Online Supplement for *Estimating structural parameters in regression models with adaptive learning*

Norbert Christopeit
University of Bonn

Michael Massmann[a]
WHU – Otto Beisheim School of Management
and
Vrije Universiteit Amsterdam

November 30, 2016

Contents

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

C Time-varying regressors

The assumption in this paper that the exogenous regressor x_t is constant serves the purpose of analytical tractability, especially in order to facilitate the examination of the asymptotic behaviour of a_t, at least to such an extent as is needed for the treatment of the EEP. Upon inspecting the proofs, however, it is apparent that time-varying regressors x_t are likely to lead to complications in the analysis. In particular, in the case of constant gain learning, the analogue of (2.2) for general time-varying x_t is

$$a_t = \left[1 - c\frac{x_t^2}{r_t}\right] a_{t-1} + c\frac{x_t^2}{r_t} + \gamma\frac{x_t}{r_t}\varepsilon_t$$

such that x_t may cause the autoregressive coefficient to switch between the stable, unit root or explosive regimes. This issue will not arise in the case of decreasing gain learning as long as a_t converges to the REE α. On the other hand, the singularity of the asymptotic second moment matrix will persist, see equation (1.6). In order derive substantial results in these settings, strong assumptions will have be imposed on the regressors. One such restriction is the case of the regressors tending to an equilibrium value which, not surprisingly, leads to essentially the same results as in Theorems 1-4.

Assumption E
The sequence x_t tends to an equilibrium value x: $\lim_{t\to\infty} x_t = x$.

Without loss of generality, we may again assume that $x = 1$. The x_t are taken to be deterministic for expositional simplicity. Identical calculations to those below would result for stochastic regressors if, for instance, *(i)* the regressors are strictly exogenous, i.e. the sequence x_t is independent of the error terms ε_t, and *(ii)* Assumption E holds with probability one.

C.1 Constant gain

Reconsider the recursion of r_t in (1.4b) with a constant gain $\gamma_t = \gamma$:

$$r_t = (1-\gamma)\, r_{t-1} + \gamma x_t^2. \tag{C.1}$$

With the solution of (C.1) given by

$$r_t = \rho^t r_0 + \gamma \sum_{n=0}^{t-1} \rho^n x_{t-n}^2$$

it follows that r_t tends to the equilibrium value $r = x^2 = 1$, provided that $\gamma \in (0,1)$. Substituting this into the dynamics of a_t in (1.4a) yields the recursion in (2.2), as indeed was obtained under the assumption of a constant $x_t = x$. As a consequence, the asymptotics of a_t are also the same.

Regarding the EEP, note that the structural equation is given by

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

or

$$y_t^x = \delta_t + \beta a_{t-1} + \varepsilon_t^x,$$

with $y_t^x = y_t/x_t$ and $\varepsilon_t^x = \varepsilon_t/x_t$. Remembering that $x_t \to 1$, it can be shown that passing from ε_t^x to ε_t does not affect the behaviour of the OLS estimator, so that all the results for the EEP in the case of constant gain remain valid.

C.2 Decreasing gain

Reconsider the recursion of r_t in (1.4b) with a decreasing gain sequence $\gamma_t = \gamma/t$, i.e.

$$r_t = \left(1 - \frac{\gamma}{t}\right) r_{t-1} + \frac{\gamma}{t} x_t^2.$$

This is of the same form as equation (B.1). Hence, performing the same analysis on r_t as is done on a_t in Appendices B.1.1 and B.1.2 shows that, for every $\gamma > 0$,

$$\lim_{t\to\infty} r_t = r\gamma x^2 = r\gamma$$

for some positive number r. Using this equilibrium value in the dynamics for a_t in (1.4a) we obtain

$$a_t = a_{t-1} + \frac{1/r}{t}\left(y_t - a_{t-1}\right),$$

which is just (1.9) with $\gamma_t = \widetilde{\gamma}/t$ and $\widetilde{\gamma} = 1/r$. Note, however, that in order to determine $\widetilde{\gamma}$ and, correspondingly, the value of $\widetilde{c} = (1 - \widetilde{\gamma})\,\beta$, one has to know r. This, however, is given by

$$r = \lim_{t\to\infty} \frac{1}{t^\gamma} \sum_{i=1}^{t} \frac{\theta_i}{i^{1-\gamma}},$$

cf. (B.14) and (B.15). Since $\theta_i \to 1$, it is clear that

$$r = \lim_{t\to\infty} \frac{1}{t^\gamma} \sum_{i=1}^{t} \frac{1}{i^{1-\gamma}} = \lim_{t\to\infty} \frac{1}{t^\gamma}\left[\int_1^t \frac{ds}{s^{1-\gamma}} + O(1)\right] = \frac{1}{\gamma}.$$

Hence, $\widetilde{\gamma} = \gamma$ and $\widetilde{c} = c$. As a consequence, up to a change in variance, we have the same asymptotics for a_t as for $x_t = 1$. The same is true for the EEP.

D Consistency

The weak consistency of the OLS estimator in Sections 2 and 3 is obtained as a byproduct of our results in Theorems 2 and 4. It is instructive, however, to look at our results in the light of the results available in the literature on consistency in models with predetermined regressors. The reason is that even the best of those conditions turn out not to be met by some of the constant and decreasing gain learning models we consider in this paper. This finding complements the failure of the Grenander condition for the decreasing gain model in Section 3, see also the discussion in the introduction.

To our knowledge, the best sufficient condition for the consistency of the OLS estimator in multivariate models with predetermined regressors is given in Lai & Wei (1982*a*). It requires that

$$\lambda_{\min}(T) \to \infty \quad \text{and} \quad \frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} \to 0 \quad \text{a.s.,} \tag{D.1}$$

where $\lambda_{\max}(T)$ and $\lambda_{\min}(T)$ are the maximal and minimal eigenvalue, respectively, of the regressors' moment matrix M_T. For the estimation of the slope parameter in a simple regression model, a slight improvement is given in Lai & Wei (1982*b*) with the condition

$$\frac{A_T}{\ln T} \to \infty \quad \text{a.s.,} \tag{D.2}$$

with A_T being the usual sum of squared mean-adjusted regressors. To illustrate the strength of (D.1), Lai & Wei (1982*a*) discuss an example in which a marginal violation of the conditions leads to the inconsistency of the OLS estimator. They hence call the conditions in (D.1) "in some sense the weakest possible" (p. 155).

For the purpose of comparing (D.1) to our results on weak consistency, note that this condition may also be used in terms of convergence in probability, in the sense that

$$\frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} \xrightarrow{p} 0 \tag{D.3}$$

implies the weak consistency of the OLS estimator, say $\widehat{\theta}_T$. This is because the basic result obtained by Lai & Wei (1982*a*) is that

$$\left\| \widehat{\theta}_T - \theta \right\|^2 = \frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} O(1) \quad \text{a.s.}$$

on the set $\left\{ \lambda_{\min}(T) > 0 \right\}$. Let us briefly discuss condition (D.3) for the various models considered in this paper.

D.1 Constant gain

Reconsider the model in (2.1)-(2.2). For the stable case, (D.3) is trivially satisfied since all entries of M_T in (2.5) satisfy a weak LLN. The same is true for the unit root case, as can be shown by some straightforward calculations on the eigenvalues, using the asymptotic behaviour of the properly normalised entries of M_T as obtained in Appendix A.3.2. For the explosive case, similar calculations making use of Theorem 1 *(iii)* show that

$$\frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} \to 4 \ln|1-c| \quad \text{a.s..}$$

Hence (D.3) is violated, but weak consistency still holds.

D.2 Decreasing gain

Turn now to model (3.1)-(3.2). For $c < 1/2$, it can be verified that condition (D.3) is met. For $c > 1/2$, however, it is shown in Appendix B.3 that

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

Hence (D.2) is not satisfied. Also, Christopeit & Massmann (2013) conclude that

$$\operatorname*{plim}_{T \to \infty} \frac{\ln \lambda_{\max}(T)}{\lambda_{\min}(T)} = \left(\alpha^2 + 1\right) \frac{2c-1}{\sigma^2 \gamma^2}$$

so that (D.3) is not satisfied either. Nevertheless, Theorem 4 implies that the slope estimator is weakly consistent.

E The Lindeberg conditions

E.1 Theorem 3 for $c \geq 1/2$

We verify the Lindeberg condition for sums of independent random variables, cf. Shiryaev (1996, Chapter III, §4, Theorem 1). Put differently, for every $\delta > 0$,

$$V_t = \frac{1}{\langle v \rangle_t} \sum_{i=1}^{t} \mathbf{E} \frac{\varepsilon_i^2}{i^{2(1-c)}} 1_{\left\{|\varepsilon_i| > \delta i^{1-c} \langle v \rangle_t^{1/2}\right\}} \to 0.$$

For $c > 1/2$, taking account of (B.17),

$$\begin{aligned} \left\{|\varepsilon_i| > \delta i^{1-c} \langle v \rangle_t^{1/2}\right\} &= \left\{|\varepsilon_i| > \frac{\sigma}{\sqrt{2c-1}} \delta i^{1-c} \sqrt{t^{2c-1} + O(1)}\right\} \\ &= \left\{|\varepsilon_i| > \frac{\sigma}{\sqrt{2c-1}} \big(1 + o(1)\big) \delta i^{1-c} t^{c-1/2}\right\} \\ &\subset \left\{|\varepsilon_i| > \kappa \big(1 + o(1)\big) t^p\right\} \end{aligned}$$

with $p = (c \wedge 1) - \frac{1}{2}$ and $\kappa > 0$. The last inclusion follows from the fact that $i^{1-c} \geq t^{1-c}$ for $c \geq 1$ and $i^{1-c} \geq 1$ for $c < 1$. Therefore, by square integrability of ε_i,

$$\mathbf{E}\varepsilon_i^2 1_{\left\{|\varepsilon_i| > \delta i^{1-c} \langle v \rangle_t^{1/2}\right\}} \leq \mathbf{E}\varepsilon_1^2 1_{\left\{|\varepsilon_1| > \kappa \left(1 + o(1)\right) t^p\right\}} = \pi_t \to 0$$

as $t \to \infty$. As a consequence,

$$V_t \leq \frac{\pi_t}{\langle v \rangle_t} \sum_{i=1}^{t} \frac{1}{i^{2(1-c)}} = \frac{\pi_t}{\sigma^2} \to 0.$$

For $c = 1/2$, the proof runs similarly, now making use of (B.22):

$$\begin{aligned} \left\{|\varepsilon_i| > \delta i^{1/2} \langle v \rangle_t^{1/2}\right\} &= \left\{|\varepsilon_i| > \sigma \big(1 + o(1)\big) \delta i^{1/2} \sqrt{\ln t}\right\} \\ &\subset \left\{|\varepsilon_i| > \kappa \big(1 + o(1)\big) \sqrt{\ln t}\right\}, \end{aligned}$$

so that

$$\mathbf{E}\varepsilon_i^2 1_{\left\{|\varepsilon_i| > \delta i^{1-c} \langle v \rangle_t^{1/2}\right\}} \leq \mathbf{E}\varepsilon_1^2 1_{\left\{|\varepsilon_1| > \kappa \left(1 + o(1)\right) \sqrt{\ln t}\right\}} = \pi_t \to 0$$

and hence

$$V_t = \frac{1}{\langle v \rangle_t} \sum_{i=1}^{t} \mathbf{E} \frac{\varepsilon_i^2}{i} 1_{\left\{|\varepsilon_i| > \delta i^{1/2} \langle v \rangle_t^{1/2}\right\}} \leq \frac{\pi_t}{\langle v \rangle_t} \sum_{i=1}^{t} \frac{1}{i} = \frac{\pi_t}{\sigma^2} \to 0.$$

E.2 Theorem 4 for $c > 1/2$

Reconsider the martingale in (B.38), reproduced here for convenience:

$$M_T = \sum_{t=1}^{T} \xi_{Tt} \varepsilon_t, \quad \xi_{Tt} = \frac{a_{t-1}}{\sqrt{\alpha_T}}.$$

We have to show that, for every $\delta > 0$,

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

cf. Christopeit & Hoderlein (2006). To this end, we make use of the elementary implication $|ab| > \delta \Rightarrow a^2 > \delta$ or $b^2 > \delta$ to obtain the inclusion $\left\{|\xi_{Tt}\varepsilon_t| > \delta\right\} = \left\{|a_{t-1}\varepsilon_t| > \delta\sqrt{\alpha_T}\right\} \subset \left\{a_{t-1}^2 > \delta\sqrt{\alpha_T}\right\} \cup \left\{\varepsilon_t^2 > \delta\sqrt{\alpha_T}\right\}$. Therefore,

$$\begin{aligned} R_T &\leq \frac{1}{\alpha_T}\sum_{t=1}^{T} \mathbf{E}\left\{a_{t-1}^2 \varepsilon_t^2 1_{\left\{a_{t-1}^2 > \delta\sqrt{\alpha_T}\right\}} | \mathcal{F}_{t-1}\right\} \\ &\quad + \frac{1}{\alpha_T}\sum_{t=1}^{T} \mathbf{E}\left\{a_{t-1}^2 \varepsilon_t^2 1_{\left\{\varepsilon_t^2 > \delta\sqrt{\alpha_T}\right\}} | \mathcal{F}_{t-1}\right\} \\ &= \frac{\sigma^2}{\alpha_T}\sum_{t=1}^{T} a_{t-1}^2 1_{\left\{a_{t-1}^2 > \delta\sqrt{\alpha_T}\right\}} + \frac{1}{\alpha_T}\sum_{t=1}^{T} a_{t-1}^2 \mathbf{E}\left\{\varepsilon_t^2 1_{\left\{\varepsilon_t^2 > \delta\sqrt{\alpha_T}\right\}}\right\} \\ &= R_T^0 + R_T^1. \end{aligned}$$

As to R_T^0, since $a_t \to \alpha$ a.s., there will be a T_0 (depending on ω) such that $a_{t-1}^2 \leq \delta\sqrt{\alpha_T}$ for all $t > T_0$. Hence the sum is finite and

$$R_T^0 \to 0 \quad \text{a.s.} \tag{E.2}$$

As to R_T^1,

$$\mathbf{E}\left\{\varepsilon_t^2 1_{\left\{\varepsilon_t^2 > \lambda_T\delta\right\}}\right\} = \pi_T \to 0.$$

Hence, taking account of (B.42),

$$R_T^1 = \frac{\pi_T}{\alpha_T}\sum_{t=1}^{T} a_{t-1}^2 = \pi_T \frac{A_T'}{\alpha_T} \xrightarrow{p} 0. \tag{E.3}$$

(E.2) and (E.3) together show (E.1).

E.3 Theorem 4 for $c < 1/2$

By definition (cf. (B.68)),

$$X_T = \sum_{t=1}^{T} \xi_{Tt}\varepsilon_t$$

with

$$\xi_{Tt} = \frac{1}{T^{1/2-c}}\left(t^{-c} - \frac{T^{-c}}{1-c}\right)$$

To show:

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

But

$$R_T = \sum_{t=1}^{T} \xi_{Tt}^2 \mathbf{E}\left\{\varepsilon_t^2 1_{\left\{\varepsilon_t^2 > \delta^2/\xi_{Tt}^2\right\}}\right\}.$$

Since

$$\max_{t \leq T} |\xi_{Tt}| \leq \frac{1}{T^{1/2-c}} + \frac{1}{1-c}\frac{1}{T^{1/2}} = m_T = o(1),$$

it follows that

$$\pi_T = \mathbf{E}\left\{\varepsilon_t^2 1_{\{\varepsilon_t^2 > \delta^2/\xi_{Tt}^2\}}\right\} \leq \mathbf{E}\left\{\varepsilon_t^2 1_{\{\varepsilon_t^2 > \delta^2/m_T^2\}}\right\} \to 0.$$

Therefore,

$$R_T \leq m_T \sum_{t=1}^{T} \xi_{Tt}^2 \to 0$$

since $\sum_{t=1}^{T} \xi_{Tt}^2 = O(1)$.

F Proof of Corollary 3

Consider the OLS residual $\widehat{\varepsilon}_t = y_t - \widehat{\delta} - \widehat{\beta} a_{t-1} = m_t + \varepsilon_t$, where

$$m_t = (\delta - \widehat{\delta}) + (\beta - \widehat{\beta}) a_{t-1}.$$

Then

$$\sum_{t=1}^{T} \widehat{\varepsilon}_t^2 = \sum_{t=1}^{T} m_t^2 + 2\sum_{t=1}^{T} m_t \varepsilon_t + \sum_{t=1}^{T} \varepsilon_t^2.$$

Since

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

it follows that

$$\frac{1}{T}\sum_{t=1}^{T} \widehat{\varepsilon}_t^2 = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2 + o(1) \to \sigma^2$$

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

G Proof of equation (B.77)

Ad R. Recall the definition of R_T in (B.76):

$$R_T = \frac{1}{T^{1/2-c}} \sum_{t=1}^{T} t^{-c} \zeta_t \varepsilon_t.$$

With a view to deriving $\mathbf{E}R_T^2$, we calculate

$$\begin{aligned}
\mathbf{E}\left[\sum_{t=1}^{T} t^{-c}\zeta_t\varepsilon_t\right]^2 &= \mathbf{E}\sum_{s,t=1}^{T} t^{-c}\zeta_t\varepsilon_t s^{-c}\zeta_s\varepsilon_s \\
&= 2\mathbf{E}\sum_{t=1}^{T} t^{-c}\zeta_t\varepsilon_t \sum_{s=1}^{t-1} s^{-c}\zeta_s\varepsilon_s + \mathbf{E}\sum_{t=1}^{T} t^{-2c}\zeta_t^2\varepsilon_t^2 \\
&= 2\sum_{t=1}^{T}\sum_{s=1}^{t-1} t^{-c}s^{-c}\mathbf{E}\zeta_t\varepsilon_t\zeta_s\varepsilon_s + \mathbf{E}\sum_{t=1}^{T} t^{-2c}\zeta_t^2\varepsilon_t^2 \\
&= R_{1T} + R_{2T}. \qquad \text{(G.1)}
\end{aligned}$$

As to R_{1T}, making use of (B.74), we obtain for $s < t$ that

$$\begin{aligned}
\mathbf{E}\zeta_t\varepsilon_t\zeta_s\varepsilon_s &= \mathbf{E}\left\{\left[\sum_{i=t}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_t\left[\sum_{i=s}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_s\right\} \\
&= \mathbf{E}\left\{\left[\sum_{i=t}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]^2\varepsilon_t\varepsilon_s\right\} \\
&\quad +\mathbf{E}\left\{\left[\sum_{i=t}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_t\left[\sum_{i=s}^{t-1}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_s\right\} \\
&= \mathbf{E}\left\{\left[\sum_{i=t}^{\infty}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_t\right\}\mathbf{E}\left\{\left[\sum_{i=s}^{t-1}\theta_i\frac{\varepsilon_i}{i^{1-c}}\right]\varepsilon_s\right\} \\
&= \sigma^4\theta_t\theta_s\frac{1}{t^{1-c}}\frac{1}{s^{1-c}}.
\end{aligned}$$

Hence, remembering that $\lim_{t\to\infty}\theta_t = 1$,

$$\begin{aligned}
R_{1T} &= 2\sigma^4\sum_{t=1}^{T} t^{-c}\theta_t\frac{1}{t^{1-c}}\sum_{s=1}^{t-1} s^{-c}\theta_s\frac{1}{s^{1-c}} \\
&= 2\sigma^4\sum_{t=1}^{T}\theta_t\frac{1}{t}\sum_{s=1}^{t-1}\theta_s\frac{1}{s} \\
&= O(1)\sum_{t=1}^{T}\frac{1}{t}\left[\ln t + O(1)\right] \\
&= O(1)\ln^2 T. \qquad \text{(G.2)}
\end{aligned}$$

As to R_{2T},

$$\begin{aligned}
\mathbf{E}\zeta_t^2\varepsilon_t^2 &= \mathbf{E}\left[\zeta_{t+1} + \theta_t \frac{\varepsilon_t}{t^{1-c}}\right]^2 \varepsilon_t^2 \\
&= \mathbf{E}\zeta_{t+1}^2 \mathbf{E}\varepsilon_t^2 + \frac{\theta_t^2}{t^{2(1-c)}} \mathbf{E}\varepsilon_t^4 \\
&= \sigma^4 \sum_{i=t+1}^{\infty} \frac{\theta_i^2}{t^{2(1-c)}} + \frac{\theta_t^2}{t^{2(1-c)}} m_4 \\
&= O\left(t^{2c-1}\right).
\end{aligned}$$

Hence

$$R_{2T} = \sum_{t=1}^{T} t^{-2c} \mathbf{E}\zeta_t^2 \varepsilon_t^2 = O(1) \sum_{t=1}^{T} \frac{1}{t} = O(1) \ln T. \tag{G.3}$$

Therefore, taking (G.1), (G.2) and (G.3) together, we find that

$$\mathbf{E}(R_T)^2 = \frac{1}{T^{1-2c}} [R_{1T} + R_{2T}] = O\left(\frac{\ln^2 T}{T^{1-2c}}\right).$$

In particular,

$$\operatorname*{plim}_{T\to\infty} R_T = 0. \tag{G.4}$$

Ad S. Recall S_T in (B.76), namely

$$S_T = \frac{1}{T^{1/2-c}} \sum_{t=1}^{T} t^{-c} w_{t-1} \varepsilon_t.$$

Since

$$w_{t-1} = \frac{1}{t} \sum_{i=1}^{t-1} \frac{O_{ti}(1)}{i^{1-c}} \varepsilon_i, \tag{G.5}$$

$$\mathbf{E}w_{t-1}^2 = O(1) \frac{1}{t^2} \sum_{i=1}^{t} \frac{1}{i^{2(1-c)}} = O\left(t^{-2}\right), \tag{G.6}$$

(cf. (B.25)) is $\mathcal{F}_{t-1}$-measurable and

$$\begin{aligned}
\mathbf{E}S_T^2 &= \frac{\sigma^2}{T^{1-2c}} \sum_{t=1}^{T} t^{-2c} \mathbf{E}w_{t-1}^2 \\
&= O(1) \frac{1}{T^{1-2c}} \sum_{t=1}^{T} \frac{1}{t^{2(1+c)}} = O\left(\frac{1}{T^{1-2c}}\right).
\end{aligned}$$

In particular,

$$\operatorname*{plim}_{T\to\infty} S_T = 0. \tag{G.7}$$

H On Remark 1

The following considerations are based on a Theorem by Lévy, cf. Kawata (1972, Theorem 13.1.1):

Theorem 1 (Lévy)
Let $X_n, n = 0, 1, \ldots,$ be a sequence of random variables with distribution functions F_k s.t. the infinite sum

$$X = \sum_{n=0}^{\infty} X_n \tag{H.1}$$

converges absolutely with probability one. Let p_n denote the maximal jump of F_n, i.e. $p_n = \sup\left\{F_n(x) - F_n(x-) : x \in \mathbb{R}\right\}$, and F the distribution function of X. Then the following is true.

(i) If one of the F_n is continuous, then also the distribution function F is continuous.

(ii) If all F_n have discontinuities, then a necessary and sufficient condition for F to be continuous is that

$$P = \prod\nolimits_{n=0}^{\infty} p_n$$

diverges to zero.

We will apply this result to the sequence $X_n = a_n \varepsilon_n$, where the ε_n are iid with finite second moments and distribution function F_0, and the sequence $(a_n) \in l^2$. The rhs of (H.1) then converges a.s. (by Kolmogorov's theorem) and in L^2. If F_0 is continuous, then so are the distribution functions $F_n(x) = F_0\left(a^{-n}x\right)$ of the X_n, and *(i)* shows that F is continuous. If the points of discontinuity of F_0 are (x_i), with $p_0 = \sup_i \left\{F_0(x_i) - F_0(x_i-)\right\}$, then the points of discontinuity of F_n are $\left(a^{-n}x_i\right)$, and the height of the jump of F_n at $a^{-n}x_i$ is just that of F_0 at x_i. Hence $p_n = p_0 = \sup_i \left\{F_0(x_i) - F_0(x_i-)\right\}$. Hence, whenever $p_0 < 1$,

$$\prod\nolimits_{n=0}^{N} p_n = p_0^{N+1} \to 0,$$

so that continuity of F follows from *(ii)*. The case $p_0 = 1$ cannot occur unless $\varepsilon_n = 0$.

This result covers all our needs for stable AR(1)-processes with *iid integrable error terms*. It shows that the stationary distribution F is *always continuous,* except for the trivial case of zero errors (in which case the stationary solution is zero). Cf. also Lai & Wei (1985, Lemma 2).

I On Remark 8

As to v_t, the predictable quadratic variation is

$$\langle v \rangle_t = \sigma^2 \sum_{i=1}^{t} \frac{\theta_i^2}{i^{2(1-c)}} = O(1) \quad a.s..$$

Therefore, by the MCT for martingales with bounded predictable quadratic variation, v_t converges a.s. to some finite random variable.

As to w_t, for every $\delta > 0$,

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

so that

$$\sum_{t=1}^{\infty} \mathbf{P}\left(|w_t| > \delta\right) < \infty.$$

Therefore, by the Borel-Cantelli Lemma,

$$\mathbf{P}\left(|w_t| > \delta \quad i.o.\right) = 0.$$

Equivalently, $\lim_{t\to\infty} w_t = 0$ a.s..

References

Christopeit, N. & Hoderlein, S. (2006), 'Local partitioned regression', *Econometrica* **74**, 787–817.

Christopeit, N. & Massmann, M. (2013), A note on an estimation problem in models with adaptive learning. TI Discussion Paper 13-151/III.

Kawata, T. (1972), *Fourier Analysis in Probability Theory*, Academic Press, New York.

Lai, T. L. & Wei, C. Z. (1982*a*), 'Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems', *Annals of Statistics* **10**, 154–166.

Lai, T. L. & Wei, C. Z. (1982*b*), 'Asymptotic properties of projections with applications to stochastic regression problems', *Journal of Multivariate Analysis* **12**, 346–370.

Lai, T. L. & Wei, C. Z. (1985), Asymptotic properties of multivariate weighted sums with applications to stochastic linear regression in linear dynamic stystems, *in* P. R. Krishnaiah, ed., 'Multivariate Analysis', Vol. VI, North-Holland, Amsterdam, pp. 375–393.

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