SFB 303, Projektbereich B

Discussion Paper No. B–303

Prediction Error Learning and Rational Expectations in Autoregressive Models with Forecast Feedback

Markus Zenner[1]

December 1994

[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

The rational expectations hypothesis is supported if rational expectations are stable with respect to reasonable learning procedures. We consider the Stochastic Gradient–Algorithm as a boundedly rational learning procedure in an univariate ARX–Model with forecast feedback. We prove that whenever there exists a stable rational expectations equilibrium and the influence of the forecast feedback is limited the learning agents cannot destabilize the model and learn to form rational expectations with probability one.

Jel Code: C22, C40, C62, D83

Key words: Rational expectations, learning, model specification

Contents

Chapter 1

Introduction

In many economic models the agents' expectations matter and are an important source for the model dynamics. But since expectations are generally unobservable one has to complete models in which expectations matter by introducing an expectations formation scheme. Clearly, any particular choice of such an expectations formation scheme is ad-hoc and open to criticism. In his famous paper MUTH (1961) introduced the rational expectations hypothesis (REH) which suggests that the agents' expectations of a variable coincide with the mathematical conditional distribution of that variable given the observable history of the model.

In spite of many attractive features the REH has a few substantial shortcomings. First of all it is extremely demanding in view of the agents' knowledge and understanding of their economic surrounding. In order to calculate rational expectations the agents have to know not only the reduced form of the model but also the exact values of its reduced form parameters. To support the REH it was argued that agents may learn somehow to form rational expectations by repeatedly observing their economic environment and collecting information. Thus rational expectations could be understood as limit outcomes of some 'reasonable' learning procedure.

Although this idea is simple and intuitively appealing its formalization and analytical study is quite difficult. Leaving aside the problem of forming expectations about other agents' expectations and the resulting infinite regress of expectations the problem of learning to become rational can be formulated as the problem of consistency of certain parameter estimates. Because of the effect of forecast feedback, i.e., the effect that agents' expectations influence via their economic actions the time series they use for their estimation, the convergence results of the classical statistical theory cannot be applied directly since, at least, some of the time series involved become non-stationary during the period of learning.

The first analytical study of the ordinary least squares (OLS)-learning procedure was carried out by M. BRAY (1982), followed by BRAY/SAVIN (1986), FOURGEAUD ET AL. (1986), MARCET/SARGENT (1989a,b,c), MOHR (1990), KOTTMANN (1990) and others. Apart from MARCET/SARGENT all studies confined themselves to the (simpler) case that the agents' expectations are based only on exogenous variables. For this case, finally, KOTTMANN and MOHR developed a complete and satisfactory theory based on stochastic approximation results by WALK (1985).

The more difficult case when agents' expectations are also based on lagged endogenous variables was first considered by CYERT/DEGROOT (1974) in a Monte-Carlo study. The first analytical results were achieved by MARCET/SARGENT (1989a,b,c). Their analysis is based on the ordinary differential equation (ODE) approach developed by LJUNG (1977) within the theory on recursive identification. Nevertheless, the results of MARCET/SARGENT are not fully satisfactory. Firstly, the analysis of LJUNG is not always convincing in the light of rigorous mathematical standards (although we believe that his results are correct). Secondly, the ODE approach, which suggests that a trajectory of a recursive estimation procedure can be approximated in a suitable manner by a trajectory of an ordinary differential equation, is basically a local concept since the approximation is only valid in the neighborhood of the equilibrium value. To obtain global convergence results LJUNG uses a certain technical trick, called 'projection facility', which is appropriate in an engineering context but not necessarily in the context of learning procedures in economics. Without this feature the analysis of MARCET/SARGENT remains a local analysis and the scope of a local convergence analysis is always limited. In addition, the ODE approach relies crucially on some stability assumption for the endogenous variable which is often not fulfilled by economic models.

In this paper we study the general case which covers the case of lagged endogenous variables among the variables used by agents for their predictions (the autoregressive case). We adopt a martingale based approach developed in ZENNER (1992b) and ZENNER (1994) to prove consistency of a recursive estimation procedure known as the Stochastic Gradient (SG)-Algorithm. This algorithm possesses a recursive structure similar to the one of the well-known OLS-algorithm and can be understood as a prediction error based learning procedure.

We will show that the SG-algorithm estimates converge with probability one to the rational expectations parameters under mild assumptions which stipulate that the influence of the agents' predictions is not too strong and that the rational expectations equilibrium (REE) is a stable process. If the REE is an unstable process then we can show that in some cases at least some components of the estimates converge to the respective components or the rational expectations parameter. These results are completey new and have no counterpart in the literature.

For the non-autoregressive case, the case that agents' expectations are based only on exogenous variables, we obtain a convergence result which relies on assumptions slightly weaker than the assumptions usually employed and therefore covers a broader class of exogenous processes.

The paper is arranged as follows. Chapter 2 presents the model setup and the assumptions our analysis is based upon. In Chapter 3 we develop our convergence analysis. The main convergence result will be formulated in terms of the order of the maximum and minimum eigenvalue of the matrix of moments which are endogenously given in the autoregressive case. In Chapter 4 we determine the order of the eigenvalues for the autoregressive case. We will show that the process of the endogenous variables will be stable whenever the REE is a stable autoregressive process. In Chapter 5 we show some convergence results. As an application we finally present a simple economic example which shows that the existence of learning agents in economic models can have a stabilizing influence.

Chapter 2

The Model

The model we consider is 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 endogenous variable which we assume to be real valued,
- $\boldsymbol{z}_t$ is an n-dimensional random vector[1] which may contain as well lagged endogenous variables as exogenous variables, i.e., $\boldsymbol{z}'_t = (\boldsymbol{y}'_t, \boldsymbol{x}'_t)$ with $\boldsymbol{y}'_t = (y_t, \ldots, y_{t-p+1})$, $\boldsymbol{x}'_t = (x_{t,1}, \ldots, x_{t,q})$, $p \geq 0, q \geq 0$, and $n = p + q$,
- $y_0, \ldots, y_{1-p}$ are the initial values of the endogenous variable,
- 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 call such a model an *ARX model with additive forecasts* or an *ARX model with linear forecast feedback.*

Before we introduce further assumptions we want to explain the main idea underlying models with forecast feedback. Possibly, this is best done giving a simple example.

Example 2.1: (Muth (1961), Cyert/DeGroot (1974))
Consider the model of an isolated market with a fixed production lag of a commodity which cannot be stored. The market equations take the form

$$\begin{aligned} C_t &= d_1 - \beta p_t && \text{(Demand)} \\ P_t &= d_2 + \gamma p^e_t + u_t && \text{(Supply)} \\ P_t &= C_t && \text{(Market equilibrium)} \end{aligned} \tag{2.2}$$

where

[1] Throughout this paper vectors and matrices will be denoted by boldface letters.

- P_t represents the number of units produced in a period lasting as long as the production lag,
- C_t is the amount consumed,
- p_t is the market price in period t,
- p_t^e is the market price expected to prevail during period t on the basis of information available through the period $t-1$,
- u_t is an error term – representing, say, variations in yields due to weather, and
- $d_1 - d_2 > 0$, $\gamma > 0$, $\beta > 0$ are fixed parameters.

Eliminating the quantity variables leads to the reduced form equation

$$p_t = \frac{d_1 - d_2}{\beta} - \frac{\gamma}{\beta} p_t^e - \frac{1}{\beta} u_t \tag{2.3}$$

which is of the form of (2.1) with $y_t = p_t$, $\boldsymbol{z}_t = 1$, and $w_t = \beta^{-1} u_t$. □

In this example the agents (or firms) are required to make a production decision in each period. Since p_t, the market price in period t, is not known in period $t-1$ they face a decision problem under uncertainty. In Example 2.1 it is assumed implicitly that agents solve this problem by maximizing their expected profit under a quadratic cost function. This leads to the supply function in (2.2). But in order to maximize the expected profit agents must have expectations of the outcome of p_t. This motivates the following general setting.

Expectations and Predictions

We suppose that in period t agents have expectations of the unknown value of y_{t+1}, and we believe that, at least for rational agents, these expectations are best modelled by probability distributions. On the other hand, expectations are generally based on experience, i.e. on previously collected information. Therefore we assume that the history of the system (2.1) is observable by agents and that their expectations are based on the information set[2].

$$I_t := \{y_t, y_{t-1}, \ldots, y_{1-p}, x_{t,1}, \ldots, x_{t,q}, \ldots, x_{0,1}, \ldots, x_{0,q}\}. \tag{2.4}$$

We thus model expectations by conditional probability distributions. Formally, we assume that there exist random variables $\tilde{y}_{t+1}$ and a (subjective) probability measure $\tilde{\mathrm{P}}$ such that the agents' expectations of y_{t+1} can be written as

[2] Nevertheless, another choice of the information set is as well possible within this framework. For example, I_t can contain some additional sun-spot variables, or it can contain only sun-spot variables. Generally, by the choice of the information set I_t it is possible to model the degree of understanding of the agents. If agents are sophisticated their information set contains all the relevant variables while the one of less clever agents, such as noise-traders, contains only part of it.

$\tilde{\mathrm{P}}[\tilde{y}_{t+1}|I_t]$. (The variables $\tilde{y}_{t+1}$ are only auxiliary mathematical objects and do not have to correspond with any observable quantity in the economic system.)

Since agents, as part of their economic surrounding, are required to take one specific action every period it is necessary for them to derive a single decision criterion from their expectations. We assume that they use the most common criterion, the mean value of their expectations, or, mathematically speaking, the expectation of their conditional probability distribution. We call this (mathematical) expectation[3] *prediction* or *forecast* and denote it by y^e_{t+1}. Hence

$$y^e_{t+1} = \tilde{\mathrm{E}}[\tilde{y}_{t+1}|I_t]. \tag{2.5}$$

Following these lines common assumptions on the behaviour of agents, like profit maximizing as in Example 2.1, often lead to a reduced form equation of the form of (2.1).

The Rational Expectations Hypothesis

Up to now we have made no specific assumptions on the expectations of agents. They could be arbitrary conditional probability distributions. MUTH (1961, p. 316), in his famous article, suggests that "*expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant theory*", and he calls this kind of expectations 'rational expectations'. More exactly, he suggests that "*expectations of firms (or, more generally, the subjective probability distributions of outcomes) tend to be distributed, for the same information set, about the prediction of the theory (or the "objective" probability distributions of outcomes)*". Within our model this means that

$$\tilde{\mathrm{P}}[\tilde{y}_{t+1}|I_t] = \mathrm{P}[y_{t+1}|I_t] \qquad \text{a.s.} \tag{2.6}$$

We will call expectations which fulfill (2.6) *strongly rational expectations.* The probability distribution on the left-hand side is the so-called subjective distribution of outcomes and the one on the right-hand side the so-called objective distribution of outcomes. Expectations which fulfill the weaker requirement

$$\tilde{\mathrm{E}}[\tilde{y}_{t+1}|I_t] = \mathrm{E}[y_{t+1}|I_t] \qquad \text{a.s.} \tag{2.7}$$

will be called *weakly rational expectations* or simply *rational expectations*[4].

Notice that the so-called objective probability of outcomes on the right-hand side of (2.6) and its mathematical expectation on the right-hand side of (2.7)

[3] The ambiguity of the term 'expectation' has caused some confusion in the literature on rational expectations. MUTH (1961), for example, speeks of agents' expectations in terms of probability distributions but uses exclusively mathematical expectations in his analysis without quoting that there exists a difference.

To avoid this confusion we distinguish strictly between the expectations of agents, which are conditional probability distributions, and the predictions or forecasts of agents, which are the mathematical expectation of their expectations.

[4] In ZENNER (1994) we called strongly rational expectations also 'rational expectations in the sense of MUTH' but since it is not clear that MUTH was aware of the difference between strongly and weakly rational expectations we omit this here.

depends on the agents' expectations in the model. Therefore, as long as agents might have arbitrary expectations, rational expectations characterized by (2.6) or (2.7) have a meaning only for an outside observer. To have a meaning also inside the model one has to claim that (2.6) and (2.7) hold when agents have already rational expectations. Hence, rational expectations are solutions of a fixed point problem and, consequently, the economic system in which agents have rational expectations is called to be in a *rational expectations equilibrium (REE).*

In our model the agents' expectations of y_{t+1} are rational expectations whenever they induce

$$y^e_{t+1} = \boldsymbol{\phi}' \boldsymbol{z}_t + a y^e_{t+1} \qquad \text{a.s.} \tag{2.8}$$

provided that $\mathrm{E}[w_{t+1}|I_t] = 0$ a.s., and it is immediate that

$$y^e_{t+1} = \frac{1}{1-a} \boldsymbol{\phi}' \boldsymbol{z}_t \qquad \text{a.s.} \tag{2.9}$$

whenever $a \neq 1$.

Rational Expectations and Learning Procedures

The rational expectations hypothesis (REH) of Muth (1961) suggests that agents have rational expectations, more exactly, it suggests that the average or market expectation is rational. Although this hypothesis is superior to every ad-hoc assumption concerning expectations since it incorporates two main concepts of economic theory, namely the concept of rationality and the concept of equilibrium, it is appropriate to ask how plausible it is. In our model the REH implies that the agents' predictions are given by (2.9). If agents know equation (2.1) and the parameters a and $\boldsymbol{\phi}$ they could calculate these rational expectations and the system described by (2.1) and (2.9) is in a rational expectations equilibrium. But in general neither (2.1) nor the parameters a and $\boldsymbol{\phi}$ will be known exactly by the agents, hence their expectations will not be rational, at least on the short-run.

But what happens on the long-run? Is it possible that agents learn to form rational expectations in following a simple and reasonable learning procedure based on observations of the history of $\boldsymbol{z}_t$? This is the question we are concerned with in this paper.

The mathematical modeling of learning processes is a quite complex and difficult problem and the subject of current research in economics as well as in information sciences, psychology, or biology. There exist a wide variety of different approaches (e.g. genetic algorithms, neuronal networks as two prominent examples in recent research), all of them facing the same problem of finding a good compromise between realism and mathematical feasability. The approach we adopt is quite simple.

As Pesaran (1987, p. 32) points out, "*there is no doubt that individuals do learn from their own experience as well as from the experience of others. Generally speaking, learning takes place through two separate but closely connected* *mechanisms, namely* repetition, *and* understanding". Therefore we assume that the agents in our model have in mind a (not necessarily fully specified) model which they believe to describe the situation they face. We thus assume that they understand, at least roughly, how the economic system which they are part of works but are uncertain of the precise system and the exact values of the parameters. They try to learn these values by repetition, i.e. by a repeated trial-and-error scheme. We model this 'learning by repetition' by a special first order adaptive prediction error mechanism.

Assumption (A.1): (Behavioural Assumption)
The predictions y^e_{t+1} made by agents at time t on the basis of the information set I_t are given as

$$y^e_{t+1} = \boldsymbol{\theta}_t{}'\boldsymbol{z}_t, \qquad t \geq 0. \tag{2.10}$$

The coefficients $\boldsymbol{\theta}_t \in I\!R^n$ are given recursively by

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + r_t^{-1}\boldsymbol{z}_t(y_{t+1} - \boldsymbol{\theta}_t{}'\boldsymbol{z}_t), \qquad t \geq 0, \tag{2.11}$$

with some initial value $\boldsymbol{\theta}_0$ which may be random and $r_t := \sum_{s=0}^{t} \boldsymbol{z}'_s\boldsymbol{z}_s$. □

The estimation procedure given by (2.11) is known in the literature on recursive identification and control as the *Stochastic Gradient (SG) Algorithm* for the reason that the time t adjustment

$$\Delta\boldsymbol{\theta}_{t+1} := \boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t = r_t^{-1}\boldsymbol{z}_t(y_{t+1} - \boldsymbol{\theta}_t{}'\boldsymbol{z}_t) \tag{2.12}$$

is, up to an scaling factor, equal to the negative gradient of the squared prediction error $(y_{t+1} - \boldsymbol{\theta}_t{}'\boldsymbol{z}_t)^2$. Therefore the SG-algorithm is also known as *steepest descent algorithm* since its adjustments are in direction of the steepest descent of the squared prediction error.

Although the learning procedure proposed by Assumption (A.1) is motivated by its mathematical suitability it is also plausible in view of agents' behaviour. Suppose that the agents believe in the auxiliary model

$$y_{t+1} = \boldsymbol{\theta}'\boldsymbol{z}_t + e_{t+1} \tag{2.13}$$

with an unknown constant (or slowly varying) parameter $\boldsymbol{\theta}$ and a zero mean disturbance term e_{t+1} which is independent from the information set I_t. If the (hypothetical) parameter $\boldsymbol{\theta}$ is known by agents at time t their expectations of y_{t+1} will be given as $\tilde{\mathrm{P}}[\boldsymbol{\theta}'\boldsymbol{z}_t + e_{t+1}|I_t]$ and, consequently, their prediction will be

$$y^e_{t+1} = \tilde{\mathrm{E}}[\boldsymbol{\theta}'\boldsymbol{z}_t + e_{t+1}|I_t] = \boldsymbol{\theta}'\boldsymbol{z}_t \qquad \text{a.s.} \tag{2.14}$$

Since the parameter $\boldsymbol{\theta}$ is not known by agents it is reasonable for them to estimate its value by a statistical estimation procedure, for instance by an ordinary least squares (OLS) regression[5] or by the SG-algorithm.

[5] If z_t is univariate the OLS-algorithm and the SG-algorithm coincide. Otherwise the OLS-estimates fulfill a recursion like (2.11) with r_t^{-1} replaced by $(\sum_0^t z_s z_s')^{-1}$.

Unlike to the OLS-procedure the computation of the SG-algorithm estimates requires no matrix inversion and could be performed with a simple pocket calculator. In addition, the SG-algorithm formalizes the intuitively appealing idea of adjusting the estimates recursively in the light of the previous prediction error, $y_{t+1} - \boldsymbol{\theta}_t{}' \boldsymbol{z}_t$. Therefore the SG-algorithm can be understood as a stylized version of real life trial-and-error learning.

Notice that although the agents believe in a generally misspecified model (the model (2.13) is correctly specified only in the REE) they are quite sophisticated in the sense that their model is based on the correct time series and also supposes a linear relationship between the variables since under Assumption (A.1) the true model (2.1) turns out to be

$$(2.15) \qquad y_{t+1} = (\boldsymbol{\phi} + a\boldsymbol{\theta}_t)' \boldsymbol{z}_t + w_{t+1}.$$

Notice furthermore that agents believing in the model (2.13) neglect the fact of forecast feedback. They do not care about the other agents' behaviour and expectations and regard the times series they face as exogenously given. Therefore these kind of agents are called *boundedly rational.* The usefulness of boundedly rational learning models was intensively discussed (see, e.g., FRYDMAN/PHELPS (1983) or PESARAN (1987)) and we do not want to go into that discussion here. We only want to remark that the assumption of boundedly rational agents reduces the mathematical complexity of the resulting model to such a degree that proper results can be achieved with some effort.

The second assumtion concerns the disturbance terms and the dependence structure between the variables.

Assumption (A.2): (Stochastic Assumption)
The disturbance terms $\{w_t\}_{t\geq 1}$ form a martingale difference sequence with respect to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ such that

$$(2.16) \quad \mathrm{E}[w_{t+1}^2 | \mathcal{F}_t] \leq \sigma^2 \qquad \text{a.s. and} \qquad \sup_{t\geq 0} \mathrm{E}[|w_{t+1}|^{2+\delta} | \mathcal{F}_t] < \infty \quad \text{a.s.}$$

with some constants $\sigma^2 < \infty$, $\delta > 0$, and

$$(2.17) \qquad \liminf_{t\geq 0} \mathrm{E}[w_{t+1}^2 | \mathcal{F}_t] > 0 \qquad \text{a.s.}$$

We assume furthermore that $\boldsymbol{x}_t$ is $\mathcal{F}_{t-p-1}$-measurable for all $t \geq p+1$ and $\mathcal{F}_0$-measurable otherwise, and that $\boldsymbol{\theta}_0$ and $\boldsymbol{y}_0$ are $\mathcal{F}_0$-measurable. □

Assumption (A.2) is standard in modern econometrics. It generalizes the assumption that $\{w_t\}$ is a white noise sequence with bounded $(2+\delta)$th moments. The conditions (2.16), (2.17) ensure that the disturbance terms introduce enough but not too much stochastic fluctuation into the model. The measurability conditions ensure that y_t and $\boldsymbol{x}_t$ are $\mathcal{F}_t$-measurable and, since

$\sigma(I_t) \subset \mathcal{F}_t$, $\mathrm{E}[w_{t+1}|I_t] = 0$ a.s. Hence there is no effect of the disturbance terms which is foreseeable by agents.

Notice that, provided that $a \neq 1$, the rational expectations with respect to the information set I_t are uniquely determined as the expectations which give rise to the predictions

$$y^*_{t+1} := \mathrm{E}[y_{t+1}|I_t] = \frac{1}{1-a}\boldsymbol{\phi}'\boldsymbol{z}_t. \tag{2.18}$$

Hence if agents' predictions are given as in Assumption (A.1), thus $y^e_{t+1} = \boldsymbol{\theta}_t{}'\boldsymbol{z}_t$, then they are rational if and only if $\boldsymbol{\theta}_t = \bar{\boldsymbol{\theta}} := (1-a)^{-1}\boldsymbol{\phi}$. So the question whether agents can learn to form rational expectations with the aid of the SG-algorithm is equivalent to the question whether the estimates $\boldsymbol{\theta}_t$ converge to the rational expectations parameter $\bar{\boldsymbol{\theta}}$.

Finally, we want to introduce an assumption concerning the exogenous variables which is necessary for some results.

Assumption (A.3): (Exogenous Inputs)
The exogenous variables $\boldsymbol{x}_t = (x_{t,1}, \ldots, x_{t,q})'$ have the following properties:

$$tr\left(\sum_{t=0}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right) = O(T) \qquad \text{a.s.} \tag{2.19}$$

and

$$\liminf_{T\to\infty} \frac{1}{T}\lambda_{min}\left(\sum_{t=0}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right) > 0 \qquad \text{a.s.} \tag{2.20}$$

where $tr(\boldsymbol{A})$ denotes the trace and $\lambda_{min}(\boldsymbol{A})$ the minimum eigenvalue (in modulus) of the matrix $\boldsymbol{A}$. □

This assumption holds for a large class of stochastic (as well as deterministic) processes including all covariance stationary and ergodic processes, such as stable ARMA-processes, ARCH-processes, i.i.d. sequences, but also some non-stationary processes. It is more general than the assumptions usually employed in this context[6] which require that $\frac{1}{T}\sum_1^T \boldsymbol{x}_t\boldsymbol{x}_t'$ converges a.s. towards some positive definite limit matrix.

[6] See, e.g., Bray/Savin (1986), Marcet/Sargent (1989a,b,c), Kottmann (1990), Mohr (1990), and Chang et al. (1991a,b,c).

Chapter 3

Convergence Analysis

In this chapter we develop our convergence analysis. In a first step we use the almost supermartingale property of the process $\{\|\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}\|^2\}$ to show that $\{\boldsymbol{\theta}_t\}$ remains bounded a.s. under certain conditions. This property holds regardless of any assumptions on the time series $\{\boldsymbol{z}_t\}$ but the measurability condition of Assumption (A.2). Then we show that $\{\boldsymbol{\theta}_t\}$ converges a.s. towards $\bar{\boldsymbol{\theta}}$ if the times series $\{\boldsymbol{z}_t\}$ has the so-called persistent excitation (PE) property. In the following chapter we analyse under what conditions the time series $\{\boldsymbol{z}_t\}$ generated by the model (2.1) has this property. If $\{\boldsymbol{z}_t\}$ fails to have the (PE) property then we can show that at least some components of $\boldsymbol{\theta}_t$ converge towards the respective components of $\bar{\boldsymbol{\theta}}$.

Our approach is closely related to the martingale difference approach of the modern theory of system identification and control (see, e.g. Lai (1989) and Chan/Guo (1991)) but due to the forecast feedback in our model we cannot apply the consistency results of that theory. Rather, we generalize these results since, letting $a = 0$, the case of no forecast feedback is a special case in our theory.

Preliminary Results

The first auxiliary result is a simple but very useful lemma. Its first part is known as the Theorem of Abel/Dini, and the second as Theorem of Pringsheim.

Lemma 3.1: (Knopp (1964))
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

(i) $\displaystyle\sum_{t=0}^{\infty} \frac{d_t}{D_t^{\alpha}} < \infty \qquad \Longleftrightarrow \qquad \alpha > 1,$

(ii) $\displaystyle\sum_{t=1}^{\infty} \frac{d_t}{D_t D_{t-1}^{\alpha}} < \infty \qquad \Longleftrightarrow \qquad \alpha > 0.$

The second result is the keystone in the martingale difference approach, a kind of deterministic reduction[1].

Lemma 3.2: (Chow (1965), Lai/Wei (1982a))
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

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

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

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

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

$$\text{(iv)} \frac{1}{T}\sum_{t=0}^{T} \left(w_{t+1}^2 - \mathrm{E}[w_{t+1}^2|\mathcal{F}_t]\right) = O(T^{-\gamma}) \quad \textit{a.s. for all } \gamma \in \left(0, \tfrac{\delta}{2+\delta}\right).$$

If, in addition, $\mathrm{E}[w_{t+1}^2|\mathcal{F}_t] = \sigma^2 > 0$ a.s. holds, then

$$\text{(v)} \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} u_t^2 = 0 \textit{ a.s. on the event } \left[\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} u_t^2 w_{t+1}^2 = 0, \sup_{t\geq 0} |u_t| < \infty\right].$$

With the following result we formalize the idea that the disturbance terms introduce persistent stochastic fluctuation into the model.

Lemma 3.3:
Suppose that (A.1) and (A.2) hold. Then

$$\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} y_t^2 > 0 \qquad \textit{a.s.} \tag{3.1}$$

[1]As Ren (1991, p. 5) points out "*in the stochastic adaptive system theory (...) the key concern is to deal with stochastically modeled disturbances and signals. However, it is possible to fully dispense with all probabilistic assumptions and adopt a completely deterministic model of such signals and disturbances. (...) Thus, instead of starting with a probabilistic set of assumptions, one could forego all stochastic assumptions and simply suppose that the noise has the properties (a–e)*". The properties (a–e) of Ren are just the properties (i)–(v) of Lemma 3.2.

Proof:
According to (2.15) we have

$$y_t^2 = [(\boldsymbol{\phi} + a\boldsymbol{\theta}_{t-1})'\boldsymbol{z}_{t-1}]^2 + 2(\boldsymbol{\phi} + a\boldsymbol{\theta}_{t-1})'\boldsymbol{z}_{t-1}w_t + w_t^2, \tag{3.2}$$

hence

$$\sum_{t=1}^T y_t^2 = \sum_{t=1}^T [(\boldsymbol{\phi} + a\boldsymbol{\theta}_{t-1})'\boldsymbol{z}_{t-1}]^2 + \sum_{t=1}^T (\boldsymbol{\phi} + a\boldsymbol{\theta}_{t-1})'\boldsymbol{z}_{t-1}w_t + \sum_{t=1}^T w_t^2. \tag{3.3}$$

Since assumption (2.17) implies $\liminf \frac{1}{T}\sum_1^T \mathrm{E}[w_{t+1}^2|\mathcal{F}_t] > 0$ a.s. we obtain with Lemma 3.2 (iv) $\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^T w_t^2 > 0$ a.s. Now if the first sum on the right-hand side of (3.3) converges the second sum converges too by Lemma 3.2 (i) and (3.1) follows directly. If the first sum diverges then Lemma 3.2 (ii) gives

$$\sum_{t=1}^T (\boldsymbol{\phi} + a\boldsymbol{\theta}_{t-1})'\boldsymbol{z}_{t-1}w_t = o\left(\sum_{t=1}^T [(\boldsymbol{\phi} + a\boldsymbol{\theta}_{t-1})'\boldsymbol{z}_{t-1}]^2\right) \qquad \text{a.s.} \tag{3.4}$$

hence $\sum_{t=1}^T y_t^2 \geq \sum_{t=1}^T w_t^2$ a.s. for T sufficiently large and, again, (3.1) follows.
■

The last auxiliary result is a convergence result for almost supermartingales[2] which was already used by KOTTMANN/KULIBERDA (1990) and ZENNER (1994) to prove consistency of the OLS-algorithm and other prediction error based learning procedures in AR(1) models with forecast feedback.

Proposition 3.4: (ROBBINS/SIEGMUND (1971))
Suppose that $\{V_t\}, \{\alpha_t\}, \{\beta_t\}$, and $\{\eta_t\}$ are sequences of non–negative random variables, adapted to a filtration $\{\mathcal{F}_t\}$ such that for $t \geq 0$

$$\mathrm{E}[V_{t+1}|\mathcal{F}_t] \leq (1+\alpha_t)V_t + \beta_t - \eta_t \qquad a.s. \tag{3.5}$$

Then on the event $[\sum_{t=0}^\infty \alpha_t < \infty, \sum_{t=0}^\infty \beta_t < \infty]$ $\lim_{t\to\infty} V_t$ exists a.s. and is finite a.s. and $\sum_{t=0}^\infty \eta_t < \infty$ a.s.

[2]This convergence result is the only probabilistic result we use within our analysis. But we conjecture that it is also possible to derive an equivalent result which does not rely on any probabilistic assumptions in the sense of the deterministic reduction mentioned in the preceding footnote.
It is an interesting fact that nearly all techniques sucsessfully employed in the analysis of forecast feedback models are of a deterministic nature. The analysis of FOURGEAUD ET AL. (1986) as well as the one of KOTTMANN (1990) is completely algebraic and applies pathwise in a stochastic environment, and the approach of MARCET/SARGENT (1989a,b,c) relies on the idea that the estimation process behaves, pathwise, like a trajectory of an ordinary differential equation.

Main Results

In the following result we summarize the convergence results we shall need in the following chapters.

Theorem 3.5:
Suppose that (A.1) and (A.2) hold and $a \neq 1$. Suppose furthermore that if $q \geq 1$ the exogenous varibles $\boldsymbol{x}_t$ fulfill

$$(3.6) \qquad \limsup_{t\to\infty} \frac{\boldsymbol{x}_t'\boldsymbol{x}_t}{\sum_{s=0}^t \boldsymbol{x}_s'\boldsymbol{x}_s} < 1 \qquad a.s. \text{ and } \qquad \sum_{t=0}^{\infty} \boldsymbol{x}_t'\boldsymbol{x}_t = \infty \qquad a.s.$$

Let $\bar{\boldsymbol{\theta}} = (1-a)^{-1}\boldsymbol{\phi}$, $r_t = \sum_{s=0}^t \boldsymbol{z}_s'\boldsymbol{z}_s$, $\boldsymbol{Z}_t = \sum_{s=0}^t \boldsymbol{z}_s\boldsymbol{z}_s'$, and $\lambda_t = r_t^{-1}\boldsymbol{z}_t'\boldsymbol{z}_t$.

(i) *If $|a| \leq 1$ then*

$$(3.7) \qquad \|\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}\|^2 \qquad \text{converges a.s.,}$$

$$(3.8) \qquad \sum_{t=0}^{T} \frac{[\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2}{r_t} < \infty \qquad \text{a.s., and}$$

$$(3.9) \qquad \boldsymbol{\theta}_t \longrightarrow \bar{\boldsymbol{\theta}} \qquad \text{a.s. on} \qquad \left[\frac{\lambda_{max}(\boldsymbol{Z}_t)}{\lambda_{min}(\boldsymbol{Z}_t)} = O(1)\right],$$

where $\lambda_{max}(\boldsymbol{Z}_t)$ and $\lambda_{min}(\boldsymbol{Z}_t)$ denote the maximum and minimum eigenvalue of the matrix $\boldsymbol{Z}_t$.

(ii) *If $a < -1$ then (3.7)–(3.9) hold on the event where*

$$(3.10) \qquad \limsup_{t\to\infty} \lambda_t < \frac{2}{1-a}.$$

Proof:
The estimates $\boldsymbol{\theta}_t$ are given recursively as

$$(3.11) \qquad \begin{aligned} \boldsymbol{\theta}_{t+1} &= \boldsymbol{\theta}_t + r_t^{-1}\boldsymbol{z}_t(y_{t+1} - \boldsymbol{\theta}_t'\boldsymbol{z}_t) \\ &= \boldsymbol{\theta}_t + r_t^{-1}\boldsymbol{z}_t\boldsymbol{z}_t'(\boldsymbol{\phi} + a\boldsymbol{\theta}_t - \boldsymbol{\theta}_t) + r_t^{-1}\boldsymbol{z}_t w_{t+1} \\ &= \boldsymbol{\theta}_t + r_t^{-1}\boldsymbol{z}_t\boldsymbol{z}_t'(\boldsymbol{\phi} - (1-a)\boldsymbol{\theta}_t) + r_t^{-1}\boldsymbol{z}_t w_{t+1}, \end{aligned}$$

hence, since $\boldsymbol{\phi} = (1-a)\bar{\boldsymbol{\theta}}$,

$$(3.12) \qquad \boldsymbol{\theta}_{t+1} - \bar{\boldsymbol{\theta}} = \boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}} - (1-a)r_t^{-1}\boldsymbol{z}_t\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) + r_t^{-1}\boldsymbol{z}_t w_{t+1}.$$

Define the matrix $\boldsymbol{A}_t$ by $\boldsymbol{A}_t = \boldsymbol{I} - (1-a)r_t^{-1}\boldsymbol{z}_t\boldsymbol{z}_t'$ and set $V_t = \|\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}\|^2$. Then

$$(3.13) \qquad \mathrm{E}[V_{t+1}|\mathcal{F}_t] = \|A_t(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})\|^2 + \frac{\boldsymbol{z}_t'\boldsymbol{z}_t}{r_t^2}\mathrm{E}[w_{t+1}^2|\mathcal{F}_t]$$

with

$$
\begin{aligned}
\|\boldsymbol{A}_t(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}})\|^2 &= \|\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}}\|^2 - 2(1-a)r_t^{-1}(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}})'\boldsymbol{z}_t\boldsymbol{z}_t'(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}}) \\
&\qquad +(1-a)^2 r_t^{-2}(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}})'\boldsymbol{z}_t\boldsymbol{z}_t'\boldsymbol{z}_t\boldsymbol{z}_t'(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}}) \\
&= \|\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}}\|^2 - \left[2(1-a)-(1-a)^2\lambda_t\right]\frac{[\boldsymbol{z}_t'(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}})]^2}{r_t}.
\end{aligned}
\tag{3.14}
$$

Define furthermore

$$
\begin{aligned}
\tilde{\eta}_t &= [2(1-a)-(1-a)^2\lambda_t]\frac{[\boldsymbol{z}_t'(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}})]^2}{r_t}, \\
\eta_t &= \tilde{\eta}_t 1_{[\tilde{\eta}_t\geq 0]}, \\
\beta_t &= \sigma^2\frac{\boldsymbol{z}_t'\boldsymbol{z}_t}{r_t^2} - \tilde{\eta}_t 1_{[\tilde{\eta}_t<0]}.
\end{aligned}
\tag{3.15}
$$

Then we have

$$
\mathrm{E}[V_{t+1}|\mathcal{F}_t] \leq V_t + \beta_t - \eta_t \qquad \text{a.s.} \tag{3.16}
$$

Now we want to apply Proposition 3.4. Since $0 \leq \lambda_t \leq 1$ and

$$
2(1-a)-(1-a)^2\lambda_t \geq 0 \iff \lambda_t \leq \frac{2}{1-a}, \tag{3.17}
$$

provided that $a < 1$, it is obvious that $\tilde{\eta}_t \geq 0$ for all $t \geq 0$ if $|a| \leq 1$, and $\tilde{\eta}_t \geq 0$ for all but finitely many $t \geq 0$ on the event (3.10) if $a < -1$. Hence

$$
\sum_{t=0}^{\infty}\beta_t = O\left(\sum_{t=0}^{\infty}\frac{\boldsymbol{z}_t'\boldsymbol{z}_t}{r_t^2}\right) < \infty \qquad \text{a.s. on (3.10)} \tag{3.18}
$$

by Lemma 3.1. (To avoid the distinction between $|a| \leq 1$ and $a < -1$ we formulate in the sequel all results on the event (3.10) since this event is equal to Ω if $|a| \leq 1$.) Then Proposition 3.4 implies that $\|\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}}\|^2$ converges a.s. on the event (3.10).

To show (3.8) we establish first that

$$
2(1-a)-(1-a)^2\lambda_t \geq \epsilon \tag{3.19}
$$

holds pathwise a.s. for some sufficiently small $\epsilon = \epsilon(\omega) > 0$. Notice that, if $a < 1$ (3.19) is equivalent to $\lambda_t \leq (2-\epsilon)/(1-a)$.

If $|a| < 1$ we can choose some constant $\epsilon > 0$ such that (3.19) holds for all $t \geq 0$ since $0 \leq \lambda_t \leq 1$. If $a = -1$ then we can use the fact that $\boldsymbol{\theta}_t$ is bounded a.s. to infer that λ_t is (pathwise) bounded away from one. This follows by (3.6) and the fact that for the autoregressive process $\{y_t\}$ the quotient $y_t^2/\sum_0^t y_s^2$ can approach one only if the parameters of the autoregressive part grow beyond every boundary. But since $\boldsymbol{\theta}_t$ is bounded a.s. this is impossible with probability one. Thus we can find some $\epsilon = \epsilon(\omega) > 0$ such that (3.19) holds for all but finitely many $t \geq 0$ for almost every $\omega \in \Omega$.

If $a < -1$ we can infer in the same way that some $\epsilon = \epsilon(\omega) > 0$ exists such that (3.19) holds for all but finitely many $t \geq 0$ on the event (3.10).

Now Proposition 3.4 implies that $\sum_0^\infty \eta_t < \infty$ a.s. on the event (3.10). Since $\epsilon > 0$ this implies that

$$\sum_{t=0}^{\infty} \frac{[\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2}{r_t} < \infty \qquad \text{a.s.} \qquad \text{on (3.10)}, \tag{3.20}$$

hence (3.8) and its analogon in (ii) are shown.

Before we show (3.9) we want to remark that, again, the following considerations are all of a pathwise nature and hold on the event (3.10). Therefore we omit from now on the additional term "on (3.10)" in the formulae. With the aid of the Kronecker Lemma we can conclude that

$$\sum_{t=0}^{T} [\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2 = o(r_T) \qquad \text{a.s.} \tag{3.21}$$

By the Cauchy Schwarz inequality we obtain

$$\begin{aligned} \left\| \frac{1}{r_T} \sum_{t=0}^{T} \boldsymbol{z}_t \boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) \right\|^2 &\leq \left(\frac{1}{r_T} \sum_{t=0}^{T} \|\boldsymbol{z}_t\| |\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})| \right)^2 \\ &\leq \left(\frac{1}{r_T} \sum_{t=0}^{T} \boldsymbol{z}_t' \boldsymbol{z}_t \right) \left(\frac{1}{r_T} \sum_{t=0}^{T} [\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2 \right) \\ &= o(1). \end{aligned} \tag{3.22}$$

Now consider the process $\{\boldsymbol{Z}_T(\boldsymbol{\theta}_T - \bar{\boldsymbol{\theta}})\}$. Firstly, notice that, since $\boldsymbol{Z}_t - \boldsymbol{Z}_{t-1} = \boldsymbol{z}_t \boldsymbol{z}_t'$, we have

$$\sum_{t=0}^{T} \boldsymbol{Z}_t(\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t) = \boldsymbol{Z}_T(\boldsymbol{\theta}_{T+1} - \bar{\boldsymbol{\theta}}) - \sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) - \boldsymbol{Z}_0(\boldsymbol{\theta}_0 - \bar{\boldsymbol{\theta}}). \tag{3.23}$$

Hence by (3.22)

$$\begin{aligned} \|\boldsymbol{Z}_T(\boldsymbol{\theta}_{T+1} - \bar{\boldsymbol{\theta}})\| &\leq \left\| \sum_{t=0}^{T} \boldsymbol{Z}_t(\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t) \right\| + \left\| \sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) + \boldsymbol{Z}_0(\boldsymbol{\theta}_0 - \bar{\boldsymbol{\theta}}) \right\| \\ &= \left\| \sum_{t=0}^{T} \boldsymbol{Z}_t(\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t) \right\| + o(r_T) \end{aligned} \tag{3.24}$$

since we know that $r_T \to \infty$ a.s., either by (3.6) or by Lemma 3.3. On the other hand the recursion (3.11) leads to

$$\sum_{t=0}^{T} \boldsymbol{Z}_t(\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t) = (a-1) \sum_{t=0}^{T} \frac{1}{r_t} \boldsymbol{Z}_t \boldsymbol{z}_t \boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) + \sum_{t=0}^{T} \frac{1}{r_t} \boldsymbol{Z}_t \boldsymbol{z}_t w_{t+1}. \tag{3.25}$$

Hence

$$\left\| \sum_{t=0}^{T} \boldsymbol{Z}_t(\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t) \right\| \leq |1-a| \left\| \sum_{t=0}^{T} \tilde{\boldsymbol{Z}}_t \boldsymbol{z}_t \boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) \right\| + \left\| \sum_{t=0}^{T} \tilde{\boldsymbol{Z}}_t \boldsymbol{z}_t w_{t+1} \right\| \tag{3.26}$$

with $\tilde{\boldsymbol{Z}}_t := r_t^{-1}\boldsymbol{Z}_t$. Since $\tilde{\boldsymbol{Z}}$ is bounded because $tr(\tilde{\boldsymbol{Z}}) = 1$ we can conclude by Lemma 3.2 that

$$\sum_{t=0}^{T} \tilde{\boldsymbol{Z}}_t \boldsymbol{z}_t w_{t+1} = o(r_T) \qquad \text{a.s.} \tag{3.27}$$

Replacing $\boldsymbol{z}_t\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})$ by $\tilde{\boldsymbol{Z}}_t\boldsymbol{z}_t\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})$ in (3.22) leads to

$$\sum_{t=0}^{T} \tilde{\boldsymbol{Z}}_t \boldsymbol{z}_t \boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) = o(r_T) \qquad \text{a.s.,} \tag{3.28}$$

hence, by (3.26) and (3.27)

$$\sum_{t=0}^{T} \boldsymbol{Z}_t(\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t) = o(r_T) \qquad \text{a.s.} \tag{3.29}$$

and, finally, by (3.23)

$$\boldsymbol{Z}_T(\boldsymbol{\theta}_{T+1} - \bar{\boldsymbol{\theta}}) = o(r_T) \qquad \text{a.s.} \tag{3.30}$$

Using the inequalities

$$\|\boldsymbol{Z}_T(\boldsymbol{\theta}_{T+1} - \bar{\boldsymbol{\theta}})\| \geq \lambda_{min}(\boldsymbol{Z}_T)\|\boldsymbol{\theta}_{T+1} - \bar{\boldsymbol{\theta}}\| \tag{3.31}$$

and

$$r_T = tr(\boldsymbol{Z}_T) \leq n\lambda_{max}(\boldsymbol{Z}_T) \tag{3.32}$$

we can conclude that

$$\|\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}\| \longrightarrow 0 \qquad \text{a.s. on} \qquad \left[\limsup_{t\to\infty} \frac{\lambda_{max}(\boldsymbol{Z}_t)}{\lambda_{min}(\boldsymbol{Z}_t)} < \infty\right]. \tag{3.33}$$

Hence Theorem 3.5 is proved. ■

Theorem 3.5 states that if the *persistent excitation condition*

$$\frac{\lambda_{max}(\boldsymbol{Z}_t)}{\lambda_{min}(\boldsymbol{Z}_t)} = O(1) \qquad \text{a.s.} \tag{3.34}$$

holds the SG-algorithm estimates $\boldsymbol{\theta}_t$ are strongly consistent, at least if $|a| < 1$. This result is closely related to the respective convergence result for the SG-algorithm in models without forecast feedback (see, e.g., Ren (1991), Theorem 2.1). Thereby the PE-condition is the usual condition to ensure consistency[3].

Now we consider the case that the PE-condition fails to hold. The following considerations are, again, of algebraic nature, hence we can argue pathwise.

There are different reasons for that the PE-condition can fail to hold. Firstly, it is possible that although

$$\liminf_{t\to\infty} \frac{1}{r_T}\sum_{t=0}^{T} z_{t,i}^2 > 0 \qquad \forall i = 1, \ldots, n, \tag{3.35}$$

[3] But the PE-condition is not a necessary condition to ensure consistency. Chen/Guo (1991) show some consistency results without the PE-condition with alternative techniques. Unfortunately, we see no way to carry over these techniques to models with forecast feedback.

with $z_{t,i}$ denoting the i'th component of $\boldsymbol{z}_t$, we have $\lambda_{min}(\boldsymbol{Z}_t) = o(\lambda_{max}(\boldsymbol{Z}_t)$. Thus all diagonal elements of the 'normed' matrix of moments $\tilde{\boldsymbol{Z}}_t = r_t^{-1}\boldsymbol{Z}_t$ remain bounded away from zero uniformly in t the minimum eigenvalue of $\tilde{\boldsymbol{Z}}_t$ (which is equal to r_t^{-1} times the minimum eigenvalue of $\boldsymbol{Z}_t$) converges to zero. This is only possible if the vectors $\tilde{\boldsymbol{z}}_i(t) := r_t^{-1}(z_{t,i}, \ldots, z_{0,i})'$, $i = 1, \ldots, n$, are asymptotically linear dependent.

Secondly, it is possible that

$$(3.36) \qquad \liminf_{T\to\infty} \frac{1}{r_T} \sum_{t=0}^{T} z_{t,i}^2 = 0 \qquad \forall i \in I$$

with some $I \subset \{1, \ldots, n\}$. Thus some diagonal elements of $\tilde{\boldsymbol{Z}}_t$ converge to zero. Since $\tilde{\boldsymbol{Z}}_t$ is positive semidefinite this implies that some rows and columns completely converge to zero. (Clearly, not all rows can converge to zero since $tr(\tilde{\boldsymbol{Z}}_t) = 1$.) This is the case if some components of $\boldsymbol{z}_t$ grow faster than the others, for example, if $\boldsymbol{z}_t = (t, 1)'$. The following theorem states that in such a case at least the components of $\bar{\boldsymbol{\theta}}$ related to the fastest growing components of $\boldsymbol{z}_t$ are properly identified by the SG-algorithm.

Theorem 3.6:
Suppose that the assumptions of Theorem 3.5 are fulfilled. Let I be a subset of $\{1, \ldots, n\}$ and let $\mathcal{B}(I)$ be the event where

$$(3.37) \qquad \liminf_{T\to\infty} \frac{1}{r_T} \sum_{t=0}^{T} z_{t,i}^2 = 0 \qquad \forall i \in I.$$

Let $\boldsymbol{Z}_t^*$ be the matrix $\tilde{\boldsymbol{Z}}_t$ with the ith row and ith column deleted for all $i \in I$ and let $\mathcal{C}(I)$ be the event where

$$(3.38) \qquad \liminf_{t\to\infty} \lambda_{min}(\boldsymbol{Z}_t^*) > 0 \qquad a.s.$$

(i) *If $|a| < 1$ then*

$$(3.39) \qquad \theta_{t,i} \to \bar{\theta}_i \qquad a.s.\ on \qquad \mathcal{B}(I) \cap \mathcal{C}(I) \qquad \forall i \notin I.$$

(ii) *If $a < -1$ then for all $i \notin I$*

$$(3.40) \quad \theta_{t,i} \to \bar{\theta}_i \quad a.s.\ on \quad \mathcal{B}(I) \cap \mathcal{C}(I) \cap \left[\limsup_{t\to\infty} \lambda_t < \frac{2}{1-a}\right].$$

Proof:
We can follow the lines of the proof of Theorem 3.5 up to (3.30) without any change. We thus have $\tilde{\boldsymbol{Z}}_T(\boldsymbol{\theta}_{T+1} - \bar{\boldsymbol{\theta}}) \to 0$ a.s. on (3.10) as $T \to \infty$. On the event $\mathcal{B}(I)$ we have

$$(3.41) \qquad \|\tilde{\boldsymbol{Z}}_T(\boldsymbol{\theta}_{T+1} - \bar{\boldsymbol{\theta}})\| = \|\boldsymbol{Z}_T^*(\boldsymbol{\theta}_{T+1}^* - \bar{\boldsymbol{\theta}}^*)\| + o(1)$$

where $\boldsymbol{\theta}_t^*, \bar{\boldsymbol{\theta}}^*$ denote the vectors $\boldsymbol{\theta}_t, \bar{\boldsymbol{\theta}}$, resp., with the ith component removed for all $i \in I$. On the event $\mathcal{C}(I)$ we have

$$\|\boldsymbol{Z}_T^*(\boldsymbol{\theta}_{T+1}^* - \bar{\boldsymbol{\theta}}^*)\| \geq \lambda_{min}(\boldsymbol{Z}_T^*)\|\boldsymbol{\theta}_{T+1}^* - \bar{\boldsymbol{\theta}}^*\| \geq \epsilon\|\boldsymbol{\theta}_{T+1}^* - \bar{\boldsymbol{\theta}}^*\| \tag{3.42}$$

with some (possibly path-dependent) $\epsilon > 0$ for T sufficiently large. But this implies $\|\boldsymbol{\theta}_{T+1}^* - \bar{\boldsymbol{\theta}}^*\| \to 0$ as $T \to \infty$. ∎

Chapter 4

Condition Analysis

In this chapter we want to determine the order of $\lambda_{max}(\boldsymbol{Z}_t)$ and $\lambda_{min}(\boldsymbol{Z}_t)$ and thus obtain an upper bound for the condition number of the matrix $\boldsymbol{Z}_t$ defined as $c(\boldsymbol{Z}_t) = \|\boldsymbol{Z}_t\| \|\boldsymbol{Z}_t^{-1}\| \leq \lambda_{max}(\boldsymbol{Z}_t)/\lambda_{min}(\boldsymbol{Z}_t)$. While the order of $\lambda_{max}(\boldsymbol{Z}_t)$ is easily obtained since

$$\lambda_{max}(\boldsymbol{Z}_t) \leq tr(\boldsymbol{Z}_t) \leq n\lambda_{max}(\boldsymbol{Z}_t) \tag{4.1}$$

the determination of the (minimum) order of $\lambda_{min}(\boldsymbol{Z}_t)$ is much more complicated and requires a technical result on asymptotic properties of certain projections. Of course, only if the model (2.1) is autoregressive, i.e. if $p \geq 1$, the determination of the maximum and minimum eigenvalues causes a problem. Otherwise the order of the eigenvalues is given exogenously.

Determination of $\lambda_{max}(\boldsymbol{Z}_t)$

Let $p \geq 1$ and suppose that the Assumptions (A.1) and (A.2) hold. Using the equality $\boldsymbol{\phi} + a\bar{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}$ we can rewrite the model (2.15) as

$$y_{t+1} = \bar{\boldsymbol{\theta}}'\boldsymbol{z}_t + a(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})'\boldsymbol{z}_t + w_{t+1} \tag{4.2}$$

or, in vector notation,

$$\boldsymbol{y}_{t+1} = \boldsymbol{A}\boldsymbol{y}_t + \boldsymbol{R}_{t+1} \tag{4.3}$$

with $\boldsymbol{y}_t = (y_t, \ldots, y_{t-p+1})'$,

$$\boldsymbol{R}_{t+1} = \left(a(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})'\boldsymbol{z}_t + \sum_{j=1}^{q} \bar{\theta}_{p+j} x_{t,j} + w_{t+1}, 0, \ldots, 0\right)', \tag{4.4}$$

and $\boldsymbol{A}$ the companion form matrix

$$\boldsymbol{A} = \begin{pmatrix} \bar{\theta}_1 & \cdots & \bar{\theta}_{p-1} & \bar{\theta}_p \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{pmatrix}. \tag{4.5}$$

It is well known that the *characteristic polynomial* defined by

$$\pi(y) = y^p - \bar{\theta}_1 y^{p-1} - \cdots - \bar{\theta}_{p-1} y - \bar{\theta}_p \tag{4.6}$$

plays a central role in the asymptotic behaviour of (constant parameter) autoregressive processes. If it possesses roots only inside the unit circle then there exists some $0 \le \rho < 1$ such that $\|\boldsymbol{A}^t\| = o(\rho^t)$ and a constant parameter autoregressive process given by (4.3) with $\boldsymbol{R}_t = (w_t, 0, \ldots, 0)'$ will be stable in the sense that $\sum_1^T y_t^2 = O(T)$ a.s. The following result shows that a similar property holds also in our model. It is a straightforward extension of Lemma 4 by Lai/Wei (1982b).

Lemma 4.1:
Suppose that (A.1) and (A.2) hold and that the exogenous variables $\{\boldsymbol{x}_t\}$ fulfill

$$\|\boldsymbol{x}_t\|^2 = o(t^\beta) \qquad a.s. \tag{4.7}$$

and

$$\sum_{t=1}^{T} \|\boldsymbol{x}_t\|^2 = O(T^\gamma) \qquad a.s. \tag{4.8}$$

for some $\beta \ge 0, \gamma \ge 0$. Suppose furthermore that the characteristic polynomial (4.6) possesses only roots inside the unit circle. Then

$$\|\boldsymbol{y}_t\|^2 = o(t^\mu) + O\left(\sum_{s=0}^{t-1} [\boldsymbol{z}_s'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2\right) \qquad a.s. \tag{4.9}$$

for every μ such that $\mu \ge \beta$ and $\mu > \frac{1}{2+\delta}$, and

$$\sum_{t=1}^{T} \|\boldsymbol{y}_t\|^2 = O(T^\nu) + O\left(\sum_{s=0}^{T-1} [\boldsymbol{z}_s'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2\right) \qquad a.s. \tag{4.10}$$

with $\nu = \max\{1, \gamma\}$.

Notice that Lemma 4.1 includes the purely autoregressive case since in that case we can set $\beta = \gamma = 0$.

Proof:
Iteration of (4.3) leads to

$$\boldsymbol{y}_t = \boldsymbol{A}^t \boldsymbol{y}_0 + \sum_{s=1}^{t} \boldsymbol{A}^{t-s} \boldsymbol{R}_s. \tag{4.11}$$

With the C_r-inequality and the Cauchy Schwarz inequality we obtain

$$\begin{aligned} \|\boldsymbol{y}_t\|^2 &\le 2\|\boldsymbol{A}^t\|^2 \|\boldsymbol{y}_0\|^2 + 2\left(\sum_{s=1}^{t} \|\boldsymbol{A}^{t-s}\| \|\boldsymbol{R}_s\|\right)^2 \\ &\le 2\|\boldsymbol{A}^t\|^2 \|\boldsymbol{y}_0\|^2 + 2\sum_{s=1}^{t} \|\boldsymbol{A}^{t-s}\| \sum_{s=1}^{t} \|\boldsymbol{A}^{t-s}\| \|\boldsymbol{R}_s\|^2. \end{aligned} \tag{4.12}$$

Since $\|\boldsymbol{A}^t\| = o(\rho^t)$ with some $0 \leq \rho < 1$ (4.12) implies

$$
\begin{aligned}
\|\boldsymbol{y}_t\|^2 &= o(1) + O\left(\sum_{s=1}^{t} \rho^{t-s} \sum_{s=1}^{t} \rho^{t-s} \|\boldsymbol{R}_s\|^2\right) \\
&= O\left(\sum_{s=0}^{t-1} \rho^{t-s} \|\boldsymbol{x}_s\|^2\right) + O\left(\sum_{s=0}^{t-1} \rho^{t-s} [\boldsymbol{z}_s'(\boldsymbol{\theta}_s - \bar{\boldsymbol{\theta}})]^2\right) \\
&\quad + O\left(\sum_{s=1}^{t} \rho^{t-s} w_s^2\right) \\
&= o(t^\beta) + O\left(\sum_{s=0}^{t-1} [\boldsymbol{z}_s'(\boldsymbol{\theta}_s - \bar{\boldsymbol{\theta}})]^2\right) + o(t^\alpha) \qquad \text{a.s.}
\end{aligned}
\tag{4.13}
$$

with $\alpha > (2+\delta)^{-1}$. The last line follows by (4.7) and, since

$$
\mathrm{P}[|w_t| > t^\alpha | \mathcal{F}_{t-1}] \leq \frac{1}{t^{\alpha(2+\delta)}} \mathrm{E}[|w_t|^{2+\delta} | \mathcal{F}_{t-1}], \tag{4.14}
$$

by the conditional Borel–Cantelli lemma (Stout (1974)) which implies $w_t = o(t^\alpha)$ a.s. for all $\alpha > (2+\delta)^{-1}$. Hence (4.9) is shown.

Now summation of (4.12) leads to

$$
\begin{aligned}
\sum_{t=1}^{T} \|\boldsymbol{y}_t\|^2 &\leq 2\|\boldsymbol{y}_0\|^2 \sum_{t=1}^{T} \|\boldsymbol{A}^t\|^2 + 2\sum_{t=1}^{T} \left(\sum_{s=1}^{t} \|\boldsymbol{A}^{t-s}\| \sum_{s=1}^{t} \|\boldsymbol{A}^{t-s}\| \|\boldsymbol{R}_s\|^2\right) \\
&= O(1) + O\left(\sum_{t=1}^{T} \sum_{s=1}^{t} \rho^{t-s} \sum_{s=1}^{t} \rho^{t-s} \|\boldsymbol{R}_s\|^2\right) \\
&= O\left(\left(\sum_{t=0}^{\infty} \rho^t\right) \sum_{t=1}^{T} \sum_{s=1}^{t} \rho^{t-s} \|\boldsymbol{R}_s\|^2\right) \\
&= O\left(\left(\sum_{t=0}^{\infty} \rho^t\right)^2 \sum_{t=1}^{T} \|\boldsymbol{R}_t\|^2\right) \\
&= O\left(\sum_{t=0}^{T-1} \|\boldsymbol{x}_t\|^2\right) + O\left(\sum_{t=0}^{T-1} [\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2\right) + O\left(\sum_{t=1}^{T} w_t^2\right),
\end{aligned}
\tag{4.15}
$$

hence also (4.10) is shown. ■

Under Assumption (A.3) we obtain the following result which determines the order of $\lambda_{max}(\boldsymbol{Z}_t)$ in the stable case.

Theorem 4.2:
Suppose that (A.1), (A.2), and (A.3) hold and $a \neq 1$. If the characteristic polynomial (4.6) possesses roots only inside the unit circle then

$$
\lambda_{max}(\boldsymbol{Z}_t) = O(t) \qquad a.s. \tag{4.16}
$$

if $|a| \leq 1$, *and*

$$\lambda_{max}(\boldsymbol{Z}_t) = O(t) \qquad a.s.\ on \qquad \left[\limsup_{t\to\infty} \lambda_t < \frac{2}{1-a}\right] \tag{4.17}$$

if $a < -1$.

Proof:
Let $|a| \leq 1$. Then by (A.3) and Lemma 4.1 we know that

$$\sum_{t=1}^{T} \|\boldsymbol{y}_t\|^2 = O(T) + O\left(\sum_{t=0}^{T-1} [\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2\right) \qquad \text{a.s.} \tag{4.18}$$

Hence

$$tr(\boldsymbol{Z}_T) = O(T) + O\left(\sum_{t=0}^{T-1} [\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2\right) \qquad \text{a.s.} \tag{4.19}$$

By Theorem 3.5

$$\sum_{t=0}^{T} [\boldsymbol{z}_t'(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})]^2 = o(tr(\boldsymbol{Z}_T)) \qquad \text{a.s.} \tag{4.20}$$

hence $tr(\boldsymbol{Z}_T) = O(T)$ a.s. Using the inequality (4.1) we finally obtain (4.16). The proof of (4.17) is completely analogous. ■

Determination of $\lambda_{min}(\boldsymbol{Z}_t)$

As already mentioned the determination of the minimum order of $\lambda_{min}(\boldsymbol{Z}_t)$ is much more complicated. LAI/WEI (1982b) have the following result.

Lemma 4.3: (LAI/WEI (1982B)
Let $\boldsymbol{Z}(t) = (\boldsymbol{z}_1(t), \ldots, \boldsymbol{z}_n(t))$ *be a matrix of dimension* $t \times n$, *then*

$$\begin{aligned} \frac{1}{n} \min_{1\leq j\leq n} \|\boldsymbol{z}_j(t) - \hat{\boldsymbol{z}}_j(t)\|^2 \quad &\leq \quad \lambda_{min}(\boldsymbol{Z}(t)'\boldsymbol{Z}(t)) \\ &\leq \quad n \min_{1\leq j\leq n} \|\boldsymbol{z}_j(t) - \hat{\boldsymbol{z}}_j(t)\|^2 \end{aligned} \tag{4.21}$$

where $\hat{\boldsymbol{z}}_j(t)$ *denotes the (orthogonal) projection of* $\boldsymbol{z}_j(t)$ *onto the linear space spanned by* $\boldsymbol{z}_1(t), \ldots, \boldsymbol{z}_{j-1}(t), \boldsymbol{z}_{j+1}(t), \ldots, \boldsymbol{z}_n(t)$.

It is worth noting that all projections appearing in this paper apply pointwise and should not be confused with the L_2–projections usually applied in the theory of stochastic processes. The following result is the keystone for the determination of $\lambda_{min}(\boldsymbol{Z}_t)$.

Proposition 4.4: (Lai/Wei (1982b))
Suppose that (A.2) holds. Let u_t, v_t and $z_{t,1}, \ldots, z_{t,n}$ be $\mathcal{F}_{t-1}$-measurable random variables. Let $\boldsymbol{Z}(t) = (z_{ij})_{1 \le i \le t, 1 \le j \le n}$, $\boldsymbol{w}_t = (w_1, \ldots, w_t)'$, $\boldsymbol{u}_t = (u_1, \ldots, u_t)'$ and $\boldsymbol{v}_t = (v_1, \ldots, v_t)'$. Let $\hat{\boldsymbol{w}}_t, \hat{\boldsymbol{u}}_t, \hat{\boldsymbol{v}}_t$ denote the projections of $\boldsymbol{w}_t, \boldsymbol{u}_t, \boldsymbol{v}_t$ onto $L(\boldsymbol{Z}(t))$, the linear space spanned by the column vectors of $\boldsymbol{Z}(t)$. Let $\hat{\hat{\boldsymbol{u}}}_t$ denote the projection of $\boldsymbol{u}_t$ onto $L(\boldsymbol{Z}(t), \boldsymbol{v}_t + \boldsymbol{w}_t)$ and let $\boldsymbol{v}_t^*$ denote the projection of $\boldsymbol{v}_t$ onto $L(\boldsymbol{Z}(t), \boldsymbol{w}_t)$. Suppose that

$$\max\left\{0, \log\left(\sum_{j=1}^{n}\sum_{s=1}^{t} z_{js}^2 + \sum_{s=1}^{t} u_s^2\right)\right\} = o(t) \qquad a.s. \tag{4.22}$$

Then

$$\|\boldsymbol{w}_t\|^2 = O(t) \qquad a.s. \text{ and } \qquad \liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{w}_t - \hat{\boldsymbol{w}}_t\|^2 > 0 \quad a.s. \tag{4.23}$$

and

$$\|\boldsymbol{u}_t - \hat{\hat{\boldsymbol{u}}}_t\|^2 = \left(\frac{\|\boldsymbol{v}_t - \boldsymbol{v}_t^*\|^2 + \|\boldsymbol{w}_t - \hat{\boldsymbol{w}}_t\|^2}{\|\boldsymbol{v}_t - \hat{\boldsymbol{v}}_t\|^2 + \|\boldsymbol{w}_t - \hat{\boldsymbol{w}}_t\|^2} + o(1)\right)\|\boldsymbol{u}_t - \hat{\boldsymbol{u}}_t\|^2 \quad a.s. \tag{4.24}$$

With the aid of Proposition 4.4 we can prove the following result which determines the minimum order of the minimum eigenvalue in the stable case.

Theorem 4.5:
Suppose that (A.1),(A.2) and (A.3) hold and $a \neq 1$. If $|a| \le 1$ and $tr(\boldsymbol{Z}_t) = O(t)$ a.s. then

$$\liminf_{t\to\infty} \frac{1}{t}\lambda_{min}(\boldsymbol{Z}_t) > 0 \qquad a.s. \tag{4.25}$$

If $a < -1$ then (4.24) holds on the event where

$$\left[tr(\boldsymbol{Z}_t) = O(t), \limsup_{t\to\infty} \lambda_t < \frac{2}{1-a}\right]. \tag{4.26}$$

Proof:
We imitate the proof of Corollary 2 in Lai/Wei (1982b). In order to maintain comparability we try to use their notation wherever possible. Define

$$\begin{array}{rcll} \boldsymbol{y}_t(\nu) & = & (y_{p+1-\nu}, \ldots, y_{t-\nu})' & \nu = 0, \ldots, 2p, \\ \boldsymbol{w}_t(\nu) & = & (w_{p+1-\nu}, \ldots, w_{t-\nu})' & \nu = 0, \ldots, p, \\ \boldsymbol{x}_{t,i}(\nu) & = & (x_{p+1-\nu,i}, \ldots, x_{t-\nu,i})' & \nu = 1, \ldots, p+1, i = 1, \ldots, n \\ \boldsymbol{\eta}_t(\nu) & = & (\eta_{p+1-\nu}, \ldots, \eta_{t-\nu})' & \nu = 1, \ldots, p+1 \end{array} \tag{4.27}$$

with $\eta_{t-\nu} = a(\boldsymbol{\theta}_{t-\nu} - \bar{\boldsymbol{\theta}})'\boldsymbol{z}_{t-\nu}$. Then (4.2) implies that for all $t \geq p+1$ and every $\nu = 0, \ldots, p$

$$\begin{aligned}(4.28)\ \boldsymbol{y}_t(\nu) &= \bar{\theta}_1 \boldsymbol{y}_t(\nu+1) + \cdots + \bar{\theta}_p \boldsymbol{y}_t(\nu+p) + \bar{\theta}_{p+1}\boldsymbol{x}_{t,1}(\nu+1) + \cdots \\ &\quad \cdots + \bar{\theta}_n \boldsymbol{x}_{t,q}(\nu+1) + \boldsymbol{\eta}_t(\nu+1) + \boldsymbol{w}_t(\nu).\end{aligned}$$

Notice that $\boldsymbol{y}_t(\nu), \boldsymbol{x}_{t,i}(\nu), \boldsymbol{\eta}_t(\nu)$ and $\boldsymbol{w}_t(\nu)$ are all $\mathcal{F}_{t-\nu}$-measurable. Furthermore we define the matrices $\boldsymbol{X}(t)$ and $\boldsymbol{Z}(t)$ by

$$\boldsymbol{X}(t) = (\boldsymbol{x}_{t,1}(1), \ldots, \boldsymbol{x}_{t,q}(1)) \tag{4.29}$$

and

$$\boldsymbol{Z}(t) = (\boldsymbol{z}_1(t), \ldots, \boldsymbol{z}_n(t)) := (\boldsymbol{y}_t(1), \ldots, \boldsymbol{y}_t(p), \boldsymbol{X}(t)). \tag{4.30}$$

Now let $|a| \leq 1$. By construction we have $\boldsymbol{Z}(t)'\boldsymbol{Z}(t) = \boldsymbol{Z}_{t-1} - \boldsymbol{Z}_{p-1}$. Since $tr(\boldsymbol{Z}_t) = O(t)$ a.s. by assumption and by Theorem 3.5 we can conclude that

$$tr\,(\boldsymbol{Z}(t)'\boldsymbol{Z}(t)) = O(t) \qquad \text{a.s. and} \qquad \|\boldsymbol{\eta}_t(\nu)\|^2 = o(t) \qquad \text{a.s.} \tag{4.31}$$

We want to show that $\liminf_{t\to\infty} \frac{1}{t}\lambda_{min}(\boldsymbol{Z}(t)'\boldsymbol{Z}(t)) > 0$ a.s. In view of (4.21) this is equivalent to show that

$$\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{z}_j(t) - \hat{\boldsymbol{z}}_j(t)\|^2 > 0 \qquad \text{a.s.} \tag{4.32}$$

for all $j = 1, \ldots, n$ with $\hat{\boldsymbol{z}}_j(t)$ the projection of $\boldsymbol{z}_j(t)$ onto $L_j^*(\boldsymbol{Z}(t))$, the linear space spanned by $\boldsymbol{z}_1(t), \ldots, \boldsymbol{z}_{j-1}(t), \boldsymbol{z}_{j+1}(t), \ldots, \boldsymbol{z}_n(t)$.

In the sequel we will repeatedly apply Proposition 4.4. Notice that the assumption $tr(\boldsymbol{Z}_t) = O(t)$ implies that condition (4.22) is fulfilled whenever we apply Proposition 4.4 to the time series of our model.

Step 1: If $q \geq 1$, thus if equation (2.1) includes exogenous variables, consider a column vector $\boldsymbol{u}_t$ of $\boldsymbol{X}(t)$ and denote by $\boldsymbol{X}^*(t)$ the matrix $\boldsymbol{X}(t)$ with the vector $\boldsymbol{u}_t$ removed and by $\hat{\boldsymbol{u}}_t$ the projection of $\boldsymbol{u}_t$ onto $L(\boldsymbol{X}^*(t))$. (If $q = 0$ we can go directly to Step 2.) By Assumption (A.3) and (4.21) we know that

$$\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{u}_t - \hat{\boldsymbol{u}}_t\|^2 > 0 \qquad \text{a.s.} \tag{4.33}$$

Now let $\hat{\boldsymbol{u}}_{t,p}$ be the projection of $\boldsymbol{u}_t$ onto $L(\boldsymbol{X}^*(t), \boldsymbol{y}_t(p))$. By (4.28) with $\nu = p$ we can infer that

$$\boldsymbol{y}_t(p) = \boldsymbol{v}_t(p) + \boldsymbol{w}_t(p) \tag{4.34}$$

with some $\boldsymbol{v}_t(p) = (v_{p+1}, \ldots, v_t)'$ such that v_i is $\mathcal{F}_{i-p-1}$-measurable for all $i = p+1, \ldots, t$. By Proposition 4.4 (with $\hat{\boldsymbol{u}}_{t,p}$ playing the role of $\hat{\hat{\boldsymbol{u}}}_t$ in (4.24))

$$\|\boldsymbol{u}_t - \hat{\boldsymbol{u}}_{t,p}\|^2 \geq \left(\frac{\|\boldsymbol{w}_t(p) - \hat{\boldsymbol{w}}_t(p)\|^2}{\|\boldsymbol{v}_t(p) - \hat{\boldsymbol{v}}_t(p)\|^2 + \|\boldsymbol{w}_t(p) - \hat{\boldsymbol{w}}_t(p)\|^2} + o(1)\right)\|\boldsymbol{u}_t - \hat{\boldsymbol{u}}_t\|^2 \quad \text{a.s.} \tag{4.35}$$

with $\hat{\boldsymbol{v}}_t(p), \hat{\boldsymbol{w}}_t(p)$ the respective projections of $\boldsymbol{v}_t(p), \boldsymbol{w}_t(p)$ onto $L(\boldsymbol{X}^*(t))$. Since $\|\boldsymbol{v}_t(p)\|^2 = O(t)$ a.s. by assumption and $\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{w}_t(p) - \hat{\boldsymbol{w}}_t(t)\|^2 > 0$ a.s. by Proposition 4.4 it follows that

$$\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{u}_t - \hat{\boldsymbol{u}}_{t,p}\|^2 > 0 \qquad \text{a.s.} \tag{4.36}$$

Now let $\hat{\boldsymbol{u}}_{t,p-1}$ be the projection of $\boldsymbol{u}_t$ onto $L(\boldsymbol{X}^*(t), \boldsymbol{y}_t(p), \boldsymbol{y}_t(p-1))$. By a similar argument as above we then obtain that $\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{u}_t - \hat{\boldsymbol{u}}_{t,p-1}\|^2 > 0$ a.s. Proceding inductively in this way we finally obtain

$$\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{u}_t - \hat{\boldsymbol{u}}_{t,1}\|^2 > 0 \qquad \text{a.s.} \tag{4.37}$$

with $\hat{\boldsymbol{u}}_{t,1}$ the projection of $\boldsymbol{u}_t$ onto $L(\boldsymbol{X}^*(t), \boldsymbol{y}_t(p), \ldots, \boldsymbol{y}_t(1))$. Hence we have shown (4.32) for all vectors $\boldsymbol{z}_j(t)$ with $j \geq p+1$.

Step 2: Now (or if equation (2.1) includes no exogenous variables) consider a column vector $\boldsymbol{y}_t(\nu)$, $\nu = 1, \ldots, p$, of $\boldsymbol{Z}(t)$. Define the matrices $\boldsymbol{X}_t(\nu)$ and $\boldsymbol{Z}_t(\nu)$ by

$$\begin{aligned} \boldsymbol{X}_t(\nu) &= (\boldsymbol{x}_{t,1}(\nu), \ldots, \boldsymbol{x}_{t,q}(\nu)), \\ \boldsymbol{Z}_t(\nu) &= (\boldsymbol{y}_t(\nu+1), \ldots, \boldsymbol{y}_t(\nu+p), \boldsymbol{X}_t(\nu+1), \ldots, \boldsymbol{X}_t(1)). \end{aligned} \tag{4.38}$$

Notice that by Assumption (A.2) the matrix $\boldsymbol{Z}_t(\nu)$ is $\mathcal{F}_{t-\nu-1}$-measurable. Then define the matrix $\boldsymbol{Z}^*(t)$ as the matrix $\boldsymbol{Z}(t)$ with the column vector $\boldsymbol{y}_t(\nu)$ removed. Let $\boldsymbol{y}_t^*(\nu)$ be the projection of $\boldsymbol{y}_t(\nu)$ onto $L(\boldsymbol{Z}^*(t))$ and $\hat{\boldsymbol{y}}_t(\nu)$ the projection of $\boldsymbol{y}_t(\nu)$ onto $L(\boldsymbol{y}_t(1), \ldots, \boldsymbol{y}_t(\nu-1), \boldsymbol{Z}_t(\nu))$. We want to show that $\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{y}_t(\nu) - \boldsymbol{y}_t^*(\nu)\|^2 > 0$ a.s. Since $L(\boldsymbol{Z}^*(t)) \subset L(\boldsymbol{y}_t(1), \ldots, \boldsymbol{y}_t(\nu-1), \boldsymbol{Z}_t(\nu))$ implies that

$$\|\boldsymbol{y}_t(\nu) - \boldsymbol{y}_t^*(\nu)\|^2 \geq \|\boldsymbol{y}_t(\nu) - \hat{\boldsymbol{y}}_t(\nu)\|^2 \tag{4.39}$$

it is sufficient to show that $\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{y}_t(\nu) - \hat{\boldsymbol{y}}_t(\nu)\|^2 > 0$ a.s.

By construction $\boldsymbol{y}_t(\nu)$ is the sum of a linear combination of some column vectors of $\boldsymbol{Z}_t(\nu)$ and the vector $\boldsymbol{w}_t(\nu)$. Let $\pi_{0,t}$ be the projection of $\boldsymbol{y}_t(\nu)$ onto $L(\boldsymbol{Z}_t(\nu))$ then, since $\boldsymbol{y}_t(\nu) - \pi_{0,t} = \boldsymbol{w}_t - \hat{\boldsymbol{w}}_t$, Proposition 4.4 gives

$$\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{y}_t(\nu) - \pi_{0,t}\|^2 > 0 \qquad \text{a.s.} \tag{4.40}$$

Now let $\pi_{1,t}$ be the projection of $\boldsymbol{y}_t(\nu)$ onto $L(\boldsymbol{Z}_t(\nu), \boldsymbol{y}_t(\nu-1))$. Since

$$L(\boldsymbol{Z}_t(\nu), \boldsymbol{y}_t(\nu-1)) = L(\boldsymbol{Z}_t(\nu), \bar{\boldsymbol{\theta}}_1\boldsymbol{y}_t(\nu) + \boldsymbol{\eta}_t(\nu) + \boldsymbol{w}_t(\nu-1)) \tag{4.41}$$

and letting $\boldsymbol{v}_t = \bar{\boldsymbol{\theta}}_1\boldsymbol{y}_t(\nu) + \boldsymbol{\eta}_t(\nu)$ Proposition 4.4 implies that with probability one

$$\begin{aligned} \|\boldsymbol{y}_t(\nu) - \pi_{1,t}\|^2 &= \left(\frac{\|\boldsymbol{v}_t - \boldsymbol{v}_t^*\|^2 + \|\boldsymbol{w}_t(\nu-1) - \hat{\boldsymbol{w}}_t(\nu-1)\|^2}{\|\boldsymbol{v}_t - \hat{\boldsymbol{v}}_t\|^2 + \|\boldsymbol{w}_t(\nu-1) - \hat{\boldsymbol{w}}_t(\nu-1)\|^2} + o(1)\right) \times \\ &\quad \times \|\boldsymbol{y}_t(\nu) - \pi_{0,t}\|^2 \end{aligned} \tag{4.42}$$

with $\hat{\boldsymbol{v}}_t$ the projection of $\boldsymbol{v}_t$ onto $L(\boldsymbol{Z}_t(\nu))$ and $\boldsymbol{v}_t^*$ the projection of $\boldsymbol{v}_t$ onto $L(\boldsymbol{Z}_t(\nu), \boldsymbol{w}_t(\nu-1))$. By assumption we have $\|\boldsymbol{v}_t\|^2 = O(t)$ a.s., hence by Proposition 4.4

$$\liminf_{t\to\infty} \frac{1}{t}\|\boldsymbol{y}_t(\nu) - \pi_{1,t}\|^2 > 0 \qquad \text{a.s.} \tag{4.43}$$

Proceding inductively in this way we finally obtain

$$\liminf_{t\to\infty} \frac{1}{t} \|\boldsymbol{y}_t(\nu) - \pi_{\nu-1,t}\|^2 > 0 \qquad \text{a.s.} \tag{4.44}$$

with $\pi_{\nu-1,t}$ the projection of $\boldsymbol{y}_t(\nu)$ onto $L(\boldsymbol{Z}_t(\nu), \boldsymbol{y}_t(\nu-1), \ldots, \boldsymbol{y}_t(1))$, hence $\pi_{\nu-1,t} = \boldsymbol{y}_t^*(\nu)$ by construction and Theorem 4.5 is shown for $|a| \leq 1$.

The proof for the case $a < -1$ is completely analogous except that all considerations hold on the event (4.26). ∎

Chapter 5

Convergence Results

Now we can reap the rewards of our work in the preceding chapters and show some proper convergence results for the SG-algorithm.

Non-Autoregressive Models

For non-autoregressive models, thus if $p = 0$ holds in our model (2.1), the situation is quite simple since all assumptions ensuring convergence of the estimates $\{\boldsymbol{\theta}_t\}$ can be formulated as exogenous.

Theorem 5.1:
Suppose that for the model (2.1) with $p = 0$ the assumptions (A.1) and (A.2) hold. Suppose furthermore that $a \neq 1$ and

$$\frac{\lambda_{max}(\boldsymbol{X}_t)}{\lambda_{min}(\boldsymbol{X}_t)} = O(1) \qquad a.s. \text{ and } \qquad tr(\boldsymbol{X}_t) \to \infty \qquad a.s. \tag{5.1}$$

with $\boldsymbol{X}_t = \sum_{t=0}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'$.

(i) *If $|a| \leq 1$ then $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s.*

(ii) *If $a < -1$ then $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s. on the event*

$$\limsup_{t\to\infty} \frac{\boldsymbol{x}_t' \boldsymbol{x}_t}{tr(\boldsymbol{X}_t)} < \frac{2}{1-a} \qquad a.s. \tag{5.2}$$

Theorem 5.1 is an immediate consequence of Theorem 3.5, hence we can omit a proof. As a corollary we get a result which is closely related to the results of BRAY/SAVIN (1986), FOURGEAUD ET AL. (1986) and KOTTMANN (1990) for the OLS-algorithm.

Corollary 5.2:
Suppose that for the model (2.1) with $p = 0$ the assumptions (A.1) and (A.2) hold. Suppose furthermore that

$$\frac{1}{T}\sum_{t=0}^{T} \boldsymbol{x}_t \boldsymbol{x}_t' \longrightarrow \boldsymbol{X} \qquad a.s. \tag{5.3}$$

with some a.s. positive definite (possibly random) matrix $\boldsymbol{X}$. Then $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s. if $a < 1$.

Proof:
We only have to show that condition (5.2) holds for arbitrary $a < 1$. This is equivalent to showing that $\lambda_t = \boldsymbol{x}_t'\boldsymbol{x}_t / \sum_0^t \boldsymbol{x}_s'\boldsymbol{x}_s \to 0$ a.s. But since

$$1 - \lambda_t = \frac{tr(\sum_{s=0}^{t-1} \boldsymbol{x}_s\boldsymbol{x}_s')}{tr(\sum_{s=0}^{t} \boldsymbol{x}_s\boldsymbol{x}_s')} = \frac{t-1}{t}\frac{\frac{1}{t-1}tr(\boldsymbol{X}_{t-1})}{\frac{1}{t}tr(\boldsymbol{X}_t)} \tag{5.4}$$

$\lambda_t \to 0$ a.s. is immediate from (5.3). ∎

The condition (5.3) is fulfilled whenever $\{\boldsymbol{x}_t\}$ is a covariance stationary and ergodic stochastic process with a non-singular covariance matrix, thus for a large class of stochastic processes. This is the assumption usually employed for the exogenous variables in models with forecast feedback. But Theorem 5.1 covers a more general situation since it assumes only that the persistent excitation condition holds and that the trace of the matrix of moments diverges a.s. to infinity.

Thus our result is more general than the results obtained by all other authors analyzing models with forecast feedback. The reason for this is that our approach does not rely on a law of large numbers while the martingale approach of Bray/Savin (1986), the ordinary differential equation approach of Marcet/Sargent (1989a,b,c), and the stochastic approximation approach of Kottmann (1990) do. Only the algebraic approach of Fourgeaud et al. (1986) relies on assumptions similar to the assumptions of Theorem 5.1 (but it results in a consistency result only for $a < 1/2$ whenever $\boldsymbol{x}_t$ is multivariate). Notice, however, that all these studies are concerned with the OLS-algorithm and not with the SG-algorithm (which coincides with the OLS-algorithm only if $\boldsymbol{x}_t$ is univariate) hence, basically, we cannot compare the results with our results, although the OLS-algorithm and the SG-algorithm have a similar structure.

To conclude this section we want to clarify the meaning of condition (5.2). As shown in the proof of Corollary 5.2 the quotient λ_t converges towards zero for all covariance stationary and ergodic processes, hence (5.2) does not imply any restriction. In general, this does not automatically hold for processes fulfilling the PE-condition. For example, consider the following univariate unstable AR(1) process given by

$$x_{t+1} = \eta x_t + \epsilon_{t+1} \tag{5.5}$$

with $|\eta| > 1$ and $\{\epsilon_t\}$ a martingale difference sequence fulfilling an assumption analogous to (A.2). It is well known (see, e.g., WEI (1987)) that in this case $\lambda_t \rightarrow (\eta^2 - 1)/\eta^2$ a.s. Then condition (5.2) is satisfied only if

$$a > -\frac{\eta^2 + 1}{\eta^2 - 1} \tag{5.6}$$

which gives a lower bound for the parameter a. In fact, this lower bound is sharp and (5.2) turns out to be a necessary condition for convergence as some computer simulations have shown (see also ZENNER (1992)).

Autoregressive Models – The Stable Case

For autoregressive models, thus for the model (2.1) with $p \geq 1$, the situation is more complicated. Since the estimates $\boldsymbol{\theta}_t$ influence the behaviour of the endogenous variable y_t the PE-condition is an endogenous property which we have to prove firstly before applying the convergence results of Chapter 3. We obtain the following result.

Theorem 5.3:
Suppose that for the model (2.1) the Assumptions (A.1)–(A.3) hold and $|a| < 1$. If the characteristic polynomial (4.6) possesses roots only inside the unit circle then $\boldsymbol{\theta}_t \rightarrow \bar{\boldsymbol{\theta}}$ a.s.

Proof:
By Theorem 4.2 we have $\lambda_{max}(\boldsymbol{Z}_t) = O(t)$ a.s. and by Theorem 4.4 we obtain $\liminf_{t\rightarrow\infty} t^{-1}\lambda_{min}(\boldsymbol{Z}_t) > 0$ a.s. Hence the PE-condition holds a.s. and by Theorem 3.5 we obtain $\boldsymbol{\theta}_t \rightarrow \bar{\boldsymbol{\theta}}$ a.s. ■

Theorem 5.3 is a proper result which has no counterpart for the OLS-algorithm in the literature since none of the employed approaches is able to handle the simple model (2.1). The reason for this lack of ability can be seen as follows. We can understand the process $\{y_t\}$ as an ARX process with time varying parameters given by

$$y_{t+1} = f(\boldsymbol{\theta}_t)'\boldsymbol{z}_t + w_{t+1} \tag{5.7}$$

with the *feedback function* $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ which maps the em perceived law of motion, $\boldsymbol{\theta}_t$, into the *actual law of motion*, $f(\boldsymbol{\theta}_t)$, of the process $\{y_t\}$. In our model the feedback function is given as $f(\boldsymbol{\theta}) = \boldsymbol{\phi} + a\boldsymbol{\theta}$.

This simple linear feedback function looks harmless since it is Lipschitz continuous and contracting if $|a| < 1$. Nevertheless, a small change in the argument of f can imply substantial changes in the qualitative long-term behaviour of the process $\{y_t\}$ since the qualitative long-term behaviour of a (constant parameter) autoregressive process differs drastically depending on whether the characteristic polynomial possesses roots only inside the unit circle or not. Hence if the range of the feedback function f is not restricted to values that imply a 'stable' characteristic polynomial the qualitative long-term behaviour of $\{y_t\}$ is by no means determined apriori but depends on the evolution of the estimates $\boldsymbol{\theta}_t$. Vice versa, the evolution of the estimates depend, sometimes also drastically via the quotient λ_t, on the qualitative long-term behaviour of $\{y_t\}$. This kind of feedback is the reason why the analysis of autoregressive models with forecast feedback is so difficult.

The ODE approach of MARCET/SARGENT (1989a,b,c) and CHANG ET AL. (1991a,b,c) tries to avoid this problem by requiring that the range of the feedback function is restricted to some kind of 'stable region'. Clearly, the feedback function in our model, $f(\boldsymbol{\theta}) = \boldsymbol{\phi} + a\boldsymbol{\theta}$, does not fulfill this assumption. Hence their approach fails to solve our problem. For this reason the scope of the ordinary differential equation (ODE) approach of LJUNG (1977) on which these studies rely is limited since, as shown in Chapter 2, even elementary economic applications do not satisfy this stability assumption. Moreover, we believe that the question of whether learning agents can destabilize an economic system or not is the most interesting one. We have shown in this paper that whenever the REE is stable (i.e., whenever the characteristic polynomial possesses roots only inside the unit circle) and the influence of agents' predictions is limited (i.e. if $|a| \leq 1$) then agents learning by the SG-algorithm cannot destabilize the system and do, in fact, learn to form rational expectations in the sense that their parameter estimates converge with probability one towards the rational expectations parameter.

As already mentioned this result is completely new and has, to our knowledge, no counterpart in the literature although KOTTMANN (1990) has shown a convergence result for the SG-algorithm in autoregressive models. But he embeds the consistency problem in a problem of adaptive control of autoregressive processes in a way that the agents' predictions are used (in a very special way) as control inputs in order to track the process $\{y_t\}$ close to some reference path (see also BAŞAR (1989) for this approach). We believe that this is not the right way to look at the problem whether agents can learn to form rational expectations.

Although Theorem 5.3 is a satisfactory convergence result it is natural to ask whether it is the most we can expect. For example one could ask whether it is possible to weaken the assumption $|a| \leq 1$ such that $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s. still holds. We claim that this is not possible. As already noticed by BRAY/SAVIN (1986) in their simulation study of the model (2.1) jointly with the OLS-algorithm a parameter $a > 1$ seems to lead with probability one to exploding estimates. We observed the same behaviour for the SG-estimates in our simulations. Unfortunately, we cannot prove this fact analytically, neither can BRAY/SAVIN, although they give an intuitive explanation for their belief which can be carried over to the SG-algorithm.

If $a < -1$ then, in order to apply Theorem 3.5, we have to ensure that $\liminf_{t\to\infty} \lambda_t < 2/(1-a)$ a.s. to obtain $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s. But, as our computer simulations have shown, this is not possible since the asymptotic behaviour of the quotient λ_t depends on the asymptotic behaviour of the estimates $\boldsymbol{\theta}_t$ in a way that, for example, a large initial value $\boldsymbol{\theta}_0$ can lead with positive probability to realizations of the process $\{y_t\}$ such that $\lambda_t > 2/(1-a)$ holds for almost all $t \geq 0$, especially if $|a|$ is large. In such a case the estimates $\boldsymbol{\theta}_t$ explode. This kind of feedback behaviour of autoregressive models was also observed by CYERT/DEGOOT (1974) in a Monte–Carlo study. Again, we cannot prove that $P(\|\boldsymbol{\theta}_t\| \to \infty) > 0$ if $a < -1$ although we believe that it is true.

Autoregressive Models – The Unstable Case

Now we consider the case that the characteristic polynomial possesses also roots on or outside the unit circle. Since in that case the convergence results which can be achieved are less satisfactory and additional technical difficulties arise we omit formal proofs. We only outline briefly and sometimes heuristically how proofs can be given.

Let $\mu_1, \ldots, \mu_k$ denote the roots of the characteristic polynomial (4.6) and $m_1, \ldots, m_k$ the respective multiplicities. Define

$$\bar{\mu} = \max_{1 \leq i \leq k} |\mu_i| \qquad \text{and} \qquad M = \max_{|\mu_j| = \bar{\mu}} \{m_j\}. \tag{5.8}$$

Then it is well known (cf. LAI/WEI (1985)) that for the companion form matrix A defined by (4.5) there exists a positive constant c such that

$$\|A^t\| \approx c t^{M-1} \bar{\mu}^t \qquad \text{as} \qquad t \to \infty. \tag{5.9}$$

Now suppose that (A.1) and (A.2) hold and $|a| \leq 1$. Then it is not difficult to show under the assumptions (4.7) and (4.8) of Lemma 4.1

$$tr(\boldsymbol{Z}_t) = O(t^\nu) \qquad \text{a.s.} \tag{5.10}$$

for some $\nu \geq 2$ if $\bar{\mu} = 1$, and

$$tr(\boldsymbol{Z}_t) = O\left(t^{2(M-1)} \bar{\mu}^{2t}\right) \qquad \text{a.s.} \tag{5.11}$$

if $\bar{\mu} > 1$. In addition, if $\bar{\mu} > 1$ it can be shown that the process $\{y_t\}$ explodes geometrically in modulus a.s. If $\bar{\mu} = 1$ the results of some computer simulations suggest that $\frac{1}{t} tr(\boldsymbol{Z}_t) \to \infty$ a.s. Hence in both cases we obtain $\frac{1}{t}\lambda_{max}(\boldsymbol{Z}_t) \to \infty$ a.s.

Now suppose that the model (2.1) with $\bar{\mu} \geq 1$ includes some exogenous variables for which (A.3) holds. Since

$$\lambda_{min}(\boldsymbol{Z}_t) \leq \lambda_{min}(\boldsymbol{X}_t) \leq \lambda_{max}(\boldsymbol{X}_t) \leq \lambda_{max}(\boldsymbol{Z}_t) \tag{5.12}$$

we face the situation that $\lambda_{min}(\boldsymbol{Z}_t)$ and $\lambda_{max}(\boldsymbol{Z}_t)$ are of different order and the PE-condition does not hold. Therefore we cannot apply Theorem 3.5 to show consistency. Instead we want to apply Theorem 3.6.

Suppose for the moment that $p = 1$ and $a \neq 1$. Then the characteristic polynomial turns out to be

$$\pi(y) = y - \bar{\theta}_1 = \frac{1}{1-a}\phi_1 \tag{5.13}$$

and possesses only the root $y^* = (1-a)^{-1}\phi_1$. Then Theorem 3.6 implies that $\theta_{t,1} \to \bar{\theta}_1$ a.s. whenever $|a| \leq 1$ and $\phi_1 \geq 1-a$.

Thus the first component of the rational expectations parameter $\bar{\boldsymbol{\theta}}$ will be identified properly by the SG-algorithm while for the other components we only know that they remain bounded a.s. Hence the limit expectations of agents can be non-rational but, since $z_{t,1}$ dominates the whole vector $\boldsymbol{z}_t$ and $\theta_{t,1} \to \bar{\theta}_1$ a.s., their predictions $\boldsymbol{\theta}_t{}'\boldsymbol{z}_t$ are close to the 'rational predictions' $\bar{\boldsymbol{\theta}}'\boldsymbol{z}_t$ in relative terms, i.e. $\|\boldsymbol{z}_t\|^{-2}[(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}})'\boldsymbol{z}_t]^2$ is small.

Now suppose that $p > 1$. Then the situation is even more complicated since we have to verify the condition (3.38) of Theorem 3.6, which requires that $\liminf_{t\to\infty} \lambda_{min}(\boldsymbol{Z}_t^*) > 0$ a.s., in order to get a convergence result. Generally, we have $\boldsymbol{Z}_t^* = r_t^{-1}\sum_0^t \boldsymbol{y}_s\boldsymbol{y}_s'$, since (3.35) holds for all $i = 1,\ldots,p$ and (3.36) for $i = p+1,\ldots,n$.

For constant parameter autoregressive processes it is well-known (see, e.g., Lai/Wei (1983)) that (3.38) holds only if all the roots of the characteristic poynomial lie outside the unit circle and have the same absolute value. Otherwise we have $\lim_{t\to\infty} \lambda_{min}(\boldsymbol{Z}_t^*) = 0$ a.s. Therefore only in that very special case the SG-algorithm will generate consistent estimates for the p first components of $\bar{\boldsymbol{\theta}}$. We believe that the same holds in our model, at least if $|a| \leq 1$. Hence, even consistency of only some components of the SG-algorithm is very unlikely to occur in the unstable case with $p > 1$ (more exactly, the set of parameter values $a, \phi_1, \ldots, \phi_p$ which induce consistency of the first p components has Lebesgue measure zero).

Examples

As an application and illustration we want to give two examples. Firstly, we continue with the example presented in Chapter 2, then we extend the model of Muth (1961) in order to obtain a non-trivial autoregressive model.

Example 5.4: (Example 2.1, continued)
The reduced form equation of this model is given as (cf. (2.3))

$$p_{t+1} = \frac{d_1 - d_2}{\beta} - \frac{\gamma}{\beta}p_{t+1}^e - \frac{1}{\beta}u_t \tag{5.14}$$

thus is of the form (2.1). We assume $\{u_t\}$ to be a martingale difference sequence such that Assumption (A.2) is fulfilled. We assume furthermore that the agents believe in the auxiliary model[1]

$$p_{t+1} = \boldsymbol{\theta}'\boldsymbol{z}_t + w_{t+1} \tag{5.15}$$

[1] Of course, this auxiliary model is quite ad hoc and its use in economic theory has to be rationalized. One argument to support it is the observation that human beings tend to look always into the past to predict future events, even in situations where it is not appropriate. However, this model serves us only as an easy to understand application of our model setup.

with $\boldsymbol{z}_t = (p_t, 1)'$ and $\boldsymbol{\theta} \in I\!R^2$ and form their expectations according to $p^e_{t+1} = \boldsymbol{\theta}'_t \boldsymbol{z}_t$ with $\boldsymbol{\theta}_t$ given by the SG–algorithm. The resulting model equation is then

$$p_{t+1} = \frac{d_1 - d_2}{\beta} - \frac{\gamma}{\beta}\theta_{t,2} - \frac{\gamma}{\beta}\theta_{t,1} p_t - \frac{1}{\beta} u_t. \tag{5.16}$$

Let $w_t = -\beta^{-1} u_t$ and

$$a = -\frac{\gamma}{\beta} \qquad \text{and} \qquad \boldsymbol{\phi} = \begin{pmatrix} 0 \\ \frac{d_1 - d_2}{\beta} \end{pmatrix}. \tag{5.17}$$

Then (5.15) can be rewritten as

$$p_{t+1} = (\boldsymbol{\phi} + a\boldsymbol{\theta}_t)'\boldsymbol{z}_t + w_{t+1}. \tag{5.18}$$

Thus the (unique) rational expectations equilibrium parameter is

$$\bar{\boldsymbol{\theta}} = \begin{pmatrix} 0 \\ \frac{d_1 - d_2}{\beta + \gamma} \end{pmatrix}. \tag{5.19}$$

Clearly, the rational expectations equilibrium is a stable ARX–process (with non-existing AR–part[2]) hence, since

$$|a| \le 1 \iff \gamma \le \beta, \tag{5.20}$$

we obtain by Theorem 5.3 that $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s. if $\gamma \le \beta$. Thus whenever the usual cobweb stability condition is fulfilled agents learn to form rational expectations.

If $\gamma > \beta$ then we cannot infer that there is a.s. convergence. In fact, as our computer simulations of this model have shown there is convergence only with a positive probability. We observed, for the same initial values, as well convergent trajectories of $\{\boldsymbol{\theta}_t\}$ as divergent trajectories, depending only on the respective trajectories of the disturbance process $\{w_t\}$. Nevertheless, the qualitative long-term behaviour of $\{\boldsymbol{\theta}_t\}$ is not completely random. The simulation results suggest that the functions

$$\begin{aligned} \left|\frac{\gamma}{\beta}\right| &\mapsto \mathrm{P}(\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}) \\ \|\boldsymbol{\theta}_0 - \bar{\boldsymbol{\theta}}\| &\mapsto \mathrm{P}(\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}) \end{aligned} \tag{5.21}$$

are both (monotonically) decreasing[3]. □

[2]Notice that this example does not fit exactely into the setup of Chapter 2 since the information set the agents' predictions are based upon contains the sun-spot variable y_t. Nevertheless, the analysis of Chapter 3 and 4 holds also for this case.

[3]We observed convergent trajectories even for the highly unstable market situation represented by $\gamma/\beta = 4$. Nevertheless, most of the trajectories diverged with that parameter configuration.

Since Example 5.4 incorporates the somewhat implausible assumption that the agents believe in an autoregressive model although the REE is an non-autoregressive process[4] we want to give another example in which this assumption is more plausible.

Example 5.5:
We extend the previous model by introducing an additional type of firms into it. More precisely, we assume that one fraction, say a_2, of the producing firms are learning about the unknown hypothetical parameter as before while the other fraction, say a_1, uses the 'classical' or 'naive' expectations formation scheme $p^e_{t+1} = p_t$. The market expectation is then given as

$$p^m_{t+1} = a_1 p_t + a_2 \boldsymbol{\theta}_t{}' \boldsymbol{z}_t \tag{5.22}$$

with $a_1, a_2 > 0$, $a_1 + a_2 = 1$, $\boldsymbol{z}_t = (p_t, 1)'$ and $\{\boldsymbol{\theta}_t\}$ given by the SG-algorithm[5].

For this model the resulting model equation is

$$p_{t+1} = \frac{d_1 - d_2}{\beta} - \frac{\gamma}{\beta} a_2 \theta_{t,2} - \frac{\gamma}{\beta}(a_1 + a_2\theta_{t,1})p_t - \frac{1}{\beta}u_t. \tag{5.23}$$

Let $w_t = -\beta^{-1} u_t$ and

$$a = -a_2 \frac{\gamma}{\beta} \qquad \text{and} \qquad \boldsymbol{\phi} = \begin{pmatrix} -\frac{a_1\gamma}{\beta} \\ \frac{d_1 - d_2}{\beta} \end{pmatrix}. \tag{5.24}$$

Then (5.23) can be rewritten as (5.18), hence the model fits into our setup. The rational expectations equilibrium parameter is easily calculated as

$$\bar{\boldsymbol{\theta}} = \begin{pmatrix} -\frac{a_1\gamma}{\beta + a_2\gamma} \\ \frac{d_1 - d_2}{\beta + a_2\gamma} \end{pmatrix}. \tag{5.25}$$

In order to derive sufficient conditions for a.s. convergence of $\{\boldsymbol{\theta}_t\}$ notice that

$$|a| \le 1 \iff a_2 \frac{\gamma}{\beta} \le 1 \tag{5.26}$$

and that the characteristic polynomial possesses only the root $-a_1\gamma/(\beta + a_2\gamma)$. Since

$$\left| -\frac{a_1\gamma}{\beta + a_2\gamma} \right| < 1 \iff (a_1 - a_2)\frac{\gamma}{\beta} < 1 \tag{5.27}$$

and

$$\frac{1}{a_2} < \frac{1}{a_1 - a_2} \iff a_2 > \frac{1}{3} \tag{5.28}$$

we obtain in view of Theorem 5.3 as sufficient conditions for $\boldsymbol{\theta}_t \to \bar{\boldsymbol{\theta}}$ a.s.

$$\frac{\gamma}{\beta} \le 1 \qquad \text{or} \qquad a_2 \frac{\gamma}{\beta} \le 1 \quad \wedge \quad a_2 > \frac{1}{3}. \tag{5.29}$$

---

[4] See Bray (1983) for a different treatment of this model.

[5] Notice that in this model the assumption that agents believe in an autoregressive auxiliary model is more plausible since the fraction of 'classical' firms establishes an autoregressive relationship between p_t and p_{t+1}.

Under these conditions the REE is a stable autoregressive process and the SG-algorithm generates consistent estimates. If

$$\frac{1}{a_2} \leq \frac{\gamma}{\beta} \leq \frac{1}{a_1 - a_2} \tag{5.30}$$

then the REE is unstable and we obtain only $\theta_{t,1} \to \bar{\theta}_1$ a.s.

These results can be interpreted in the following way. Firstly, under the usual cobweb stability condition the price process is stable and the firms following the SG-learning procedure learn to form rational expectations with probability one.

Secondly, the learning firms have a stabilizing effect in the sense that even for parameter configurations which reflect an unstable market situation, i.e., if $\gamma/\beta > 1$, the resulting price process can be a stable autoregressive process. Therefore it is necessary that, on the one hand, the fraction of learning firms is sufficiently small to ensure $\gamma/\beta \leq 1/a_2$ (so that the estimates converge) and, on the other hand, that it is sufficiently large to ensure $(1-2a_2)\frac{\gamma}{\beta} < 1$ (so that the REE is stable). Notice that whenever $a_2 > 1/2$, thus whenever the learning firms 'dominate' the market, the REE is always stable.

But also the 'classical' firms have a stabilizing effect. Suppose that β and γ are given such that $\gamma/\beta > 1$. Then we know by Example 5.4 that without 'classical' firms in the market the SG-learning procedure does not converge with probability one. But if there are 'classical' firms in the market in a quantity such that $\gamma/\beta \leq 1/a_2 = 1/(1-a_1)$ and $(2a_2 - 1)\frac{\gamma}{\beta} = (1-2a_2)\frac{\gamma}{\beta} < 1$ holds the SG-learning procedure converges with probability one.

Finally, we want to remark that in an REE neither the market prediction nor the predictions of the 'classical' firms are rational, only the predictions of the learning firms are rational[6]. Hence the 'classical' firms do not participate in the success of learning of the other firms. □

[6] This property seems to contradict the definition of the REE in Chapter 2. But we understand the 'classical' firms as part of the economic surrounding since they are completely ignorant with repsect to the evolution of the price process and persist robot-like in their expectations formation scheme.

References

Başar, T. (1989)
Some Thoughts on Rational Expectations Models, and Alternate Formulations
Computers Math. Applic. **18**, 591–604

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

Bray, M.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

Chang, M.C., C.Y. Cyrus Chu and K. Lin (1991a)
The Global Stability of Least Square Learning
Discussion Paper 9106, Academia Sinica, Taipei

Chang, M.C., C.Y. Cyrus Chu and K. Lin (1991b)
A Reformulation of the Least Square Learning Mechanism
Discussion Paper 9107, Academia Sinica, Taipei

Chang, M.C., C.Y. Cyrus Chu and K. Lin (1991c)
The Recursive Prediction Error Learning: An Alternative to Least Square Learning
Discussion Paper, Academia Sinica, Taipei

Chen, H. and L. Guo (1991)
Identification and Stochastic Adaptive Control
Birkhäuser, Boston

Chow, Y.S. (1965)
Local convergence of martingales and the law of large numbers
Ann. Math. Stat., **36**, 552–558

Cyert, R.M. and M.H. DeGroot (1974)
Rational Expectations and Bayesian Analysis
Journal of Political Economy, **82**, 521–536

Fourgeaud, C., C. Gourieroux, and J. Pradel (1986)
Learning Procedures and Convergence to Rationality
Econometrica, **54**, 845–868

Frydman, R. and E.S. Phelps (1983)
Individual forecasing and aggregate outcomes
Cambridge University Press, Cambridge

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

KOTTMANN, T. and I. KULIBERDA (1990)
Autoregressive Models with Forecast Feedback: A Monte–Carlo–Study and First Theoretical Results
Discussion Paper B–160, Universität Bonn

LAI, T.L. (1989)
Extended Stochastic Ljapunov functions and recursive algorithms in stochastic linear 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 (1982a)
Least squares estimates in stochastic regression models with application to identification and control of dynamic systems
The Annals of Statistics, **10**, 154–166

LAI, T.L. and C.Z. WEI (1982b)
Asymptotic Properties of Projections with Applications to Stochastic Regression Problems
Journal of Multivariate Analysis, **12**, 346–370

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

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

PESARAN, M.H. (1987)
The Limits to Rational Expectations
Basil Blackwell, Oxford

REN, W. (1991)
Stochastic Adaptive System Theory for Identification, Filtering, Prediction and Control
Thesis, University of Illinois at Urbana-Champaign

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

STOUT, W.F. (1974)
Almost Sure Convergence
Academic Press, New York

WALK, H. (1985)
Almost Sure Convergence of Stochastic Approximation Processes
Statistics and Decisions, Supplement Issue No. 2, 137–141

WEI, C.Z. (1987)
Adaptive prediction by least squares predictors in stochastic regression models with application to time series
The Annals of Statistics, **15**, 1667–1682

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

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