TI 2013-151/III
Tinbergen Institute Discussion Paper

A Note on an Estimation Problem in Models with Adaptive Learning

Norbert Christopeit[1]
Michael Massmann[2]

[1] *University of Bonn, Germany;*
[2] *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, The Netherlands.*

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

A note on an estimation problem in models with adaptive learning

Norbert Christopeit
University of Bonn

Michael Massmann*
Vrije Universiteit Amsterdam
and Tinbergen Institute

September 24, 2013

Abstract

This paper provides an example of a linear regression model with predetermined stochastic regressors for which the sufficient condition for strong consistency of the ordinary least squares estimator by Lai & Wei (1982, *Annals of Statistics*) is not met. Nevertheless, the estimator is strongly consistent, as shown in a companion paper, cf. Christopeit & Massmann (2013b). This is intriguing because the Lai & Wei condition is the best currently available and is referred to as "in some sense the weakest possible". Moreover, the example discussed in this paper arises naturally in a class of macroeconomic models with adaptive learning, the estimation of which has recently gained popularity amongst researchers and policy makers.

keywords: least-squares regression, stochastic regressors, strong consistency, minimal sufficient condition, adaptive learning.

JEL codes: C22, C51, D83

*corresponding author: Michael Massmann, Department of Econometrics, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, phone: +31 (0)20 5986014, email: `m.massmann@vu.nl`

Motivation and result

Consider the general linear regression model

$$y_t = \theta' z_t + \varepsilon_t, \quad t = 1, 2, \ldots, \tag{1}$$

where the ε_t are unobservable errors, $\theta = (\theta_1, \ldots, \theta_K)'$ is a vector of unknown parameters and y_t is the observed response to the inputs $z_t = (z_{t1}, \ldots, z_{tK})'$. For serially uncorrelated ε_t, the parameter vector θ is usually estimated by the ordinary least squares (OLS) estimator

$$\widehat{\theta}_T = \left(Z_T' Z_T\right)^{-1} Z_T' y_{(T)}$$

based on the observations $y_{(T)} = (y_1, \ldots, y_T)'$ and $Z_T = (z_{tk})_{1 \le t \le T, 1 \le k \le K}$. For *deterministic* regressors z_{tk}, there exists a well established theory for strong consistency (i.e. convergence with probability one) of the OLS estimator, cf. Anderson & Taylor (1976), Drygas (1976), Lai & Robbins (1977) and Lai, Robbins & Wei (1978, 1979), providing both sufficient and necessary conditions. For *stochastic* regressors, the situation is more delicate. The best result obtained so far is by Lai & Wei (1982*a*, 1982*b*) and draws on two assumptions:

(A1) (ε_T) is a martingale difference sequence with respect to some basic filtration $(\mathcal{F}_T)$.

(A2) (z_T) is a predetermined sequence with respect to $(\mathcal{F}_T)$, i.e. z_T is $\mathcal{F}_{T-1}$-measurable for each T.

Denote the sample second moment matrix of the regressors by

$$M_T = Z_T' Z_T,$$

and let $\lambda_{\max}(T)$ and $\lambda_{\min}(T)$ be the maximal and the minimal eigenvalue of M_T, respectively. Lai & Wei (1982*a*) then prove the following.

Theorem (Lai & Wei, 1982a) *If the disturbances satisfy the condition* $\sup_T E\left(\varepsilon_T^2 | \mathcal{F}_{T-1}\right) < \infty$ *a.s., a sufficient condition for strong consistency of the OLS estimators of* θ *is the following:*

$$\lambda_{\min}(T) \to \infty \quad \text{and} \quad [\log \lambda_{\max}(T)]^{1+\rho} = o\left(\lambda_{\min}(T)\right) \quad \text{a.s.} \tag{2}$$

for some $\rho > 0$. *If, in addition,* $\sup_T E\left(\varepsilon_T^{2+\eta} | \mathcal{F}_{T-1}\right) < \infty$ *for some* $\eta > 0$, *then it suffices to require* (2) *for* $\rho = 0$.

Lai & Wei (1982*a*) refer to the condition in (2) as "in some sense the weakest possible" (p. 155), since even a marginal violation of it can lead to inconsistency. For $\rho = 0$, they give an example to that effect. The purpose of this Note is to provide a different example, see Result below, which may be considered as a converse of that in Lai & Wei (1982a) in the sense that, despite a marginal violation of (2), the strong consistency of $\widehat{\theta}_T$ continues to hold. This example arises in a typical macroeconomic model in which agents are not assumed to be fully rational but to form expectations by means of an adaptive learning algorithm. Estimation of models in this model class have recently gained popularity amongst researchers and policy makers; see for instance the New Keynesian Phillips curve models estimated by Milani (2007) and Chevillon, Massmann & Mavroeidis (2010), the European Central Bank's New Multi-Country Model by Dieppe, González Pandiella, Hall & Willman (2011), and the inflation model by Malmedier & Nagel (2012).

Specifically, consider the simple linear regression model

$$y_t = \delta + \beta a_{t-1} + \varepsilon_t, \tag{3}$$

where ε_t is an i.i.d. Gaussian disturbance with mean 0 and variance σ^2, where the regressor a_t is determined by the recursion

$$a_t = a_{t-1} + \gamma_t \left(y_t - a_{t-1} \right), \tag{4}$$

$a_0 = 0$, and where

$$\gamma_t = \frac{\gamma}{t} \tag{5}$$

is a weighting, or gain, sequence. It is assumed throughout that $\gamma > 0$ and $\beta < 1$. This model is derived from the more general specification $y_t = \delta w_t + \beta y^e_{t|t-1} + \varepsilon_t$ where $y^e_{t|t-1}$ denotes agents' expectations of y_t given the information set $\mathcal{F}_{t-1} = \sigma \left(y_s, s \leq t-1; w_s, s \leq t \right)$, see for instance the classical cobweb model in Bray & Savin (1986) or the Lucas (1973) aggregate supply model. Importantly, in the present setting, agents would no longer possess sufficient knowledge to form rational expectations $y^e_{t|t-1} = \mathbf{E} \left(y_t | \mathcal{F}_{t-1} \right)$ but, instead, are assumed to know only the structure of the rational expectations equilibrium $y_t = \alpha w_t + \varepsilon_t$ and to estimate the unknown parameter α recursively by means of a stochastic approximation algorithm such that their forecast is $y^e_{t|t-1} = a_{t-1} w_t$. Assuming for analytical tractability that w_t is constant and, then, setting $w = 1$ without loss of generality, the model in (3)-(4) obtains. The weighting sequence in (5) makes of the updating mechanism in (4) an instance of a so-called decreasing-gain recursion, since $\gamma_t \to 0$. The asymptotic properties of this system are analysed in detail in two companion papers, viz. Christopeit & Massmann (2013a, 2013b). The properties turn out to depend crucially on $c = \gamma \left(1 - \beta \right)$. In particular, Christopeit & Massmann (2013b) prove that the OLS estimators of δ and β are strongly consistent for $c > 1/2$. Nevertheless, the following result shows that Lai & Wei's condition (2) is violated; the proof is presented in the following section.

Result *For the model in (3)-(5), the two eigenvalues of the sample second moment matrix M_T are given by*

$$\begin{aligned}\lambda_{\min} &= \frac{A_T + O(1)}{\alpha^2 + 1 + o(1)}\,(1 + o(1)) \\ \lambda_{\max} &= T\left[\alpha^2 + 1 + o(1)\right](1 + o(1))\end{aligned}$$

where

$$A_T = \sum_{t=1}^{T} (a_t - \overline{a}_T)^2 .$$

When $c > 1/2$, although $\lambda_{\min}(T) \to \infty$ it turns out that

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

Consequently, Lai & Wei's condition (2) is violated for $c > 1/2$.

It will also be seen below that condition (2) *does* hold when $c < 1/2$; accordingly, the estimators in this case are strongly consistent, cf. Christopeit & Massmann (2013*a*). Investigation of the boundary case $c = 1/2$ is left to future research.

Proof and discussion

Calculating the eigenvalues

Model (3) is a special bivariate case of (1) with $\theta = (\delta, \beta)'$ and $z_t = (1, x_t)'$, where $x_t = a_{t-1}$. The sample second moment matrix is, then,

$$M_T = T \begin{pmatrix} 1 & p_T \\ p_T & r_T \end{pmatrix}$$

where

$$p_T = \overline{x}_T = \overline{a}_T - \frac{1}{T} a_T, \tag{6}$$

$$r_T = \overline{x_T^2} = \overline{a_T^2} - \frac{1}{T} a_T^2. \tag{7}$$

We adopt the usual notation $\overline{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t$ and $\overline{x_T^2} = \frac{1}{T}\sum_{t=1}^{T} x_t^2$. Recall also that $a_0 = 0$. For ease of notation, we suppress the subscript T of p_T and r_T for a moment. Simple calculation shows that the eigenvalues of M_T are given by

$$\begin{aligned}\lambda_{\pm} &= \frac{T}{2}\left[r + 1 \pm \sqrt{(r+1)^2 - 4\left(r - p^2\right)}\right] \\ &= \frac{T}{2}\left[r + 1 \pm \sqrt{(r-1)^2 + 4p^2}\right].\end{aligned}$$

Therefore $\lambda_{\max} \geq T|p|$. To obtain the exact rate, consider the expansion

$$\begin{aligned}
\lambda_{\max} &= \frac{T}{2}\left[r+1+\sqrt{(r+1)^2-4\left(r-p^2\right)}\right] \\
&= \frac{T(r+1)}{2}\left[1+\sqrt{1-4\frac{r-p^2}{(r+1)^2}}\right] \\
&= \frac{T(r+1)}{2}\left[1+(1-2d-2R(d))\right] \\
&= T(r+1)\left[1-d-R(d)\right], \qquad (8)
\end{aligned}$$

where we have put

$$d = d_T = \frac{r-p^2}{(r+1)^2} \qquad (9)$$

and where $R(d)$ is the residual in the Taylor expansion of the square root function about one:

$$\sqrt{1-4d} = 1-2d-\frac{2d^2}{(1-4\theta d)^{3/2}} = 1-2d-2R(d),$$

with $0<\theta<1$. Similarly, for $\lambda_{\min}$,

$$\begin{aligned}
\lambda_{\min} &= \frac{T}{2}\left[r+1-\sqrt{(r+1)^2-4\left(r-p^2\right)}\right] \\
&= \frac{T(r+1)}{2}\left[1-\sqrt{1-4\frac{r-p^2}{(r+1)^2}}\right] \\
&= \frac{T(r+1)}{2}\left[1-(1-2d-2R(d))\right] \\
&= T(r+1)\left[d+R(d)\right], \qquad (10)
\end{aligned}$$

for the same d and $R(d)$ as above. Note that

$$R(d) = O\left(d^2\right) \text{ as } d \to 0. \qquad (11)$$

Asymptotics of eigenvalues

It is shown in Christopeit & Massmann (2013a) that $a_T \to \alpha$ with probability one. Therefore, p_T and r_T in (6)-(7) satisfy, with probability one,

$$\begin{aligned}
p_T &= \overline{a}_T + O(T^{-1}) \\
r_T &= \overline{a_T^2} + O(T^{-1})
\end{aligned}$$

as well as

$$r_T - p_T^2 = \overline{a_T^2} - \overline{a}_T^2 + O(T^{-1}). \qquad (12)$$

Due to the convergence of a_T it follows that $p_T = \alpha + o(1)$ and $r_T = \alpha^2 + o(1)$ such that d_T and $R(d_T)$ in (9) and (11) satisfy, almost surely,

$$\begin{aligned}
d_T &= o(1), \\
R(d_T) &= o(1)
\end{aligned}$$

and

$$TR(d_T) = Td_T \cdot O(d_T) = Td_T \cdot o(1).$$

To ascertain the convergence rate of d_T, introduce

$$A_T = \sum_{t=1}^{T} (a_t - \overline{a}_T)^2 = T\left[\overline{a_T^2} - \overline{a}_T^2\right].$$

This is the quantity whose behaviour turns out crucial for the consistency of the OLS estimator of δ and β, see Christopeit & Massmann (2013*a*). By virtue of (12),

$$r_T - p_T^2 = \frac{1}{T} A_T + O(T^{-1}),$$

such that Td_T may be written as

$$Td_T = \frac{A_T + O(1)}{(r_T + 1)^2} = \frac{A_T + O(1)}{(\alpha^2 + 1 + o(1))^2} = \frac{A_T}{(\alpha^2 + 1)^2}\left[1 + O\left(A_T^{-1}\right) + o(1)\right].$$

Therefore, by (10),

$$\begin{aligned}
\lambda_{\min} &= (r_T + 1)\, T\, [d_T + R(d_T)] \\
&= (r_T + 1)\, Td_T\, [1 + o(1)] \\
&= \left[\alpha^2 + 1 + o(1)\right] \frac{A_T}{(\alpha^2 + 1)^2} \left[1 + O\left(A_T^{-1}\right) + o(1)\right] \\
&= \frac{A_T}{\alpha^2 + 1}\left[1 + O\left(A_T^{-1}\right) + o(1)\right]
\end{aligned}$$

and by (8),

$$\begin{aligned}
\lambda_{\max} &= T\,(r_T + 1)\,[1 - d_T - R(d_T)] \\
&= T\left[\alpha^2 + 1 + o(1)\right](1 + o(1)).
\end{aligned}$$

As a consequence,

$$\frac{\log \lambda_{\max}(T)}{\lambda_{\min}(T)} = \left(\alpha^2 + 1\right) \frac{\log T + O(1)}{A_T\left[1 + O\left(A_T^{-1}\right) + o(1)\right]}. \tag{13}$$

It is shown in Christopeit & Massmann (2013*a*, equation (B.14)) that, for $c > 1/2$,

$$\frac{A_T}{\log T} \xrightarrow{P} \frac{\gamma^2\sigma^2}{2c - 1}. \tag{14}$$

Since A_T is monotone increasing, this implies that $\lim_{T\to\infty} A_T = \infty$ a.s.. It is not known whether (14) also holds with probability one. As a consequence of (14),

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

In other words, Lai & Wei's condition (2) is violated for $c > 1/2$, as claimed in Result. Nevertheless, Christopeit & Massmann (2013*b*) prove that strong consistency of the OLS estimator for δ and β *does* hold for $c > 1/2$.

Remark 1 *For* $c < 1/2$*, it is shown in Christopeit & Massmann (2013*a*, equation (B.22)) that the exact divergence rate of* A_T *is*

$$A_T = \frac{\gamma^2 v^2}{1-2c} T^{1-2c}(1+o(1)) \quad a.s.,$$

where v *is a random variable which is nonzero with probability one. Hence, from (13),*

$$\lim_{T\to\infty} \frac{\log \lambda_{\max}(T)}{\lambda_{\min}(T)} = 0 \quad a.s.$$

such that Lai & Wei's condition (2) is satisfied for $c < 1/2$*.*

Remark 2 *The divergence rate of* A_T *for the boundary case* $c = 1/2$ *is not known. Hence the question of whether or not Lai & Wei's condition (2) is satisfied is left to future research.*

Remark 3 *For the bivariate case of simple regression models like (3), a somewhat weaker condition for strong consistency of the slope estimator is given in Lai & Wei (1982*a*), namely,*

$$\frac{A_T}{\log T} \to \infty \quad a.s.. \tag{15}$$

In view of (14), (15) is not satisfied for $c > 1/2$*. For* $c < 1/2$*, it is since (15) is implied by (2).*

References

Anderson, T. W. & Taylor, J. B. (1976), 'Strong consistency of least squares estimates in normal linear regression', *Annals of Statistics* **4**, 788–790.

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

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

Christopeit, N. & Massmann, M. (2013*a*), Estimating structural parameters in regression models with adaptive learning. TI Discussion Paper 13-111/III.

Christopeit, N. & Massmann, M. (2013*b*), Strong consistency of the least-squares estimator in simple regression models with stochastic regressors. TI Discussion Paper 12-109/III.

Dieppe, A., González Pandiella, A., Hall, S. & Willman, A. (2011), The ECB's new multi-country model for the Euro area, Working Paper 1316, European Central Bank.

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

Lai, T. L. & Robbins, H. (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., Robbins, H. & Wei, C. Z. (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., Robbins, H. & Wei, C. Z. (1979), 'Strong consistency of least squares estimates in multiple regression II', *Journal of Multivariate Analysis* **9**, 343–361.

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.

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

Malmedier, U. & Nagel, S. (2012), Learning from inflation experience. Mimeo.

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