $T, i$ . By Kolmogorov's theorem, the second term in brackets converges with probability one to some finite random variable as  $T \rightarrow \infty$ . As to the third term, call it  $v_T$  for a moment, write

$$|v_T| \leq O(1) \sum_{i=s}^T \frac{|\varepsilon_i| - m}{i^{1+c}} + O(1),$$

with  $m = \mathbf{E}|\varepsilon_i|$ . To the first term on the right hand side, again Kolmogorov's theorem applies to show that it converges with probability one to some finite random variable as  $T \rightarrow \infty$ . Therefore,  $v_T = O(1)$  with probability one. As to the first term,

$$M_T = \sum_{i=s}^T \frac{\varepsilon_i}{i^{1-c}}$$

is a martingale with quadratic variation

$$\langle M \rangle_T = \sigma^2 \sum_{i=s}^T \frac{1}{i^{2(1-c)}} = \frac{\sigma^2}{2c-1} T^{2c-1} + O(1).$$

Hence, by the standard martingale convergence theorem, the first term tend to zero with probability one. Summarizing, we find that

$$\eta_{sT} = O(1) \text{ a.s.},$$

which shows (6.45).

**Ad (6.46)** For this, we evoke again the result already cited in Remark 4, namely the equivalence

$$\mathbf{P}(X_\infty < \infty) = 1 \Leftrightarrow \mathbf{E}X_\infty < \infty.$$

But

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

so that

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

Therefore,  $\mathbf{E}X_\infty = \infty$ , implying that  $\mathbf{P}(X_\infty < \infty) < 1$ . Equivalently,  $\mathbf{P}(\Gamma) > 0$ .  $\square$

## References

- Anderson, T. W. and J. B. Taylor (1976). Strong consistency of least squares estimates in normal linear regression. *Annals of Statistics* 4, 788–790.
- Anderson, T. W. and J. B. Taylor (1979). Strong consistency of least squares estimates in dynamic models. *Annals of Statistics* 7, 484–489.
- Benveniste, A., M. Métivier, and P. Priouret (1990). *Adaptive Algorithms and Stochastic Approximation*. Berlin: Springer. Originally 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.
- Christopeit, N. (1986). Starke Konsistenz von Regressionsschätzern. In W. Krelle (Ed.), *Ökonomische Prognose-, Entscheidungs- und Gleichgewichtsmodelle*, pp. 262–272. Weinheim: VCH-Verlagsgesellschaft.
- Christopeit, N. and K. Helmes (1980). Strong consistency of least squares estimators in linear regression models. *Annals of Statistics* 8, 778–788.
- Christopeit, N. and M. Massmann (2010). Consistent estimation of structural parameters in regression models with adaptive learning. TI discussion paper no. 10-077/4.
- Christopeit, N. and M. Massmann (2012a). Asymptotic normality of ordinary least squares estimators in regression models with adaptive learning. Mimeo.
- Christopeit, N. and M. Massmann (2012b). A note on an estimation problem in models with adaptive learning. Mimeo.
- Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.



- 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. and S. Honkapohja (2001). *Learning and Expectations in Macroeconomics*. Princeton: Princeton University Press.
- Kottmann, T. (1990). *Learning Procedures and Rational Expectations in Linear Models with Forecast Feedback*. Ph. D. thesis, University of Bonn.
- Lai, T. L. (2003). Stochastic approximation. *Annals of Statistics* 31, 391–406.
- 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 of 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). 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 (1982b). Asymptotic properties of projections with applications to stochastic regression problems. *Journal of Multivariate Analysis* 12, 346–370.
- Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *American Economic Review* 63, 326–334.
- Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica* 29, 315–335.
- 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.
- Sargent, T. J. (2008). Rational expectations. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (2nd ed.). Palgrave Macmillan. 1st edition 1987.
- Shiryaev, A. N. (1996). *Probability* (2nd ed.). New York: Springer. 1st edition 1984.