SFB 303, Projektbereich B

Discussion Paper No. B–315

OLS–Learning in Non-Stationary Models with Forecast Feedback

Markus Zenner[1]

May 1995

[1]Institut für Ökonometrie und Operation Research
Adenauerallee 24–42, 53113 Bonn, Germany
E-Mail: zenner@eos.or.uni-bonn.de
The research was supported by the Deutsche Forschungsgemeinschaft (SFB 303) at the University of Bonn

Abstract

In this study we consider a linear model with forecast feedback in which boundedly rational agents are learning the parameter values of the rational expectations equilibrium by the OLS learning procedure. We show strong consistency of the OLS estimates under much weaker assumptions on the involved time series than the ones usually employed. This result extends the boundedly rational learning approach to models including non-stationary time series, like processes with polynomial trends or unit root autoregressive processes, and indicates that the idea that agents can learn only stationary rational expectations equilibria is misleading.

JEL Code: C22, C40, C62, D83

Keywords: Rational expectations equilibrium, boundedly rational learning, stochastic approximation, non-stationary time series.

Contents

1 Introduction **2**

2 The Model **4**

3 Convergence Analysis **7**

4 Convergence Results **13**

Stationary and Ergodic Processes 13

Stable Processes 14

Non-Stable Processes 15

5 Conclusions and Remarks **21**

6 References **23**

Chapter 1

Introduction

A specific feature of economic models is that expectations of agents matter and have to be modeled in order to complete the model. But since the process which generates individual expectations is neither observable nor completely understood and empirical results do not suggest one specific expectations formation scheme it is tempting for a model builder to employ an ad hoc expectations scheme. But, as it is well known, the choice of an expectations formation scheme is not only a matter of taste or convenience. Dynamic models incorporating different expectations formation schemes can show substantially different dynamic behaviour. Moreover, by introducing a suitable expectations formation scheme it is possible to construct models which show almost any dynamical behaviour and are able "to explain almost everything — and thus nothing" as GRANDMONT (1992, p. 13) remarks.

The rational expectations hypothesis (REH) introduced by MUTH (1961) which suggests that agents' expectations "are essentially the same as the predictions of the relevant theory" (p. 315) points out a way out of this dilemma. At first sight the REH seems to be an attractive way to model agents' expectations since it is free from any ad-hoc assumptions and reflects the common point of view that "information is scarce, and the economic system does not waste it" (MUTH (1961, p. 316)). But the microeconomic foundation of the REH on the level of individual agents has turned out to be problematic since it imposes extreme informational assumptions on agents.

To support the REH it was frequently argued that agents can learn somehow to form rational expectations. More precisely, it was argued that "by observing the history of a stationary world, people can eventually learn the objective probability distributions" of real outcomes "by using Bayesian or classical statistical techniques" (BRAY (1983, p. 124)). Unfortunately, this problem is more difficult than the formulation suggest since agents, during the learning phase, change their behaviour in the light of what they have learned and thereby introduce non-stationarities into the model. For that reason standard results of Bayesian theory and classical statistics cannot provide a rigorous mathematical proof of the assertion that agents can learn to form rational expectations. Since the late seventies several studies were contributed to the question whether agents can learn to form rational expectations following a 'reasonable' learning procedure but there are still some questions open.

A point which was frequently stressed in the literature is that learning, in order to be successful, requires a certain amount of stationarity of the economic environment. Therefore the first analytical studies (BRAY (1982, 1983), BRAY/SAVIN (1986), KOTTMANN (1990), and MOHR (1990)) were restricted to the case that the variables agents use to form their predictions (i.e. the regressors of the OLS-procedure) are exogenous and follow a stationary and ergodic process[1].

[1]Only FOURGEAUD ET AL. (1986) consider the slightly more general case that the involved time series are stable in the sense that the condition number of the matrix of moments remains bounded.

In models in which agents' predictions are based also on lagged endogenous variables stationarity and ergodicity of the involved times series cannot be assumed a priori since these properties are endogenous. Nevertheless, since some kind of stability is needed in order to ensure convergence of the estimates towards the RE parameters it is generally assumed that at least the exogenous variables are stationary and ergodic processes (see, e.g. MARCET/SARGENT (1989a,b) and KUAN/WHITE (1994)). In ZENNER (1994b) the assumption of stationarity and ergodicity is dropped but it is assumed that the exogenous variables are stable processes in the sense that the condition number of the matrix of moments remains bounded. Hence the case that the exogenous variables follow an integrated process or a process with some trend is excluded as well as it is excluded that the RE solution possesses roots on or outside the unit circle[2].

In summary it may be said that virtually nothing is known about the performance of boundedly rational learning procedures in models in which the exogenous variables are not necessarily stable processes in the sense mentioned above.

On the other hand it is well known that the least squares estimates in linear regression models are strongly consistent under much more general conditions than stationarity and ergodicity of the involved time series (see, e.g., LAI/WEI (1982)). It is only necessary that the orders of growth of the several regressor variables (measured by the minimum and maximum eigenvalues of the matrix of moments) do not differ 'too much'. Hence it is worthwhile to analyze whether this property carries over to models with forecast feedback.

In this paper we show that this is, in fact, the case. We show that in a simple linear model with a single forecast term agents can learn to form rational expectations with probability one under assumptions corresponding to the weakest possible assumptions in linear regression models. This result which has, to our knowledge, no counterpart in the literature extends the boundedly rational learning approach to a larger class of models. It covers processes with polynomial trends as well as autoregressive processes with unit roots for the exogenous variables. This extension might prove to be useful especially in macro-economics.

The paper is arranged as follows. In Chapter 2 we present the model, specify the learning procedure, and list the assumptions underlying our convergence analysis. The main convergence result is proved in Chapter 3. Since the conditions ensuring convergence differ from standard assumptions Chapter 4 is devoted to some applications for which these conditions are met. These applications include processes with polynomial trend variables and autoregressive models with unit roots for the exogenous variables. Finally, in Chapter 5 we give some concluding remarks.

[2]Only the studies by MARCET/SARGENT (1989c) and ZENNER (1992a,b, 1994a) consider the case that the RE solution is an autoregressive process with roots on or outside the unit circle. But these studies are restricted to univariate first order autoregressive models and the convergence analysis relies crucially on that special structure.

Chapter 2

The Model

We consider a *linear model with forecast feedback* given by its reduced form equation

$$y_{t+1} = \boldsymbol{\phi}' \boldsymbol{z}_t + a y^e_{t+1} + w_{t+1}, \qquad t \geq 0, \tag{2.1}$$

where

- y_t is the time t real valued endogenous variable,
- $\boldsymbol{z}_t$ is an n-dimensional random vector with $n \geq 1$,
- w_t is the time t disturbance term,
- y^e_t is the aggregate or market prediction of y_t made by agents at time $t-1$, and
- $\boldsymbol{\phi} \in \mathbb{R}^n$ and $a \in \mathbb{R}$ are model parameters.

We make the following two assumptions specifying the required properties of the disturbance terms w_t and the random variables $\boldsymbol{z}_t$.

Assumption (A.1):
The disturbance terms $\{w_t\}_{t \geq 1}$ form a martingale difference sequence with respect to a filtration $\{\mathcal{F}_t\}_{t \geq 0}$, i.e. $\mathrm{E}[w_{t+1}|\mathcal{F}_t] = 0$ a.s., such that

$$\sup_{t \geq 0} \mathrm{E}[|w_{t+1}|^{2+\delta}|\mathcal{F}_t] < \infty \qquad \text{a.s.} \tag{2.2}$$

for some fixed $\delta > 0$. $\square$

Assumption (A.2):
The random variables $\boldsymbol{z}_t$ are $\mathcal{F}_t$-measurable for all $t \geq 0$ and satisfy

$$\sum_{t=0}^{\infty} \|\boldsymbol{z}_t\|^2 = \infty \qquad \text{a.s.} \tag{2.3}$$

and

$$T^* := \inf\{t \in \mathbb{N} | \boldsymbol{Z}_t \text{ is nonsingular}\} < \infty \qquad \text{a.s.} \tag{2.4}$$

where $\boldsymbol{Z}_T$ is defined as $\boldsymbol{Z}_T := \sum_{t=0}^{T} \boldsymbol{z}_t \boldsymbol{z}'_t$. $\square$

Assumption (A.1) is a standard assumption on disturbance terms in econometrics. It generalizes the common assumption that the disturbance terms form a white noise process with bounded $(2+\delta)$th moments[1].

Assumption (A.2) is a minimum assumption on the variables $\boldsymbol{z}_t$ and it should be noted that it is not sufficient for our convergence results. But since the additional assumptions ensuring convergence arise quite naturally in the course of the convergence analysis in Chapter 3 we decided to introduce them lateron. The conditions (2.3) and (2.4) are very weak in that they only require that $\boldsymbol{z}_t$ does not vanish too rapidly as $t \to \infty$ and that there are no exact linear dependencies among the variables contained in $\boldsymbol{z}_t$.

Notice that the random variable T^* is measurable. In fact, it is a stopping time with respect to the canonical filtration of $\{\boldsymbol{Z}_t\}$.

Since we adopt the point of view that the σ-algebra $\mathcal{F}_t$ contains all information about the model (2.1) available at time t the measurability assumption in (A.2) imposes no restriction on the variables $\boldsymbol{z}_t$. Nevertheless, Assumptions (A.1) and (A.2) together rule out the possibility of autocorrelated disturbance terms and correlations between w_{t+1} and $\{\boldsymbol{z}_s\}_{s\leq t}$.

Notice that Assumption (A.2) does not require the variables $\boldsymbol{z}_t$ to be exogenous. The vector $\boldsymbol{z}_t$ may contain lagged endogenous variables but, unlike to our recent study (ZENNER (1994b)), we are not able to verify the conditions required for convergence in that case.

In order to complete the model we still have to specify the predictions of agents. We assume that at time t, the time when agents form their predictions y^e_{t+1}, the realization of $\boldsymbol{z}_t$ is observable and known by agents and that the predictions y^e_{t+1} are based on the information set[2]

$$I_t = \{\boldsymbol{z}_t, y_t, \ldots, \boldsymbol{z}_1, y_1, \boldsymbol{z}_0\} \subset \mathcal{F}_t. \tag{2.5}$$

If all agents know the equation (2.1) and the parameters a and $\boldsymbol{\phi}$ and, additionally, it is common knowledge that the knowledge of (2.1), a, and $\boldsymbol{\phi}$ is common knowledge then agents can calculate the so-called *rational expectations* $\mathrm{E}[y_{t+1}|I_t]$, thus the conditional mathematical expectation of y_{t+1} with respect to the information set I_t. In our model the rational expectations are uniquely given as

$$\mathrm{E}[y_{t+1}|I_t] = \frac{1}{1-a}\boldsymbol{\phi}'\boldsymbol{z}_t, \tag{2.6}$$

provided that $a \neq 1$. If the agents' predictions in (2.1) are given as $y^e_{t+1} = (1-a)^{-1}\boldsymbol{\phi}'\boldsymbol{z}_t$ then the model (2.1) is called to be in a *rational expectations equilibrium* and the process $\{y_t\}$ is called a *rational expectations solution.*

Since we assume the variables $\boldsymbol{z}_t$ to be known by agents at time t agents have to learn only the parameter

$$\bar{\boldsymbol{\theta}} = \frac{1}{1-a}\boldsymbol{\phi} \tag{2.7}$$

in order to form rational expectations. The learning procedure is made explicit in

---

[1] The reason why we require w_t to have slightly more than bounded second (conditional) moments is, more or less, a didactical one. The slightly stronger assumption enables us to obtain a slightly stronger convergence result and reduces the notational complexity in some proofs. Nevertheless, the technique used to show convergence applies also for the weaker assumption of bounded second (conditional) moments, but in that case we need slightly stronger assumptions on the variables z_t. This tradeoff between the different assumptions will be described in more detail in Chapter 3.

[2] We thus assume that the agents know which variables determine the endogenous variable. Hence we rule out the possibility that agents are poorly informed. But since we do not require that the parameter vector ϕ is non-zero in each component we include the possibility that agents consider also sun-spot variables which influence the endogenous variable only via the agents' expectations. These sun-spot variables are then contained in z_t.

Assumption (A.3):
Agents' predictions y^e_{t+1} are given by

$$y^e_{t+1} = \boldsymbol{\theta}_t{}'\boldsymbol{z}_t, \qquad t \geq 0, \tag{2.8}$$

with

$$\boldsymbol{\theta}_t = \left(\sum_{s=0}^{t-1} \boldsymbol{z}_s\boldsymbol{z}'_s\right)^{-1} \sum_{s=0}^{t-1} \boldsymbol{z}_s y_{s+1} \tag{2.9}$$

where the stochastic inverse is defined as the Moore–Penrose inverse if necessary[3]. □

In order to motivate an assumption like (A.3) is is usually assumed that agents believe in an auxiliary model

$$y_{t+1} = \boldsymbol{\theta}'\boldsymbol{z}_t + e_{t+1} \tag{2.10}$$

with some parameter $\boldsymbol{\theta} \in I\!R^n$ and a disturbance term e_{t+1} which is independent from the information set I_t. Based on this model agents estimate the unknown (hypothetical) parameter $\boldsymbol{\theta}$ by the ordinary least squares (OLS) procedure and predict y_{t+1} by $y^e_{t+1} = \boldsymbol{\theta}_t{}'\boldsymbol{z}_t$, where $\boldsymbol{\theta}_t$ is the time t OLS-estimate of $\boldsymbol{\theta}$. Given the belief of agents that (2.10) is the correctly specified model the use of the OLS–learning procedure is rational.

One can interpret the auxiliary model as reflecting the agents' beliefs concerning their economic environment. Hence agents do not fully understand the economic system they are part of. In particular, they do not take into account that the market expectation affects the outcome of the endogenous variable[4]. But one can also think of the agents' behaviour in a more naive way. Suppose that agents are aware that there exists a relationship between y_{t+1} and $\boldsymbol{z}_t$, either because they know roughly how the economic system works or that explorative data analysis has revealed such a relationship. In order to extract the information on y_{t+1} contained in $\boldsymbol{z}_t$ agents simply carry out a regression of y_{t+1} on $\boldsymbol{z}_t$.

---

[3] The convention that the stochastic inverse be given by the Moore–Penrose inverse ensures that θ_t is well-defined and measurable on the full ω–space. In practice other conventions might be more convenient such as replacing a singular matrix Z_t by $\epsilon I + Z_t$ with some fixed $\epsilon > 0$. Since T^* is a stopping time with respect to the filtration $\{\mathcal{F}_t\}$ one can easily allow for a split-case definition of θ_t for ω such that $t \geq T^*$ and $t < T^*$, resp.

[4] The main methodological problem of the boundedly rational learning approach consists in justifying that agents neglect the forecast feedback. Clearly, this problem cannot be solved without imposing some amount of irrationality on agents' behaviour. Some arguments to support this lack of rationality are the following.

Since expectations are generally not observable agents cannot find out what other agents expect and, in general, they cannot even deduce ex post what the other agents did expect. In addition, gathering information about other agents' expectations can be too costly. Hence it is impossible to include expectations into an auxiliary model like (2.10). Even if there exists some institution which reports ex post the past market expectation there remains the problem to predict the current market expectation.

In a competitive market in which a single agents' prediction has no impact on real outcomes agents may believe that the market expectation, as an aggregate, is already rational, either because agents believe that the other agents are more sophisticated or that the average expectation is more accurate than an individual one (an argument which was already used by Muth (1961) based on empirical evidence).

But agents can also be aware of the problem of infinite regress of expectations yet do not know how to solve it since they do not know the parameter values of (2.1). (In the rational learning approach it is usually assumed that agents know at least the parameter values which have to be known in order to solve the problem of infinite regress.)

Finally, since we allow sun-spot variables among the variables contained in z_t the auxiliary model (2.10) can contain some kind of 'rumour variables' reflecting some economy-wide collective beliefs about future events.

Chapter 3

Convergence Analysis

We comply with the following notational conventions. Multivariate variables and multivariate parameters are denoted by boldface letters. The transpose of a (column) vector $\boldsymbol{z}$ will be denoted by $\boldsymbol{z}'$. By $\|\boldsymbol{z}\|^2$ we denote the Euclidian norm, thus $\|\boldsymbol{z}\|^2 = \boldsymbol{z}'\boldsymbol{z}$. For a matrix $\boldsymbol{A}$, $\lambda_{min}(\boldsymbol{A}), \lambda_{max}(\boldsymbol{A})$ denote the minimum and maximum eigenvalue in modulus. By $\|\boldsymbol{A}\|$ we denote the operator norm, thus $\|\boldsymbol{A}\|^2 = \lambda_{max}(\boldsymbol{A}'\boldsymbol{A})$. By $|\boldsymbol{A}|$ we denote the determinant of $\boldsymbol{A}$ and by $tr(\boldsymbol{A})$ its trace. Let a_t and b_t be two sequences of non-negative real numbers. We employ the Landau symbols $o(\cdot)$ and $O(\cdot)$ in the usual sense, thus $a_t = o(b_t) \Leftrightarrow a_t/b_t \to 0$ and $a_t = O(b_t) \Leftrightarrow sup_t|a_t/b_t| < \infty$.

In this chapter we show the following result which determines the order of convergence or divergence of the OLS-estimates.

Theorem 3.1:
Suppose that (A.1)–(A.3) *hold and* $a < 1/2$. *Let* λ_t *be given by*

$$\lambda_t = \boldsymbol{z}_t' \boldsymbol{Z}_t^{-1} \boldsymbol{z}_t. \tag{3.1}$$

If

$$\limsup_{t\to\infty} \lambda_t < \frac{1-2a}{(1-a)^2} \qquad a.s. \tag{3.2}$$

then

$$\|\boldsymbol{\theta}_{t+1} - \bar{\boldsymbol{\theta}}\|^2 = O\left(\frac{\log \lambda_{max}(\boldsymbol{Z}_t)}{\lambda_{min}(\boldsymbol{Z}_t)}\right) \qquad a.s.. \tag{3.3}$$

The following corollary is then obvious.

Corollary 3.2:
Suppose that (A.1)–(A.3) *hold and* $a < 1/2$. *If the condition* (3.2) *is satisfied and*

$$\log \lambda_{max}(\boldsymbol{Z}_t) = o\left(\lambda_{min}(\boldsymbol{Z}_t)\right) \qquad a.s. \tag{3.4}$$

then $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ *a.s.*

The proof of Theorem 3.1 is based on the analysis of the quadratic form

$$V_t = (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})' \boldsymbol{Z}_{t-1} (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}). \tag{3.5}$$

We show that the non-negative process V_t is an extended stochastic Lyapunov function in the sense of LAI (1989), i.e. we show that V_t has the 'almost supermartingale' property as introduced by ROBBINS/SIEGMUND (1971). This enables us to use a refinement of the almost supermartingale convergence theorem by LAI (1989) in order to show that $V_t = O(\log \lambda_{max}(\boldsymbol{Z}_{t-1}))$ a.s. under the assumptions of Theorem 3.1.

Proposition 3.3: (LAI (1989))
Let $\{w_t\}_{t\geq 1}$ be a martingale difference sequence with respect to some filtration $\{\mathcal{F}_t\}_{t\geq 0}$ such that $\sup_{t\geq 0} \mathrm{E}[w_{t+1}^2|\mathcal{F}_t] < \infty$ a.s. Let V_t, α_t, β_t, η_t, and u_t be non-negative $\mathcal{F}_t$-measurable random variables such that $\sum_{t=0}^{\infty} \alpha_t < \infty$ a.s. Suppose that for $t \geq 0$

$$V_{t+1} \leq (1+\alpha_t) V_t + \beta_{t+1} - \eta_{t+1} + u_t w_{t+1} \qquad a.s. \tag{3.6}$$

Then

(i) *On the event $[\sum_{t=1}^{\infty} \mathrm{E}[\beta_t|\mathcal{F}_{t-1}] < \infty]$, V_t converges a.s. and $\sum_{t=1}^{\infty} \mathrm{E}[\eta_t|\mathcal{F}_{t-1}] < \infty$ a.s.*

(ii) *For every $\alpha > 0$*

$$\max\left\{V_T, \sum_{t=1}^{T} \eta_t\right\} = O\left(\sum_{t=1}^{T} \beta_t + \left(\sum_{t=0}^{T-1} u_t^2\right)^{1/2+\alpha}\right) \qquad a.s. \tag{3.7}$$

We need some more auxiliary results. The first one is a simple lemma known as the theorem of ABEL/DINI (see KNOPP (1964)).

Lemma 3.4:
Let $(d_t)_{t\geq 0}$ be a sequence of non-negative real numbers with $d_0 > 0$ such that $D_t := \sum_{s=0}^{t} d_s \to \infty$ as $t \to \infty$. Let $\alpha \geq 0$, then

$$\sum_{t=0}^{\infty} \frac{d_t}{D_t^{\alpha}} < \infty \qquad \Longleftrightarrow \qquad \alpha > 1. \tag{3.8}$$

The following result is the keystone in the martingale difference approach. It is based on the local convergence theorem for martingales by CHOW (1965) and provides a kind of deterministic reduction for weighted sums of martingale difference sequences.

Lemma 3.5: (LAI/WEI (1982))
Suppose that $\{w_t\}$ *is a martingale difference sequence with respect to some filtration* $\{\mathcal{F}_t\}$ *such that* $\sup_{t\geq 0} \mathrm{E}[w_{t+1}^2|\mathcal{F}_t] < \infty$ *a.s. and* $\{u_t\}$ *is a sequence of* $\mathcal{F}_t$*-adapted random variables. Then*

(i) $\sum_{t=0}^{T} u_t w_{t+1}$ *converges a.s. on the event* $\left[\sum_{t=0}^{\infty} u_t^2 < \infty\right]$,

(ii) $\sum_{t=0}^{T} u_t w_{t+1} = o\left(\sum_{t=1}^{T} u_t^2\right)$ *a.s. on the event* $\left[\sum_{t=0}^{\infty} u_t^2 = \infty\right]$,

(iii) $\sum_{t=0}^{\infty} |u_t| w_{t+1}^2 < \infty$ *a.s. on the event* $\left[\sum_{t=0}^{\infty} |u_t| < \infty\right]$,

(iv) $\sum_{t=0}^{T} |u_t| w_{t+1}^2 = o\left(\sum_{t=0}^{T} |u_t|\right)^{1+\alpha}$ *a.s. on* $\left[\sum_{t=0}^{\infty} |u_t| = \infty\right]$ *for every* $\alpha > 0$.

If $\{w_t\}$ *satisfies also* $\sup_{t\geq 0} \mathrm{E}[|w_{t+1}|^{2+\delta}|\mathcal{F}_t] < \infty$ *a.s. for some* $\delta > 0$ *then*

(v) $\sum_{t=0}^{T} |u_t| w_{t+1}^2 = O\left(\sum_{t=0}^{T} |u_t|\right)$ *a.s. on* $\left[\sup_{t\geq 0} |u_t| < \infty\right]$.

The last auxiliary result concerns the quadratic form λ_t.

Lemma 3.6: (LAI/WEI (1982))
Let $(\boldsymbol{z}_t)$ *be a sequence of n-dimensional vectors. Let* $\boldsymbol{Z}_t$ *and* T^* *be defined as in Assumption* (A.2) *and* λ_t *as in (3.1) with* $\boldsymbol{Z}_t^{-1}$ *the Moore–Penrose inverse if necessary. Suppose that* $T^* < \infty$. *Then*

$$\lambda_t = \frac{|\boldsymbol{Z}_t| - |\boldsymbol{Z}_{t-1}|}{|\boldsymbol{Z}_t|} \qquad \textit{for} \qquad t \geq T^* \tag{3.9}$$

and

$$\sum_{t=0}^{T} \lambda_t = O\left(\log|\boldsymbol{Z}_T|\right) = O\left(\log \lambda_{max}(\boldsymbol{Z}_T)\right). \tag{3.10}$$

In the following lemma we show that the sum $\sum_{t=0}^{\infty} \lambda_t$ is a.s. unbounded under Assumption (A.2), a property we need lateron.

Lemma 3.7:
Let $\{\boldsymbol{z}_t\}$ *be a sequence of random vectors in* $\mathbb{R}^n$ *such that* (A.2) *is satisfied. Then*

$$\sum_{t=0}^{\infty} \boldsymbol{z}_t' \boldsymbol{Z}_t^{-1} \boldsymbol{z}_t = \infty \qquad a.s. \tag{3.11}$$

with $\boldsymbol{Z}_t^{-1}$ *the Moore–Penrose inverse if necessary.*

Proof:
We argue pathwise. Since $T^* < \infty$ a.s. by Assumption (A.2) it is sufficient to show (3.11) with the sum starting at $t = T^*$. Let $t \geq T^*$. Since $\boldsymbol{Z}_t$ is symmetric and positive definite we know that

$$z_t' \boldsymbol{Z}_t^{-1} z_t \geq z_t' z_t \lambda_{min}(\boldsymbol{Z}_t^{-1}) = \frac{z_t' z_t}{\lambda_{max}(\boldsymbol{Z}_t)} \tag{3.12}$$

and since $\lambda_{max}(\boldsymbol{Z}_t) \leq tr(\boldsymbol{Z}_t)$ we obtain

$$\sum_{t=T^*}^{\infty} z_t' \boldsymbol{Z}_t^{-1} z_t \geq \sum_{t=T^*}^{\infty} \frac{z_t' z_t}{\sum_{s=T^*}^{t} z_s' z_s} = \infty \tag{3.13}$$

where the equality follows by Lemma 3.4. ■

Proof of Theorem 3.1:
Due to the split-case definition of $\boldsymbol{\theta}_t$ in (A.3) we have to argue pathwise. For fixed t let $t \geq T^*$. Then we have the following recursive representation for the least squares estimates

$$\begin{aligned}
\boldsymbol{\theta}_{t+1} &= \boldsymbol{Z}_t^{-1} \sum_{s=0}^{t} z_s y_{s+1} \\
&= \boldsymbol{Z}_t^{-1} \left(\sum_{s=0}^{t-1} z_s y_{s+1} + z_t z_t' (\boldsymbol{\phi} + a\boldsymbol{\theta}_t) + z_t w_{t+1} \right) \\
&= \boldsymbol{Z}_t^{-1} \left(\boldsymbol{Z}_{t-1} \boldsymbol{\theta}_t + z_t z_t' (\boldsymbol{\phi} + a\boldsymbol{\theta}_t) + z_t w_{t+1} \right) \\
&= \boldsymbol{\theta}_t - \boldsymbol{Z}_t^{-1} z_t z_t' (\boldsymbol{\theta}_t - (\boldsymbol{\phi} + a\boldsymbol{\theta}_t)) + \boldsymbol{Z}_t^{-1} z_t w_{t+1} \\
&= \boldsymbol{\theta}_t - (1-a) \boldsymbol{Z}_t^{-1} z_t z_t' (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) + \boldsymbol{Z}_t^{-1} z_t w_{t+1}
\end{aligned} \tag{3.14}$$

using the equality $\boldsymbol{\phi} = (1-a)\bar{\boldsymbol{\theta}}$. Hence

$$\boldsymbol{\theta}_{t+1} - \bar{\boldsymbol{\theta}} = \left[\boldsymbol{I} - (1-a) \boldsymbol{Z}_t^{-1} z_t z_t' \right] (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) + \boldsymbol{Z}_t^{-1} z_t w_{t+1}. \tag{3.15}$$

Define

$$V_t = (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})' \boldsymbol{Z}_{t-1} (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}). \tag{3.16}$$

Then we obtain

$$\begin{aligned}
V_{t+1} &= (\boldsymbol{\theta}_{t+1} - \bar{\boldsymbol{\theta}})' \boldsymbol{Z}_t (\boldsymbol{\theta}_{t+1} - \bar{\boldsymbol{\theta}}) \\
&= (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})' \left[\boldsymbol{I} - (1-a) \boldsymbol{Z}_t^{-1} z_t z_t' \right]' \boldsymbol{Z}_t \left[\boldsymbol{I} - (1-a) \boldsymbol{Z}_t^{-1} z_t z_t' \right] (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) \\
&\quad + 2(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})' \left[\boldsymbol{I} - (1-a) \boldsymbol{Z}_t^{-1} z_t z_t' \right]' z_t w_{t+1} + z_t' \boldsymbol{Z}_t^{-1} z_t w_{t+1}^2 \\
&= V_t + (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})' z_t z_t' (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) + \lambda_t w_{t+1}^2 + u_t w_{t+1} \\
&\quad - (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})' \left[2(1-a) z_t z_t' - (1-a)^2 \lambda_t z_t z_t' \right] (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) \\
&= V_t + \lambda_t w_{t+1}^2 + u_t w_{t+1} \\
&\quad - \left[1 - 2a - (1-a)^2 \lambda_t \right] [z_t (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2
\end{aligned} \tag{3.17}$$

with

$$u_t = 2[1 - (1-a)\lambda_t] (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})' z_t. \tag{3.18}$$

Furthermore define

$$
\begin{aligned}
\tilde{\eta}_{t+1} &= [1-2a-(1-a)^2\lambda_t][\boldsymbol{z}_t'(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}})]^2, \\
\eta_{t+1} &= \tilde{\eta}_{t+1}1_{[\tilde{\eta}_{t+1}\geq 0]}, \\
\beta_{t+1} &= \lambda_t w_{t+1}^2-\tilde{\eta}_{t+1}1_{[\tilde{\eta}_{t+1}<0]}
\end{aligned}
\tag{3.19}
$$

to obtain

$$V_{t+1}=V_t+\beta_{t+1}-\eta_{t+1}+u_t w_{t+1}. \tag{3.20}$$

on the event $[t\geq T^*]$.

If $0\leq t<T^*$ we define V_{t+1} by (3.16) and set $\beta_{t+1}=V_{t+1}$, $\eta_{t+1}=V_t$, and $u_t=0$. Also set $V_0=0$ for all ω. Then V_t, β_{t+1}, η_{t+1}, and u_t are well-defined random variables satisfying (3.20) and the measurability assumption of Proposition 3.3.

In order to verify the assumptions of Proposition 3.3 we remark first that $\tilde{\eta}_t(\omega)<0$ holds only for finitely many t for almost all $\omega\in\Omega$. This is easily seen because

$$\tilde{\eta}_{t+1}\geq 0 \iff \lambda_t\leq\frac{1-2a}{(1-a)^2}, \tag{3.21}$$

hence by assumption (3.2) we can infer that $\tilde{\eta}_t<0$ holds only for finitely many t.

But this implies

$$\sum_{t=1}^{T}\beta_t=O\left(\sum_{t=0}^{T-1}\lambda_t w_{t+1}^2\right)+O(1) \qquad \text{a.s.} \tag{3.22}$$

By Lemma 3.7 we know that $\sum_0^\infty\lambda_t=\infty$ a.s. Hence, since $\lambda_t\leq n$, Lemma 3.5 (v) and Lemma 3.6 imply that

$$\sum_{t=1}^{T}\beta_t=O\left(\log\lambda_{max}(\boldsymbol{Z}_{T-1})\right) \qquad \text{a.s.} \tag{3.23}$$

Now we determine the order of $\sum_0^T u_t^2$. By assumption (3.2) we can infer that there exists a positive constant $\epsilon=\epsilon(\omega)$ such that $1-2a-(1-a)^2\lambda_t>\epsilon$ for all but finitely many t. This implies that $u_t^2\leq C\eta_{t+1}$ for all but finitely many t with some positive constant C. Hence

$$\sum_{t=0}^{T-1}u_t^2=O\left(\sum_{t=1}^{T}\eta_t\right) \qquad \text{a.s.} \tag{3.24}$$

Now we can apply Proposition 3.3. We obtain

$$\max\left\{V_T,\sum_{t=1}^{T}\eta_t\right\}=O\left(\sum_{t=1}^{T}\beta_t+\left(\sum_{t=1}^{T}u_t^2\right)^{1/2+\alpha}\right) \qquad \text{a.s.} \tag{3.25}$$

for every $\alpha>0$. This implies, using (3.23) and (3.24) with some $\alpha<1/2$, that

$$\max\left\{V_T,\sum_{t=1}^{T}\eta_t\right\}=O\left(\log\lambda_{max}(\boldsymbol{Z}_{T-1})\right) \qquad \text{a.s.} \tag{3.26}$$

Since

$$V_{T+1}=\|\boldsymbol{Z}_T^{1/2}(\boldsymbol{\theta}_{T+1}-\bar{\boldsymbol{\theta}})\|^2\geq\lambda_{min}(\boldsymbol{Z}_T)\|\boldsymbol{\theta}_{T+1}-\bar{\boldsymbol{\theta}}\|^2 \tag{3.27}$$

we finally obtain

$$\|\boldsymbol{\theta}_{T+1}-\bar{\boldsymbol{\theta}}\|^2=O\left(\frac{\log\lambda_{max}(\boldsymbol{Z}_T)}{\lambda_{min}(\boldsymbol{Z}_T)}\right) \qquad \text{a.s.} \tag{3.28}$$

and the proof is complete. ■

Since the proof of Theorem 3.1 applies in a pathwise manner we can formulate Theorem 3.1 also in the following way.

Theorem 3.8:
Suppose that (A.1)–(A.3) *hold and* $a < 1/2$. *Then, on the event where*

$$\limsup_{t\to\infty} \lambda_t < \frac{1-2a}{(1-a)^2} \qquad \textit{and} \qquad \frac{\log \lambda_{max}(\boldsymbol{Z}_t)}{\lambda_{min}(\boldsymbol{Z}_t)} = o(1) \qquad a.s. \tag{3.29}$$

holds, $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ *a.s.*

Remark:
There exists a tradeoff between the assumptions imposed on the disturbance terms and the order of $\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}$ in the following way. If we relax Assumption (A.1) in that we only require

$$\sup_{t\geq 0} \mathrm{E}[w_{t+1}^2|\mathcal{F}_t] < \infty \qquad \text{a.s.} \tag{3.30}$$

we obtain by the same reasoning as in the proof of Theorem 3.1, but with Lemma 3.5 (iv) instead of (v), the slightly weaker result

$$\|\boldsymbol{\theta}_{T+1} - \bar{\boldsymbol{\theta}}\|^2 = O\left(\frac{(\log \lambda_{max}(\boldsymbol{Z}_T))^{1+\delta}}{\lambda_{min}(\boldsymbol{Z}_T)}\right) \qquad \text{a.s.} \tag{3.31}$$

for every $\delta > 0$.

It is worth noting that Theorem 3.1 and Theorem 3.8 apply for any choice of random variables $\boldsymbol{z}_t$ such that Assumption (A.2) is satisfied. We do not a priori restrict $\boldsymbol{z}_t$ to contain only exogenous variables. Nevertheless, whenever $\boldsymbol{z}_t$ contains also lagged endogenous variables the conditions (3.2) and (3.4) are endogenous and have to be verified. Unfortunately, we failed in doing so although computer simulations suggest that (3.2) and (3.4) are satisfied for some range of parameter values a and $\boldsymbol{\phi}$. See Chapter 5 for some details.

Chapter 4

Convergence Results

Although the convergence results of the preceding chapter are satisfactory from the mathematical point of view it is not at all clear which kind of time series $\{\boldsymbol{z}_t\}$ satisfy the assumptions of Theorem 3.1 or Corollary 3.2, especially the condition (3.2) for the quadratic form λ_t. In this chapter we present some examples, including stationary and ergodic processes, processes with a polynomial trend, and autoregressive processes with unit roots. As already mentioned we confine ourselves to the case that $\boldsymbol{z}_t$ contains only exogenous variables.

Stationary and Ergodic Processes

In models with forecast feedback the exogenous variables are usually assumed to be stationary and ergodic. In such a case we obtain the following result.

Corollary 4.1:
Suppose that (A.1)–(A.3) *hold and* $a < 1/2$. *If*

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t' \longrightarrow \boldsymbol{Z}^* \qquad a.s. \tag{4.1}$$

with $\boldsymbol{Z}^*$ *a (possibly pathdependent) positive definite matrix. Then* $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ *a.s.*

Proof:
We show first that $\lambda_t \to 0$ a.s. By Lemma 3.6 we know that for t sufficiently large

$$\lambda_t = \boldsymbol{z}_t' \boldsymbol{Z}_t^{-1} \boldsymbol{z}_t = 1 - \frac{|\boldsymbol{Z}_{t-1}|}{|\boldsymbol{Z}_t|} = 1 - \left(\frac{t-1}{t}\right)^n \frac{|\frac{1}{t-1}\boldsymbol{Z}_{t-1}|}{|\frac{1}{t}\boldsymbol{Z}_t|}. \tag{4.2}$$

Hence $\lambda_t \to 0$ a.s. by (4.1) and condition (3.2) holds true. Moreover, (4.1) implies that $\lambda_{max}(\boldsymbol{Z}_t) = O(\lambda_{min}(\boldsymbol{Z}_t))$ a.s. Hence also condition (3.4) is met and Corollary 3.2. applies. ■

Corollary 4.1 gives a result which is somewhat weaker than the results of BRAY/SAVIN (1986), MARCET/SARGENT (1989a,b), and KOTTMANN (1990) which show a.s. convergence if $a < 1$. This indicates that the approach proposed in this paper is not the optimal one for this class of processes. Nevertheless, Theorem 3.1 gives a slightly stronger result determining the rate of convergence ($\|\boldsymbol{\theta}_{t+1} - \bar{\boldsymbol{\theta}}\|^2 = O(\log t/t)$) than the result $\|\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}\|^2 = o(t)$ obtained by FOURGEAUD ET AL. (1986) and KOTTMANN (1990).

Corollary 4.1 indicates that the condition (3.2) for the quadratic form λ_t does not mean any restriction in a stationary and ergodic environment. Therefore it is worthwhile to ask whether condition (3.2) can be dispensed with under more general conditions than (4.1). The answer is negative as the following example shows. If we relax the condition (4.1) slightly to $\lambda_{max} = O(\lambda_{min})$ a.s. we cannot conclude that $\lambda_t \to 0$ a.s. without imposing further assumptions on $\{\boldsymbol{z}_t\}$.

Example 4.2:
Consider the (deterministic) process z_t defined by

$$z_t = \begin{cases} \sqrt{\frac{9}{10}t} & \text{if } t = 10^m, m \in I\!N, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

It it easy to see that

$$\limsup_{t\to\infty} \lambda_t = \frac{9}{10} \qquad \text{and} \qquad \frac{1}{T}\sum_{t=1}^{T} z_t^2 \in [0.1, 0.9]. \tag{4.4}$$

Hence $\lambda_{max}(Z_t) = O(\lambda_{min}(Z_t))$ but $\lambda_t \not\to 0$. Nevertheless, Corollary 3.2 implies that $\theta_t \to \bar{\theta}$ a.s. if

$$-\left(\sqrt{\frac{10}{81}} + \frac{1}{9}\right) < a < \sqrt{\frac{10}{81}} - \frac{1}{9}. \tag{4.5}$$

This shows that our approach provides convergence results in situations where almost all other approaches fail. □

Stable Processes

FOURGEAUD ET AL. (1986) have shown that if $a < 1/2$ and

(i) $\lambda_{min}(\boldsymbol{Z}_T) \to \infty$ a.s.,

(ii) $\lambda_{max}(\boldsymbol{Z}_T) = O(\lambda_{min}(\boldsymbol{Z}_T))$ a.s.,

(iii) $\|\boldsymbol{Z}_T^{-1}(\boldsymbol{Z}_T - \boldsymbol{Z}_{T-1})\| \to 0$ a.s., and

(iv) $\|\boldsymbol{Z}_T^{-1}\sum_{t=0}^{T} \boldsymbol{z}_t w_{t+1}\| \to 0$ a.s.

then $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s. We show now that their result is a special case of Corollary 3.2.

Assumption (i) is stronger than Assumption (A.2), and (ii) is stronger than (3.4). In view of Lemma 3.5 Assumption (iv) is implied by (i) and (ii). (FOURGEAUD ET AL. make the slightly different assumption that the disturbance terms satisfy the orthogonality condition $\mathrm{E}[\boldsymbol{z}_t w_{t+1}] = 0$.) Since

$$
(4.6)\qquad \begin{aligned}
\|\boldsymbol{Z}_T^{-1}(\boldsymbol{Z}_T - \boldsymbol{Z}_{T-1})\|^2 &= \|\boldsymbol{Z}_T^{-1}\boldsymbol{z}_T\boldsymbol{z}_T'\|^2 \\
&= \lambda_{max}(\boldsymbol{z}_T\boldsymbol{z}_T'\boldsymbol{Z}_T^{-2}\boldsymbol{z}_T\boldsymbol{z}_T') \\
&= \boldsymbol{z}_T'\boldsymbol{z}_T\boldsymbol{z}_T'\boldsymbol{Z}_T^{-2}\boldsymbol{z}_T \\
&\geq (\boldsymbol{z}_T'\boldsymbol{Z}_T^{-1}\boldsymbol{z}_T)^2 \\
&= \lambda_T^2,
\end{aligned}
$$

where the inequality follows by the Cauchy-Schwarz inequality, the condition (iii) implies $\lambda_t \to 0$ a.s.

We define a process $\{\boldsymbol{z}_t\}$ as *stable* if it satisfies

$$
(4.7)\qquad \lambda_{max}(\boldsymbol{Z}_t) = O(\lambda_{min}(\boldsymbol{Z}_t)) = O(t) \qquad \text{a.s.}
$$

and

$$
(4.8)\qquad \|\boldsymbol{z}_t\|^2 = o(t) \qquad \text{a.s.}
$$

Then, since $\lambda_t \leq \|\boldsymbol{z}_t\|^2/\lambda_{min}(\boldsymbol{Z}_t)$, we obtain

Corollary 4.3:
Suppose that (A.1)–(A.3) *hold and* $a < 1/2$. *Suppose furthermore that* $\{\boldsymbol{z}_t\}$ *is a stable process, thus* (4.7) *and* (4.8) *hold. Then* $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ *a.s.*

It is an interesting question whether $a < 1/2$ is a necessary condition for a.s. convergence or just a sufficient one. We do not know. The coincidence of the condition $a < 1/2$ in our result and the result of FOURGEAUD ET AL. (1986) may rely on the fact that both approaches are, more or less, based on algebraic properties of the involved time series. But it is as well possible that it relies on deeper mathematical or probabilistic properties.

To conclude this section we want to remark that the preceding results hold also if the disturbance terms satisfy (3.30) instead of (2.2). Clearly, in that case the rate of convergence reduces to $\|\boldsymbol{\theta}_{t+1} - \bar{\boldsymbol{\theta}}\|^2 = O((\log t)^{1+\delta}/t)$ a.s. for every $\delta > 0$.

Non-Stable Processes

Now we consider processes $\{\boldsymbol{z}_t\}$ which do not satisfy (4.7)and hence are not stable. Instead, we assume that

$$
(4.9)\qquad \liminf_{T\to\infty} \frac{1}{T}\lambda_{min}(\boldsymbol{Z}_T) > 0 \qquad \text{a.s. and} \qquad \lambda_{max}(\boldsymbol{Z}_T) = O(T^{\alpha}) \qquad \text{a.s.}
$$

with some $\alpha > 1$. Since $n^{-1}tr(\boldsymbol{Z}_t) \leq \lambda_{max}(\boldsymbol{Z}_t) \leq tr(\boldsymbol{Z}_t)$ this implies that some components of $\boldsymbol{z}_t$ grow with at most an algebraic (i.e. polynomial) order. In the sequel we prove a result which implies that $\lambda_t \to 0$ a.s. can hold also for these processes and Corollary 3.2 applies.

Our analysis relies on the following partitioning. Suppose that there exists a partition $\boldsymbol{z}_t' = (\boldsymbol{z}_t^{m\prime}, \ldots, \boldsymbol{z}_t^{1\prime})$ where $\boldsymbol{z}_t^i$ is $q_i \times 1$ and $\sum_1^m q_i = n$. For $i = 0, \ldots, m-1$ define $\boldsymbol{u}_t^{i\prime} = (\boldsymbol{z}_t^{m\prime}, \ldots, \boldsymbol{z}_t^{i+1\prime})$ and

$$\boldsymbol{P}_T^i := \sum_{t=0}^T \boldsymbol{u}_t^i \boldsymbol{u}_t^{i\prime}. \tag{4.10}$$

Also define

$$\boldsymbol{Q}_T^i := \sum_{t=0}^T \boldsymbol{z}_t^i \boldsymbol{z}_t^{i\prime} \tag{4.11}$$

for $i = 1, \ldots, m-1$.

The following result shows that the behaviour of the quadratic form λ_T is determined by the behaviour of the same kind of quadratic forms involving only the subvectors $\boldsymbol{u}_T^i$, $\boldsymbol{z}_T^i$ and the submatrices $\boldsymbol{P}_T^i$, $\boldsymbol{Q}_T^i$, provided that the matrix $\boldsymbol{Z}_T$ can be partitioned in an appropriate way.

Proposition 4.4:
Let $\{\boldsymbol{z}_t\}$ be a sequence of n-dimensional random vectors and $\boldsymbol{Z}_T = \sum_{t=0}^T \boldsymbol{z}_t \boldsymbol{z}_t'$. Suppose that the vectors $\boldsymbol{z}_T$ and the matrices $\boldsymbol{Z}_T$ can be partitioned as above such that

(i) $\lambda_{max}(\boldsymbol{Q}_T^i) = O\left(\lambda_{min}(\boldsymbol{P}_T^{i-1})\right)$ *a.s. for $i = 1, \ldots, m-1$,*

(ii) $\|\boldsymbol{z}_T^i\|^2 = o\left(\sum_{t=0}^T \|\boldsymbol{z}_t^i\|^2\right)$ *a.s. for $i = 1, \ldots, m-1$, and*

(iii) $\lim_{T\to\infty} \boldsymbol{u}_T^{m-1\prime}(\boldsymbol{P}_T^{m-1})^{-1}\boldsymbol{u}_T^{m-1} = 0$ *a.s.*

Then $\lambda_T \to 0$ a.s.

The proof of Proposition 4.4 relies crucially on the following result.

Lemma 4.5: (Lai/Wei (1983))
Let $\boldsymbol{A}$ be a positive definite symmetric $n \times n$ matrix. Let $\boldsymbol{A}$ be partitioned as

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{P} & \boldsymbol{H} \\ \boldsymbol{H}' & \boldsymbol{Q} \end{pmatrix} \tag{4.12}$$

where $\boldsymbol{P}$, $\boldsymbol{Q}$ are $p \times p$ and $q \times q$ matrices such that $n = p + q$. Then for every $\boldsymbol{u} \in I\!R^p$

$$\begin{pmatrix} \boldsymbol{u} \\ \mathbf{o} \end{pmatrix}' \boldsymbol{A}^{-1} \begin{pmatrix} \boldsymbol{u} \\ \mathbf{o} \end{pmatrix} \le \boldsymbol{u}'\boldsymbol{P}^{-1}\boldsymbol{u}\left(1 + \|\boldsymbol{A}\|^{-1} tr(\boldsymbol{Q})\right). \tag{4.13}$$

Proof of Proposition 4.4:
Let $\boldsymbol{u}_T^i$, $\boldsymbol{z}_T^i$, $\boldsymbol{P}_T^i$, and $\boldsymbol{Q}_T^i$ be defined as above. Then

$$\begin{aligned} \boldsymbol{z}_T'\boldsymbol{Z}_T^{-1}\boldsymbol{z}_T &= \begin{pmatrix} \boldsymbol{u}_T^1 \\ \boldsymbol{z}_T^1 \end{pmatrix}' \boldsymbol{Z}_T^{-1} \begin{pmatrix} \boldsymbol{u}_T^1 \\ \boldsymbol{z}_T^1 \end{pmatrix} \\ &\le 2\begin{pmatrix} \boldsymbol{u}_T^1 \\ \mathbf{o} \end{pmatrix}' \boldsymbol{Z}_T^{-1} \begin{pmatrix} \boldsymbol{u}_T^1 \\ \mathbf{o} \end{pmatrix} + 2\begin{pmatrix} \mathbf{o} \\ \boldsymbol{z}_T^1 \end{pmatrix}' \boldsymbol{Z}_T^{-1} \begin{pmatrix} \mathbf{o} \\ \boldsymbol{z}_T^1 \end{pmatrix} \\ &= 2\begin{pmatrix} \boldsymbol{u}_T^1 \\ \mathbf{o} \end{pmatrix}' (\boldsymbol{P}_T^0)^{-1} \begin{pmatrix} \boldsymbol{u}_T^1 \\ \mathbf{o} \end{pmatrix} + 2\begin{pmatrix} \mathbf{o} \\ \boldsymbol{z}_T^1 \end{pmatrix}' (\boldsymbol{P}_T^0)^{-1} \begin{pmatrix} \mathbf{o} \\ \boldsymbol{z}_T^1 \end{pmatrix}. \end{aligned} \tag{4.14}$$

Since $\boldsymbol{Z}_T = \boldsymbol{P}_T^0$ we obtain for the latter quadratic form by (i) and (ii)

$$
\begin{aligned}
\begin{pmatrix} \mathbf{o} \\ \boldsymbol{z}_T^1 \end{pmatrix}' (\boldsymbol{P}_T^0)^{-1} \begin{pmatrix} \mathbf{o} \\ \boldsymbol{z}_T^1 \end{pmatrix} &\leq \|\boldsymbol{z}_T^1\|^2 \frac{1}{\lambda_{min}(\boldsymbol{P}_T^0)} \\
&= O\left(\frac{\|\boldsymbol{z}_T^1\|^2}{tr(\boldsymbol{Q}_T^1)} \frac{\lambda_{max}(\boldsymbol{Q}_T^1)}{\lambda_{min}(\boldsymbol{P}_T^0)} \right) \qquad (4.15) \\
&= o(1) \qquad \text{a.s.}
\end{aligned}
$$

since $tr(\boldsymbol{Q}_T^1) \leq q_1 \lambda_{max}(\boldsymbol{Q}_T^1)$.

For the former quadratic form Lemma 4.5 implies

$$
\begin{aligned}
\begin{pmatrix} \boldsymbol{u}_T^1 \\ \mathbf{o} \end{pmatrix}' (\boldsymbol{P}_T^0)^{-1} \begin{pmatrix} \boldsymbol{u}_T^1 \\ \mathbf{o} \end{pmatrix} &\leq \boldsymbol{u}_T^{1\prime} (\boldsymbol{P}_T^1)^{-1} \boldsymbol{u}_T^1 \left(1 + \|\boldsymbol{P}_T^0\|^{-1} tr(\boldsymbol{Q}_T^1)\right) \\
&= O\left(\boldsymbol{u}_T^{1\prime} (\boldsymbol{P}_T^1)^{-1} \boldsymbol{u}_T^1 \left(1 + \frac{\lambda_{max}(\boldsymbol{Q}_T^1)}{\lambda_{min}(\boldsymbol{P}_T^0)}\right)\right) \qquad (4.16) \\
&= O\left(\boldsymbol{u}_T^{1\prime} (\boldsymbol{P}_T^1)^{-1} \boldsymbol{u}_T^1 \right).
\end{aligned}
$$

Since

$$
\boldsymbol{P}_T^1 = \begin{pmatrix} \boldsymbol{P}_T^2 & \boldsymbol{H}_T^2 \\ (\boldsymbol{H}_T^2)' & \boldsymbol{Q}_T^2 \end{pmatrix} \qquad (4.17)
$$

with some matrix $\boldsymbol{H}_T^2$ we can proceed inductively in the same way and finally obtain

$$
\boldsymbol{z}_T' \boldsymbol{Z}_T^{-1} \boldsymbol{z}_T = O\left(\boldsymbol{u}_T^{m-1\prime} (\boldsymbol{P}_T^{m-1})^{-1} \boldsymbol{u}_T^{m-1} \right) + o(1) \qquad \text{a.s.} \qquad (4.18)
$$

Now condition (iii) implies the desired result. ■

Proposition 4.4 can be applied to the following kind of situations. Suppose that $\boldsymbol{z}_t$ consists of random variables with different asymptotic behaviour, for example, $\boldsymbol{z}_t = (\boldsymbol{z}_t^{3\prime}, \boldsymbol{z}_t^{2\prime}, \boldsymbol{z}_t^{1\prime})'$ such that $\boldsymbol{z}_t^1$ follows a stable process and $\boldsymbol{z}_t^2, \boldsymbol{z}_t^3$ show some different trends. In that case we have $m = 3$.

In the first step we set $\boldsymbol{u}_t^1 = (\boldsymbol{z}_t^{3\prime}, \boldsymbol{z}_t^{2\prime})'$. Then condition (i) of Proposition 4.4 requires that

$$
\lambda_{max}\left(\sum_{t=0}^{T} \boldsymbol{z}_t^1 \boldsymbol{z}_t^{1\prime} \right) = O\left(\lambda_{min}(\boldsymbol{Z}_T)\right) \qquad \text{a.s.} \qquad (4.19)
$$

and condition (ii) requires that

$$
\|\boldsymbol{z}_T^1\|^2 = o\left(\sum_{t=0}^{T} \|\boldsymbol{z}_t^1\|^2 \right) \qquad \text{a.s.} \qquad (4.20)
$$

Suppose that $\lambda_{min}(\boldsymbol{Z}_T)$ is of minimum order $O(T)$ a.s. Then (4.19) is satisfied while (4.20) is satisfied since $\boldsymbol{z}_t^1$ follows a stable process.

In the second step we set $\boldsymbol{u}_t^2 = \boldsymbol{z}_t^3$. Then condition (i) requires that

$$
\lambda_{max}\left(\sum_{t=0}^{T} \boldsymbol{z}_t^2 \boldsymbol{z}_t^{2\prime} \right) = O\left(\lambda_{min}\left(\sum_{t=0}^{T} \begin{pmatrix} \boldsymbol{z}_t^3 \\ \boldsymbol{z}_t^2 \end{pmatrix} \begin{pmatrix} \boldsymbol{z}_t^3 \\ \boldsymbol{z}_t^2 \end{pmatrix}' \right) \right) \qquad \text{a.s.} \qquad (4.21)
$$

and condition (ii) requires that

$$\|z_T^2\|^2 = o\left(\sum_{t=0}^{T} \|z_t^2\|^2\right) \qquad \text{a.s.} \tag{4.22}$$

While (4.22) it is generally not very difficult to verify the verification of (4.21) can cause some problem since the minimum order of the minimum eigenvalue is generally difficult to determine. Nevertheless, in some applications (4.21) can be shown with some effort, for example if $\boldsymbol{z}_t^2$ and $\boldsymbol{z}_t^3$ are two different trends. See also Example 4.6.

Finally, condition (iii) requires that

$$\boldsymbol{z}_T^{3\prime} \left(\sum_{t=0}^{T} \boldsymbol{z}_t^3 \boldsymbol{z}_t^{3\prime}\right)^{-1} \boldsymbol{z}_T^3 \longrightarrow 0 \qquad \text{a.s.} \tag{4.23}$$

Although this is, basically, the same kind of problem as the original one (i.e. to show that $\lambda_t \to 0$) it is generally easier to solve since $\boldsymbol{z}_t^3$ is of a lower dimension than $\boldsymbol{z}_t$, possibly of dimension one. If the minimum and maximum eigenvalues of the matrix in (4.23) are of the same order then (4.23) reduces to a problem like in (4.20) and (4.22).

Now we present two examples of non-stable processes which satisfy the assumptions of Corollary 3.2.

Example 4.6:
Let x_t be a real valued stochastic process such that

$$\bar{x}_T := \frac{1}{T}\sum_{t=0}^{T} x_t \longrightarrow \mu \qquad \text{a.s.} \tag{4.24}$$

and

$$\frac{1}{T}\sum_{t=0}^{T}(x_t - \bar{x}_T)^2 \longrightarrow \sigma^2 \qquad \text{a.s.} \tag{4.25}$$

with some fixed $\mu \in I\!R$, $\sigma^2 > 0$. Set $\boldsymbol{z}_t = (x_t, t)'$, thus $\boldsymbol{z}_t$ contains x_t and a linear trend. We claim that

$$\lambda_{max}(\boldsymbol{Z}_T) = O(T^3) \qquad \text{a.s.} \tag{4.26}$$

and

$$\liminf_{T\to\infty} \frac{1}{T}\lambda_{min}(\boldsymbol{Z}_T) > 0 \qquad \text{a.s.} \tag{4.27}$$

Hence the eigenvalue condition of Corollary 3.2 will be satisfied.

In order to show (4.26) and (4.27) we define $\widetilde{\boldsymbol{Z}}_T = \frac{1}{T}\boldsymbol{Z}_T$. Since

$$\begin{aligned} \lambda_{max}(\widetilde{\boldsymbol{Z}}_T) &\leq tr(\widetilde{\boldsymbol{Z}}_T) \\ &= \frac{1}{T}\sum_{t=0}^{T} x_t^2 + \frac{1}{T}\sum_{t=1}^{T} t^2 \\ &= O(1) + \frac{(T+1)(2T+1)}{6} \qquad \text{a.s.} \end{aligned} \tag{4.28}$$

the property (4.26) is obvious. To show (4.27) we consider the inverse of $\widetilde{\boldsymbol{Z}}_T$ which is easily calculated as

$$\widetilde{\boldsymbol{Z}}_T^{-1} = \frac{1}{d_T}\begin{pmatrix} \frac{(T+1)(2T+1)}{6} & -\frac{1}{T}\sum_{t=1}^T tx_t \\ -\frac{1}{T}\sum_{t=1}^T tx_t & \frac{1}{T}\sum_{t=0}^T x_t^2 \end{pmatrix} \tag{4.29}$$

with

$$d_T = \frac{(T+1)(2T+1)}{6}\frac{1}{T}\sum_{t=0}^T x_t^2 - \left(\frac{1}{T}\sum_{t=1}^T tx_t\right)^2. \tag{4.30}$$

We shall show that

$$d_T = \frac{(T+1)(2T+1)}{6}(C+o(1)) \qquad \text{a.s.} \tag{4.31}$$

with some positive constant C and

$$\widetilde{\boldsymbol{Z}}_T^{-1} \longrightarrow \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{a.s.} \tag{4.32}$$

Hence $\lambda_{max}(\widetilde{\boldsymbol{Z}}_T^{-1}) \to C$ a.s. and (4.27) is shown.

To show (4.31) we consider the sum $\frac{1}{T}\sum_{t=1}^T tx_t$. By partial summation we obtain

$$\frac{1}{T}\sum_{t=1}^T tx_t = \bar{x}_T(T+1) - \frac{1}{T}\sum_{t=1}^T t\bar{x}_t. \tag{4.33}$$

Using assumption (4.24) it is not difficult to show that

$$\frac{1}{T}\sum_{t=1}^T t\bar{x}_t = \mu\frac{T+1}{2} + o(T) \qquad \text{a.s.} \tag{4.34}$$

as $T \to \infty$. Hence

$$\frac{1}{T}\sum_{t=1}^T tx_t = \mu\frac{T+1}{2} + o(T) \qquad \text{a.s.} \tag{4.35}$$

On the other hand we have by (4.25)

$$\left(\frac{1}{T}\sum_{t=1}^T t^2\right)\left(\frac{1}{T}\sum_{t=1}^T x_t^2\right) = \frac{(T+1)(2T+1)}{6}(\sigma^2+\mu^2) + o(T^2) \qquad \text{a.s.} \tag{4.36}$$

With some elementary calculations we finally obtain

$$d_T = \frac{(T+1)(2T+1)}{6}\left(\sigma^2 + \frac{1}{4}\mu^2 + o(1)\right) \qquad \text{a.s.} \tag{4.37}$$

Hence

$$\frac{\log \lambda_{max}(\boldsymbol{Z}_T)}{\lambda_{min}(\boldsymbol{Z}_T)} = O\left(\frac{\log T^3}{T}\right) = o(1) \qquad \text{a.s.} \tag{4.38}$$

In addition Proposition 4.4 implies $\lambda_t \to 0$ a.s. by (4.24) and Corollary 4.1. Hence $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s. if $a < 1/2$. □

Remark:
Example 4.6 can be generalized in several directions. Firstly, the trend variable in $\boldsymbol{z}_t$ may be any polynomial trend. Secondly, the variable x_t may be multivariate as long as conditions equivalent to (4.24) and (4.25) are satisfied. Finally, the vector $\boldsymbol{z}_t$ may include several trend variables, provided that they are not linear dependent. Of course, if $\boldsymbol{z}_t$ is of dimension larger than two the proof of (3.2) and (3.4) will be not as simple as in our example.

Since autoregressive processes with unit roots are quite popular, especially in macro-economics, we show in the following example that these processes satisfy the assumptions of Corollary 3.2.

Example 4.7:
Let x_t be an $AR(p)$ process, i.e.

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + e_t \tag{4.39}$$

with $\{e_t\}$ a martingale difference sequence with respect to some filtration $\{\mathcal{G}_t\}$ such that

$$\sup_{t\geq 1} \mathrm{E}[|e_t|^\alpha | \mathcal{G}_{t-1}] < \infty \qquad \text{a.s.} \tag{4.40}$$

for some $\alpha > 2$ and

$$\liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathrm{E}[e_t^2 | \mathcal{G}_{t-1}] > 0 \qquad \text{a.s} \tag{4.41}$$

Suppose furthermore that the initial values $x_0, \ldots, x_{1-p}$ are $\mathcal{G}_0$-measurable and that the characteristic polynomial

$$\pi(x) = x^p - \phi_1 x^{p-1} - \cdots - \phi_{p-1} x - \phi_p \tag{4.42}$$

possesses roots inside as well as on the unit circle. Let ρ be the largest multiplicity of all the distinct roots on the unit circle and set $\boldsymbol{x}_t = (x_t, \ldots, x_{t-p+1})'$. Then it is well-known (cf. LAI/WEI (1983) or (1985)) that

$$\lambda_{max}\left(\sum_{t=0}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right) = O\left(T^{2\rho} \log\log T\right) \qquad \text{a.s.,} \tag{4.43}$$

$$\liminf_{T\to\infty} \frac{1}{T} \lambda_{min}\left(\sum_{t=0}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right) > 0 \qquad \text{a.s.} \tag{4.44}$$

and

$$\lim_{T\to\infty} \boldsymbol{x}_T' \left(\sum_{t=0}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \boldsymbol{x}_T = 0 \qquad \text{a.s.} \tag{4.45}$$

Now let $\boldsymbol{z}_t = \boldsymbol{x}_t$ and suppose that $\mathcal{G}_t \subset \mathcal{F}_t$ for all $t \geq 0$. Then the assumptions of Corollary 3.2 are satisfied and we can conclude that $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s. if $a < 1/2$. □

We wish to remark that Example 4.7 can be generalized without further considerations to multivariate autoregressive processes with possibly multiple unit roots.

Chapter 5

Conclusions and Remarks

In this study we have shown that agents can learn to form rational expectations even in an environment in which the rational expectations equilibrium is a non-stationary process. This result extends the class of models for which the boundedly rational learning approach applies and indicates that the presumption that learning, in order to be successful, requires a stationary environment is misleading.

From the mathematical point of view our result is not surprising since it is well known that the OLS estimator in linear regression models is strongly consistent under more general conditions than the 'classical' ones which require that the regressors are stationary and ergodic[1]. Therefore it was to be expected that this feature carries over to linear models with forecast feedback if the effect of forecast feedback is sufficiently small.

But also from the economic point of view it is not surprising that agents can learn rational expectations in non-stationary models since learning in the context of the boundedly rational learning approach means learning of parameter values, thus learning of the law of motion of the endogenous variable. Although in our model the REE can be a non-stationary process the law of motion of the REE is time invariant and the law of motion of $\{y_t\}$ is time dependent only because agents persistently change their predictive behaviour during the learning phase. Therefore, if the feedback between the perceived law of motion and the true law of motion is damped, it is not surprising that agents can learn this law of motion by following a learning procedure which would be appropriate to learn the law of motion in the REE.

Since agents are learning about relationships between economic variables which they observe it is not necessary that the variables themselves are stationary. It is only necessary that these relationships are 'stable' and the observed variables satisfy the requirements of the information extraction mechanism by which agents learn. As mentioned above the OLS procedure is able to deal with non-stationary variables. Other learning procedures, like the Stochastic Gradient (SG) procedure[2] analyzed by ZENNER (1994b), lack this ability[3].

To summarize this discussion we want to point out that our results do not actually refute

[1]If we set $a = 0$ in our model, thus if there is no forecast feedback, then Corollary 3.2 gives the strong consistency of the OLS estimates in a linear regression model satisfying the eigenvalue condition (3.4). Notice that if $a = 0$ the condition (3.2) is always satisfied. As LAI/WEI (1982) have shown the condition (3.4) is in some sense the weakest possible ensuring strong consistency.

[2]The SG procedure is a parameter estimation procedure which possesses a recursive representation like (3.14) with the matrix Z_t replaced by its trace. It is well known and popular in the theory on recursive identification and control since it is computationally less demanding than the OLS procedure and sometimes easier to analyze theoretically.

[3]On the other hand ZENNER (1994b) recently showed for the SG learning procedure that if $|a| < 1$ and the characteristic polynomial of the REE possesses roots only inside the unit circle then $\{z_t\}$ is a stable process whenever the exogenous variables follow a stable process.

the intuitively appealing idea that learning can be successful only in an environment which is stable or regular in some sense. What they show is that one has to be careful in using terms like 'stationary' which possess different meanings in different contexts. In statistics 'stationarity' means a specific and well defined property of a time series while in colloquial language and in economics 'stationary' is sometimes used synonymously to 'time invariant' or 'regular'.

Since the performance of the least squares learning procedures in autoregressive models with forecast feedback is not yet solved in a satisfactory manner[4] we want to add some remarks on the behaviour of the OLS procedure in our model when the random vector $\boldsymbol{z}_t$ contains lagged endogenous variables. Since we failed in verifying the assumptions of Corollary 3.2 in that case these remarks are based on some computer simulations.

Firstly, we observed that if $|a| < 1$ (thus if the *feedback function* which maps the perceived law of motion $\boldsymbol{\theta}_t$ into the true law of motion $\boldsymbol{\phi} + a\boldsymbol{\theta}_t$ is contracting) and if the REE is a stable process (thus if the characteristic polynomial of the REE possesses roots only inside the unit circle and the exogenous variables follow a stable process) agents following the OLS learning procedure cannot destabilize the system and learn the RE parameter values with probability one. Even if agents start with totally implausible initial estimates the process $\{y_t\}$ stabilizes after some transient phase and the estimates converge. This finding corresponds to the convergence result for the SG learning procedure shown by ZENNER (1994b).

Secondly, we observed the same kind of stability of rational expectations with respect to the OLS procedure if the characteristic polynomial possesses unit roots or the exogenous variables incorporate trend variables. But in the latter case $a < 1/2$ seems to be a necessary condition for a.s. convergence.

Finally, we also observed convergence if the REE possesses roots outside the unit circle. In that case the condition (3.2) is crucial since λ_t does not converge to zero but remains positive. The limit point of λ_t then depends on the roots of the REE. If this limit point satisfies the inequality (3.2) the OLS procedure seems to converge a.s. for stable exogenous variables as well as for variables with a trend.

To conclude this study we want to remark that it is straightforward to generalize the results to multivariate (simultaneous equations) models. For notational convenience and since this generalization incorporates no new ideas we decided to omit it.

[4]The ordinary differential equation (ODE) approach by LJUNG (1977) adopted by MARCET/SARGENT (1989a,b) seems to be widely accepted in economic theory (see, e.g., WOODFORD (1990), MOORE (1993), and FUHRER/HOOKER (1993)) although it has a few short-comings. The main problem consists in justifying the use of the so-called 'projection facility' on the level of individual agents. Without this facility no global convergence results can be obtained for autoregressive models.

ZENNER (1994a) provides global convergence results of the OLS procedure in autoregressive models without using this facility but the underlying analysis is restricted to univariate generalized AR(1) models.

Chapter 6

References

BRAY, M. (1982)
Learning, Estimation, and the Stability of Rational Expectations
Journal of Economic Theory, **26**, 318–339

BRAY, M. (1983)
Convergence to Rational Expectations Equilibrium; in R. Frydman and E.S. Phelps (eds): *Individual Forecasting and Aggregate Outcomes,* 123–132
Cambridge University Press, Cambridge

BRAY, M.M. and N.E. SAVIN (1986)
Rational Expectations Equilibria, Learning, and Model Specification
Econometrica, **54**, 1129–1160

CHOW, Y.S. (1965)
Local Convergence of Martingales and the Law of Large Numbers
Ann. Math. Stat., **36**, 552–558

FOURGEAUD, C., C. GOURIEROUX, and J. PRADEL (1986)
Learning Procedures and Convergence to Rationality
Econometrica, **54**, 845–868

FUHRER, J.C. and M.A. HOOKER (1993)
Learning about Monetary Regime Shifts in an Overlapping Wage Contract Model
Journal of Economic Dynamics and Control, **17**, 531–553

GRANDMONT, J.-M. (1992)
Aggregation, Learning and Rationality
Technical Report No. 52, Stanford University

KNOPP, K. (1964)
Theorie und Anwendung der unendlichen Reihen
Springer-Verlag, Berlin

KOTTMANN, T. (1990)
Learning Procedures and Rational Expectations in Linear Models with Forecast Feedback
Ph.D. Thesis, Universität Bonn

KUAN, C.-M. and H. WHITE (1994)
Adaptive Learning with Nonlinear Dynamics Driven by Dependent Processes
Econometrica, **62**, *1087–1114*

LAI, T.L. (1989)
Extended Stochastic Lyapunov Functions and Recursive Algorithms in Linear Stochastic Systems; in N. Christopeit et al. (eds): *Stochastic Differential Systems, Lecture Notes in Control and Information Sciences,* **126**, 206–220
Springer-Verlag, Heidelberg

LAI, T.L. and C.Z. WEI (1982)
Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems
The Annals of Statistics, **10**, 154–166

LAI, T.L. and C.Z. WEI (1983)
Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters
Journal of Multivariate Analysis, **13**, 1–23

LAI, T.L. and C.Z. WEI (1985)
Asymptotic Properties of Multivariate Weighted Sums with Applications to Stochastic Regression in Linear Dynamic Systems; in P. R. Krishnaiah (ed): *Multivariate Analysis – VI,* 375–393
North-Holland, Amsterdam

LJUNG, L. (1977)
Analysis of Recursive Stochastic Algorithms
IEEE Transactions on Automatic Control, **AC-22**, 551–575

MARCET,A. and T.J. SARGENT (1989a)
Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models
Journal of Economic Theory, **48**, 337–368

MARCET,A. and T.J. SARGENT (1989b)
Convergence of Least-Squares Learning in Environments with Hidden State Variables and Private Information
Journal of Political Economy, **97**, 1306–1322

MARCET, A. and T.J. SARGENT (1989c)
Least-Squares Learning and the Dynamics of Hyperinflation; in W. A. Barnett, et al. (eds): *Economic Complexity: Chaos, Sunspots, Bubbles, and Nonlinearity,* 119–137
Cambridge University Press, Cambridge

MOHR, M. (1990)
Asymptotic Theory for Ordinary Least Squares Estimators in Regression Models with Forecast Feedback
Ph.D. Thesis, Universität Bonn

MOORE, B.J. (1993)
Least-Squares Learning and the Stability of Equilibria with Externalities
Review of Economic Studies, **60**, 197–208

MUTH, J.F. (1961)
Rational Expectations and the Theory of Price Movements
Econometrica, **29**, 315–335

ROBBINS, H. and D. SIEGMUND (1971)
A Convergence Theorem for Non Negative Almost Supermartingales and some Applications; in J. S. Rustagi (ed): *Optimizing Methods in Statistics,* 233–257
Academic Press, New York

WOODFORD, M. (1990)
Learning to Believe in Sunspots
Econometrica, **58**, 277–307

ZENNER, M. (1992a)
Performance of Least Squares Learning in Autoregressive Models with Forecast Feedback – The Deterministic Case
Discussion Paper B–224, Universität Bonn

ZENNER, M. (1992b)
Performance of Least Squares Learning in Autoregressive Models with Forecast Feedback – The Stochastic Case
Discussion Paper B–225, Universität Bonn

ZENNER, M. (1994a)
Prediction Error Learning in Univariate Generalized Autoregressive Models
Discussion Paper B–273, Universität Bonn

ZENNER, M. (1994b)
Prediction Error Learning and Rational Expectations in Autoregressive Models with Forecast Feedback
Discussion Paper B–303, Universität Bonn