

# LEARNING IN LINEAR MODELS WITH EXPECTATIONAL LEADS

By Jan Wenzelburger\*

Fakultät für Wirtschaftswissenschaften  
Universität Bielefeld  
Postfach 100 131  
D-33501 Bielefeld  
Germany

Discussion paper No. 478

## Abstract

This paper provides an adaptive learning algorithm for linear stochastic models with expectational leads in which forecasts for an arbitrary period ahead of the current state feed back into the economic system. The concept of an unbiased forecasting rule which generates rational expectations equilibria is introduced. It is shown that the learning scheme provides an arbitrary precise approximation of an unbiased forecasting rule such that all trajectories generated by the scheme converge to rational expectations equilibria globally for all initial conditions. We strengthen convergence results in relaxing standard assumptions and in providing conditions ensuring algorithm convergence which are much easier to verify and to interpret as compared to those known previously in the literature.

Keywords: Adaptive learning, rational expectations, recursive estimation, stochastic approximation.

JEL Classification: C13, C30, D83

First version: Nov. 1999, this version: Jan. 2002.

---

\*ACKNOWLEDGEMENT. This research was supported by the Deutsche Forschungsgemeinschaft under contract No. Bo 635/9-3. I am indebted to Volker Böhm, Ludwig Elsner, Jochen Jungeilges, and Thorsten Pampel for many valuable discussions.

# 1 Introduction

Models with expectational leads describe economic scenarios in which forecasts about state variables further than one period ahead of the current state feed back into the system. Expectational leads are intrinsic in situations in which the planning horizon of agents' intertemporal decisions is longer than two periods and occur naturally in overlapping generations with more than two-period lives. In macroeconomics, these may be related to announcement effects, describing the impact of possible future actions by an authority such as a government or a bank on the current state, where the pursued policy rule is unknown to agents. As optimal decisions of agents require a detailed knowledge of all the relevant economic relations, much of the literature, with the pioneering work of Bray (1982), Bray & Savin (1986) and others, is concerned with analyzing adaptive learning algorithms designed to converge to rational expectations solutions.

For models with two-period-ahead forecasts, the existence of such rational-expectations solutions and methods to learn these from historical data are understood quite well. However, most of the convergence results in the literature are of a local nature such that initial guesses of the system's parameters may have to be quite close to their true values in order to assure convergence, cf. Evans & Honkapohja (2001). A few contributions, as Chen & White (1998), Chatterji & Chattopadhyay (2000), or Wenzelburger (2001a, b), succeeded in providing global results for a specialized class of nonlinear Cobweb-type models in which the endogenous variables do not feed back into the economic law and which contain at most two-period-ahead forecasts. The special nature of these models indicate that there is still need for more general convergence results. Moreover, the learning literature has rarely addressed the question of generalizing these concepts to multivariate systems with expectational leads of an arbitrarily finite length, an issue which is much more complex than in a one or two period-ahead situation.

This paper attempts to fill this gap to some extent by applying a geometric approach presented in Böhm & Wenzelburger (1999, 2000a,b) to linear multivariate models with expectational leads of an arbitrarily finite length. We distinguish between an *economic law* describing the basic market mechanism of an economy and a *forecasting rule* which models the way in which a (forecasting) agency forms expectations. The combination of both ingredients may be seen as an (*economic*) *random dynamical system* in the sense of Arnold (1998) which is defined explicitly and globally on the whole state space. Following previous work, we generalize the notion of an unbiased forecasting rule which by definition generates trajectories with rational expectations in the classical sense that all forecast errors conditional on the available information vanish on average. In other words, unbiased forecasting rules should be seen as the generators for rational expectations equilibria and as such provide an alternative way of characterizing these, an issue which is still unresolved in the literature.

A structural key feature of systems with expectational leads is the fact that several forecasts, each formed at a different time, pertain to the same realization of an endogenous

variable. This is a primary cause for an inherent multiplicity of unbiased forecasting rules and thus of solutions to rational expectations models. Extending Böhm & Wenzelburger (2000a), we shed new light on this phenomenon by showing that a mere invertibility condition of the economic law (nonlinear with additive noise) is responsible for the existence of an unbiased forecasting rule which never updates forecasts formed at previous dates. In the linear case, the existence for such a *no-updating rule* depends on the non-singularity of a particular parameter matrix and is surprisingly easy to compute whenever it exists. It yields the most precise forecasts in the sense that all forecast errors conditional on information available at a particular date vanish, including those which have been formed at a stage in which that information was not available yet. This phenomenon is caused by the feedback nature of the forecasts and indeed holds true only for forecasts which feed into to system in a non-trivial manner. We relate the concept of a (linear) unbiased forecasting rule to minimal-state-variable solutions in the sense of McCallum (1983, 1998, 1999) which, in the context of this paper, will be generated by special forecasting rules. The main advantage of this approach is that it distinguishes in a clear-cut fashion between the quality of the forecasts and the dynamic stability of the economic system in which the forecasts feed into.

The proposed setup is the basis for developing an adaptive learning scheme for linear stochastic models which is based on the recursive extended-least-squares algorithm well-established in the engineering literature, cf. Caines (1988). The main difference between our learning scheme and the approach of the literature is that at each point in time, we first estimate the whole system including the feedback of all relevant forecasts and then compute an approximation of an unbiased forecasting rule. This allows us to keep the system dynamically stable at all stages of the learning procedure without any a-priori choice of a particular forecasting rule which initially is unknown. Contrary to Evans & Honkapohja (2001, p. 173), an approximation of the preferred forecasting rule can, in principle, be selected on the basis of an adequate economic reasoning, as soon as parameter estimates are sufficiently precise. In particular, (approximations to) minimal-state-variable solutions, which are dynamically stable, may be chosen whenever these exist, cf. McCallum (1999) for a further discussion of this issue. The idea of estimating the expectations feedback is already implicit in Kuan & White (1994) who provide a local convergence result for a class of nonlinear Cobweb-type models (i.e., models in which the endogenous variables do not feed back into the economic law). Unfortunately they give no clue on how to select between multiple solutions inherent in models with expectational leads.

Using ideas from the literature on adaptive control and optimal tracking, this paper applies results of several papers by Lai and Wei, notably their (1986b) paper. These authors pursue a martingale approach paralleling and extending the original approach of Ljung (1977) who instead uses a limiting non-random ordinary differential equation (ODE) for his convergence analysis. We relax standard assumptions of the learning literature in several respects. First, we admit leads of arbitrary length. Second, we

generalize the class of stochastic processes which governs the exogenous observable variables, where, in addition, we do not assume that the process is known. Third, we admit colored noise for the unobservable stochastic perturbations acting on the system. This generalizes the usual white-noise assumption and seems to be well known in the literature on adaptive control, cf. Caines (1988). Under simple and standard identifiability conditions (which in the white-noise case are generically satisfied), we show convergence of our learning scheme globally for all initial conditions. This strengthens well-known convergence results from the literature, because the conditions require relatively mild technical prerequisites and are much easier to verify and interpret as compared to those based on Ljung's ODE approach.

The focus of the present paper is on linear models, as they provide a benchmark dynamics for their nonlinear generalizations the appropriate treatment of which is beyond the scope of the present paper. Linear models are popular in many economic applications, either as exact formulations or as linear approximations of originally nonlinear specifications. Nevertheless, our results show that the proposed methodology carries over to the nonlinear case (for a first example, see Wenzelburger 2001b) and in future work should be combined with nonparametric estimation techniques such as provided in Chen & White (1998).

The plan of the paper is as follows. The general assumptions of the model are given in Sec. 2. Unbiased forecasting rules and their relation to minimal-state-variable solutions are discussed in Secs. 3 and 4. The adaptive learning scheme is introduced in Sec. 5, convergence of the scheme is investigated in Sec. 6, and conclusions are given in Sec. 7. All technical proofs which could distract from the main issues of the paper are collected in an appendix.

## 2 Linear Models

Consider the case of a linear economic law of the form

$$(1) \quad y_t = F(y_{t-1}, \dots, y_{t-n_1}, y_{t-1, t+n_2}^e, \dots, y_{t-1, t}^e) + \xi_t,$$

where

$$F(y_{t-1}, \dots, y_{t-n_1}, y_{t-1, t+n_2}^e, \dots, y_{t-1, t}^e) := \sum_{i=1}^{n_1} A^{(i)} y_{t-i} + \sum_{i=0}^{n_2} B^{(n_2-i)} y_{t-1, t+i}^e$$

is a linear function of all arguments.  $y_t \in \mathbb{R}^d$  denotes a vector of endogenous variables describing the state of the economy at date  $t$ ,  $A^{(i)}$ ,  $B^{(j)}$  are non-random  $d \times d$  matrices, and  $\{\xi_t\}_{t \in \mathbb{N}}$  is a  $\mathbb{R}^d$ -valued stochastic process on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  adapted to the filtration  $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$ .  $y_{t-1, t+i}^e \in \mathbb{R}^d$ ,  $i = 0, \dots, n_2$  are forecasts for future realizations

$y_{t+i}$ ,  $i = 0, \dots, n_2$ , respectively, which are based upon information available at date  $t - 1$  and hence are assumed to be  $\mathcal{F}_{t-1}$ -measurable.

The stochastic perturbations acting on the system are given by

$$(2) \quad \xi_t = b + \sum_{i=1}^{n_3} C^{(i)} w_{t-i} + \epsilon_t + \sum_{i=1}^{n_4} D^{(i)} \epsilon_{t-i},$$

where  $\{w_t\}_{t \in \mathbb{N}}$  describes the process of  $\mathbb{R}^d$ -valued exogenous variables, which is assumed to be observable,  $\{\epsilon_t\}_{t \in \mathbb{N}}$  is a  $\mathbb{R}^d$ -valued martingale difference sequence describing an unobservable noise process, the  $C^{(i)}$ ,  $D^{(i)}$  are non-random  $d \times d$  matrices, and  $b \in \mathbb{R}^d$  is a constant non-random vector. (1) together with (2) provide a quite general specification of a linear economic model with rational expectations, the complete treatment of which seems as yet unaccomplished, cf. Evans & Honkapohja (2001, p. 173). Most models treated therein have  $n_1 = 1$ ; a relatively straightforward extension of a financial market model treated in Böhm & Chiarella (2000) or Wenzelburger (2001b) may serve as a first example for  $n_1 \geq 1$ . (Consider traders with a planning horizon which is longer than two periods and who trade just once in a lifetime.)

To fix the probabilistic assumptions on the two exogenous processes, let  $\lambda_{\min}(B)$  denote the minimum eigenvalue of a symmetric matrix  $B$ .

**Assumption 2.1** *The noise process  $\{\epsilon_t\}_{t \in \mathbb{N}}$  is a  $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$ -adapted martingale difference sequence on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with values in  $\mathbb{R}^d$  which satisfies the following.<sup>1</sup>*

- (i)  $\sup_{t \geq 1} \mathbb{E}[\|\epsilon_t\|^\alpha \mid \mathcal{F}_{t-1}] < \infty$   $\mathbb{P}$ -a.s. for some  $\alpha > 2$ .
- (ii)  $\sum_{t=1}^T \|\epsilon_t\|^2 = O(T)$  and<sup>2</sup>  $\liminf_{T \rightarrow \infty} T^{-1} \lambda_{\min} \left( \sum_{t=1}^T \epsilon_t \epsilon_t^\top \right) > 0$   $\mathbb{P}$ -a.s.
- (iii)  $D^{(1)}, \dots, D^{(n_4)}$ ,  $n_4 \geq 0$  are non-random  $d \times d$  matrices and the matrix polynomial  $\Gamma(s) = I_d + D^{(1)}s + \dots + D^{(n_4)}s^{n_4}$ ,  $s \in \mathbb{C}$ , where  $I_d$  denotes the  $d \times d$  identity matrix, is strictly positive real, that is,
  - (a)  $\det \Gamma(s) \neq 0$  for all  $s \in \mathbb{C}$  with  $|s| \leq 1$ .
  - (b) For each  $s \in \mathbb{C}$  with  $|s| \leq 1$ , the matrix  $\Gamma(s) + \Gamma^\top(\bar{s})$ , with  $\bar{s} \in \mathbb{C}$  denoting the complex conjugate of  $s$ , is strictly positive definite, that is,  $x^\top (\Gamma(s) + \Gamma^\top(\bar{s})) x > 0$  for all  $0 \neq x \in \mathbb{R}^d$ .

---

<sup>1</sup>By  $a_t = O(b_t)$   $\mathbb{P}$ -a.s. for two sequences of random numbers  $(a_t)$  and  $(b_t)$ , we mean that  $\sup_{t \in \mathbb{N}} \left| \frac{a_t}{b_t} \right| \leq c$   $\mathbb{P}$ -a.s. for some positive constant  $c$ .

<sup>2</sup> It is shown in Lai & Wei (1983) that (ii) follows from (i) and the weaker assumption

$$(ii') \quad \liminf_{t \rightarrow \infty} \lambda_{\min}(\mathbb{E}[\epsilon_t \epsilon_t^\top \mid \mathcal{F}_{t-1}]) > 0 \quad \mathbb{P} - \text{a.s.}$$

Assumption 2.1 includes the white-noise case  $\Gamma(s) = I_d$  ( $n_4 = 0$ ) which is assumed in most economic models, cf. Evans & Honkapohja (2001). This generalization seems to be quite standard in the engineering literature (see Caines 1988) and is referred to as colored noise.

**Assumption 2.2** *The exogenous variables  $\{w_t\}_{t \in \mathbb{N}}$  are given by a  $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$ -adapted stochastic process on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with values in  $\mathbb{R}^d$ . Let  $\mathcal{F}'_t := \sigma(\mathcal{F}_{t-1}, \epsilon_t)$  denote the  $\sigma$ -algebra generated by  $\mathcal{F}_{t-1}$  and  $\epsilon_t$ . Setting  $\tilde{w}_t = w_t - \mathbb{E}[w_t | \mathcal{F}'_t]$ , the process  $\{w_t\}_{t \in \mathbb{N}}$  is assumed to satisfy the following.*

- (i)  $\sup_{t \geq 1} \mathbb{E}[\|\tilde{w}_t\|^\beta | \mathcal{F}'_t] < \infty \quad \mathbb{P} - \text{a.s. for some } \beta > 2.$
- (ii)  $\sum_{t=1}^T \|\tilde{w}_t\|^2 = O(T) \quad \text{and} \quad \liminf_{T \rightarrow \infty} T^{-1} \lambda_{\min} \left( \sum_{t=1}^T \tilde{w}_t \tilde{w}_t^\top \right) > 0 \quad \mathbb{P} - \text{a.s.}$
- (iii)  $\sum_{t=1}^T \|\mathbb{E}[w_t | \mathcal{F}'_t]\|^2 = O \left( \lambda_{\min} \left( \sum_{t=1}^T \tilde{w}_t \tilde{w}_t^\top \right) \right) \quad \mathbb{P} - \text{a.s.}$

An analogous remark to that of Footnote 2 applies to Assumption 2.2. Assumption 2.2 (ii) implies  $\lambda_{\min} \left( \sum_{t=1}^T \tilde{w}_t \tilde{w}_t^\top \right) = O(T)$  and thus  $\sum_{t=1}^T \|w_t\|^2 = O(T)$   $\mathbb{P}$ -a.s. Observe also, that  $w_t$  may have constant non-random entries.

In the sequel, we will restrict our analysis to forecasting rules  $\varphi = (\varphi^{(1)}, \dots, \varphi^{(n_2+1)})$  of a particular functional form, where for each  $i = 1, \dots, n_2 + 1$ ,

$$(3) \quad \begin{aligned} \varphi^{(i)} : \mathbb{R}^d \times \mathbb{R}^{dn_1} \times \mathbb{R}^{d(n_2+1)} &\rightarrow \mathbb{R}^d, \\ y_{t-1, t-1+i}^e &:= \varphi^{(i)}(\mathbb{E}_{t-1}[\xi_t], y_{t-1}, \dots, y_{t-n_1}, y_{t-2, t-1+n_2}^e, \dots, y_{t-2, t-1}^e) \end{aligned}$$

with  $\mathbb{E}_{t-1}[\xi_t]$  denoting the expected value of the exogenous perturbations conditional on  $\mathcal{F}_{t-1}$ . A forecasting rule  $\varphi$  could, of course, depend on an information set larger than implied by (3). It will turn out in Sec. 3 that the functional form (3) includes all information necessary to obtain rational expectations along a trajectory of the system.

Assumptions 2.1 and 2.2 imply that the system (1) under the exogenous perturbations (2) is stable in the sense that all realizations  $y_t$  are  $\mathbb{P}$ -a.s. bounded, provided that all forecasts  $y_{t,t+j}^e$  are a.s. bounded and that the matrix

$$(4) \quad \mathcal{A} = \begin{pmatrix} A^{(1)} & \dots & \dots & \dots & A^{(n_1)} \\ I_d & 0 & \dots & \dots & 0 \\ 0 & I_d & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & I_d & 0 \end{pmatrix}$$

associated with (1) has all eigenvalues inside the unit circle. This is the contents of the following.

**Theorem 2.1** Consider the feedback system defined by (1), (2), and (3). Assume that the following hypotheses are satisfied.

(i) All eigenvalues of  $\mathcal{A}$  given in (4) have modulus less or equal to unity.

(ii) Assumption 2.1 (ii) and Assumptions 2.2 (ii) are fulfilled.

(iii)  $\sum_{t=1}^T \|y_{t,t+j}^e\|^2 = O(T)$   $\mathbb{P}$  - a.s. for all  $j = 1, \dots, n_2 + 1$ .

Then the system (1)-(3) is stable in the sense that  $\mathbb{P}$ -a.s.

$$(i) \|x_t\| = O(t^\alpha) \text{ for some } \alpha > 0 \quad \text{and} \quad (ii) \sum_{t=1}^T \|x_t\|^2 = O(T),$$

where  $x_t = (y_t^\top, \dots, y_{t-n_1+1}^\top, y_{t,t+1+n_2}^{e\top}, \dots, y_{t,t+1}^{e\top}, w_t^\top, \dots, w_{t-n_3+1}^\top, \epsilon_t^\top, \dots, \epsilon_{t-n_4+1}^\top)^\top$ .

The proof of Theorem 2.1 is given in Appendix A.1. Theorem 2.1 will become important when introducing our learning scheme, because it guarantees dynamic stability of the whole economic system as long as the forecasts are on average bounded.

### 3 Unbiased forecasting rules

In this section we will develop the notion of an unbiased forecasting rule which is designed to generate rational expectations along all trajectories of the system under consideration or, in other words, rational expectations equilibria (REE). In order to derive such a forecasting rule, observe that in each period  $t$  there are  $n_2$  forecasts  $y_{t-i,t+1}^e$ ,  $i = 1, \dots, n_2$  for  $y_{t+1}$  which have been set at dates prior to  $t$ . The forecast errors for  $y_{t+1}$  conditioned on information available at the respective dates vanish, if and only if

$$(5) \quad \mathbb{E}_{t-i}[y_{t+1} - y_{t-i,t+1}^e] = 0 \quad \mathbb{P} - \text{a.s.}, \quad i = 0, \dots, n_2,$$

where  $\mathbb{E}_j$  denotes the conditional expectations operator with respect to  $\mathcal{F}_j$ . Condition (5) and the law of iterated expectations ( $\mathbb{E}_{t-i}[\mathbb{E}_t[y_{t+1}]] = \mathbb{E}_{t-i}[y_{t+1}]$ ,  $i \geq 0$ ) imply that all forecasts  $y_{t-n_2,t+1}^e, \dots, y_{t,t+1}^e$  for  $y_{t+1}$  have to satisfy the *consistency conditions*

$$(6) \quad \mathbb{E}_{t-i}[y_{t,t+1}^e] = y_{t-i,t+1}^e \quad \mathbb{P} - \text{a.s.}, \quad i = 1, \dots, n_2.$$

On the other hand, if all forecasts  $y_{t-n_2,t+1}^e, \dots, y_{t,t+1}^e$  obey (6), then the law of iterated expectations implies that (5) is automatically satisfied, provided that the most recent forecast  $y_{t,t+1}^e$  has a vanishing conditional forecast error, that is,  $\mathbb{E}_t[y_{t+1}] - y_{t,t+1}^e = 0$

$\mathbb{P}$ -a.s. Since (6) has to hold for all times  $t$  and analogously for all forecasts  $y_{t,t+j}^e$ , we can transform (6) to consistency conditions for the first  $n_2$  forecasts  $y_{t,t+1}^e, \dots, y_{t,t+n_2}^e$  formed at date  $t$ . These take the form

$$(7) \quad \mathbb{E}_{t-1}[y_{t,t+i}^e] - y_{t-1,t+i}^e = 0 \quad \mathbb{P} - \text{a.s.}, \quad i = 1, \dots, n_2,$$

where  $y_{t-1,t+1}^e, \dots, y_{t-1,t+n_2}^e$  are the respective forecasts formed at date  $t - 1$ .

In order to express our consistency conditions (7) in terms of forecasting rules, let  $Y_t = (y_t, \dots, y_{t-n_1+1}) \in \mathbb{R}^{dn_1}$  denote the vector of past realizations and  $Y_{t-1}^e = (y_{t-1,t+n_2}^e, \dots, y_{t-1,t}^e) \in \mathbb{R}^{d(n_2+1)}$  the vector of past forecasts. Consider forecasting rules of the form  $\varphi = (\varphi^{(1)}, \dots, \varphi^{(n_2+1)})$ , where for each  $i = 1, \dots, n_2 + 1$ ,

$$(8) \quad \varphi^{(i)} : \mathbb{R}^d \times \mathbb{R}^{dn_1} \times \mathbb{R}^{d(n_2+1)} \rightarrow \mathbb{R}^d, \quad y_{t,t+i}^e := \varphi^{(i)}(\mathbb{E}_t[\xi_{t+1}], Y_t, Y_{t-1}^e).$$

The applicability of the forecasting rule (8) depends crucially on the observability of  $\xi_t$  and hence of  $\epsilon_t$  which is problem in the case of general colored noise ( $n_4 > 0$ ). It will be shown in Sec. 5 how this missing information can be obtained from historical data.<sup>3</sup> Consistent forecasting rules can now be defined as follows.

**Definition 3.1** *A forecasting rule  $\varphi = (\varphi^{(1)}, \dots, \varphi^{(n_2+1)})$  of type (8) is called consistent at  $(Y_t, Y_{t-1}^e) \in \mathbb{R}^{d(n_1+n_2+1)}$  in period  $t$ , if for each  $i = 1, \dots, n_2$ ,*

$$(9) \quad \mathbb{E}_{t-1}[\varphi^{(i)}(\mathbb{E}_t[\xi_{t+1}], Y_t, Y_{t-1}^e)] = y_{t-1,t+i}^e, \quad \mathbb{P} - \text{a.e.}$$

$\varphi$  is called consistent, if (9) is satisfied for all  $(Y_t, Y_{t-1}^e) \in \mathbb{R}^{d(n_1+n_2+1)}$  and all times  $t$ .

Definition 3.1 provides a minimum requirement a forecasting rule has to meet in order to generate rational expectations, since otherwise (6) cannot be satisfied for all forecasts. Using the economic law  $F$  and a consistent forecasting rule  $\varphi$ , the future state of the economy is determined by the *time-one map*  $F_\varphi(\xi_t, \mathbb{E}_t[\xi_{t+1}], \cdot)$ , defined by

$$(10) \quad \begin{aligned} F_\varphi(\xi_t, \mathbb{E}_t[\xi_{t+1}], \cdot) : \mathbb{R}^{dn_1} \times \mathbb{R}^{d(n_2+1)} &\rightarrow \mathbb{R}^{dn_1} \times \mathbb{R}^{d(n_2+1)} \\ \begin{pmatrix} Y_{t-1} \\ Y_{t-1}^e \end{pmatrix} &\mapsto \begin{pmatrix} (F(Y_{t-1}, Y_{t-1}^e) + \xi_t, \text{pr}_{-1} Y_{t-1}) \\ \varphi(\mathbb{E}_t[\xi_{t+1}], F(Y_{t-1}, Y_{t-1}^e) + \xi_t, Y_{t-1}^e) \end{pmatrix}, \end{aligned}$$

where  $\text{pr}_{-1}$  denotes the projection onto the first  $n_1$  components. Thus, given  $(Y_{t-1}, Y_{t-1}^e)$ ,

$$(Y_t, Y_t^e) = F_\varphi(\xi_t, \mathbb{E}_t[\xi_{t+1}], Y_{t-1}, Y_{t-1}^e)$$

---

<sup>3</sup> By (2),  $\mathbb{E}_t[\xi_{t+1}] = b + \sum_{i=1}^{n_3} C^{(i)} w_{t+1-i} + \sum_{i=1}^{n_4} C^{(i)} \epsilon_{t+1-i}$  such that  $\mathbb{E}_t[\xi_{t+1}]$  in the expression for  $\varphi$  could be replaced by the list  $(w_t, \dots, w_{t-n_3+1}, \epsilon_t, \dots, \epsilon_{t-n_4+1})$ . This remark is valid for all subsequent functions involving  $\mathbb{E}_t[\xi_{t+1}]$ .



describes the next (lagged) state  $Y_t$  of the economy together with all relevant forecasts  $Y_t^e$  at date  $t$ .

To see which forecasting rules generate rational expectations, notice first that the expected value of the future state  $y_{t+1}$  is

$$\mathbb{E}_t[y_{t+1}] = F(Y_t, Y_t^e) + \mathbb{E}_t[\xi_{t+1}].$$

The forecast error  $\mathbb{E}_t[y_{t+1}] - y_{t,t+1}^e$  conditioned on information available at date  $t$  is then obtained from the (*mean*) *error function* associated with the economic law (1) which is given by a function  $e_F : \mathbb{R}^d \times \mathbb{R}^{dn_1} \times \mathbb{R}^{d(n_2+1)} \rightarrow \mathbb{R}^d$ , defined by

$$(11) \quad e_F(\mathbb{E}_t[\xi_{t+1}], Y_t, Y_t^e) := F(Y_t, Y_t^e) + \mathbb{E}_t[\xi_{t+1}] - y_{t,t+1}^e,$$

where  $Y_t^e = (y_{t,t+1+n_2}^e, \dots, y_{t,t+1}^e)$ . Given an arbitrary state of the economy  $Y_t$ , the error function describes all possible mean forecasts errors, regardless of which forecasting rule has been applied. Geometrically, the graph of the error function (11) is a hyper-plane in  $\mathbb{R}^{d(n_1+n_2+3)}$ . Given arbitrary forecasts  $Y_t^e = (y_{t,t+1+n_2}^e, \dots, y_{t,t+1}^e)$ , the conditional forecast error  $\mathbb{E}_t[y_{t+1}] - y_{t,t+1}^e$  vanishes, if and only if

$$(12) \quad \mathbb{E}_t[y_{t+1}] - y_{t,t+1}^e = e_F(\mathbb{E}_t[\xi_{t+1}], Y_t, Y_t^e) = 0 \quad \mathbb{P} - \text{a.s.}$$

A forecasting rule which for given  $(Y_t, Y_{t-1}^e) \in \mathbb{R}^{d(n_1+n_2+1)}$  sets  $Y_t^e = \varphi(\mathbb{E}_t[\xi_{t+1}], Y_t, Y_{t-1}^e)$  such that (12) holds is called unbiased at  $(Y_t, Y_{t-1}^e)$ . A forecasting rule which is unbiased along all trajectories can now be defined as follows.

**Definition 3.2** *A consistent forecasting rule  $\varphi = (\varphi^{(1)}, \dots, \varphi^{(n_2+1)})$  is called (globally) unbiased, if there exists a family  $\mathcal{U} = \{\mathcal{U}(\eta)\}_{\eta \in \mathbb{R}^d}$  of non-empty subsets  $\mathcal{U}(\eta) \subset \mathbb{R}^{dn_1} \times \mathbb{R}^{d(n_2+1)}$ , such that the following holds true:*

- (i) *For each  $\eta \in \mathbb{R}^d$ ,  $e_F(\eta, Y, Y^e) = 0$  for all  $(Y, Y^e) \in \mathcal{U}(\eta)$ .*
- (ii) *For each  $\eta \in \mathbb{R}^d$ ,  $\mathcal{U}(\eta)$  contains  $\mathbb{R}^{dn_1} \times \{0\}$  as a subset.*
- (iii)  *$\mathcal{U}$  is invariant under  $F_\varphi$  defined in (10), that is, for each  $(Y, Y^e) \in \mathcal{U}(\mathbb{E}_{t-1}[\xi_t])$ ,*

$$F_\varphi(\xi_t, \mathbb{E}_t[\xi_{t+1}], Y, Y^e) \in \mathcal{U}(\mathbb{E}_t[\xi_{t+1}]) \quad \mathbb{P} - \text{a.s. for all } t \in \mathbb{N}.$$

Condition (i) describes states  $(\mathbb{E}_{t-1}[\xi_t], Y, Y^e)$  in which the forecast error (12) vanishes, Condition (ii) states that (12) can be satisfied globally for all  $Y \in \mathbb{R}^{dn_1}$ , and Condition (iii) implies that (12) can be satisfied for all times  $t$ . If  $\varphi_\star$  is unbiased in the sense of Definition 3.2, then all trajectories  $\gamma(Y_0, Y_0^e) := \{Y_t, Y_t^e\}_{t \in \mathbb{N}}$  with arbitrary initial conditions  $(Y_0, Y_0^e) \in \mathcal{U}(\mathbb{E}_0[\xi_1])$  which are generated by the induced map  $F_{\varphi_\star}$  given in (10) correspond precisely to rational expectations equilibria (REE).

Definitions 3.1 and 3.2 are valid for general non-linear economic laws  $F$  with additive perturbations. The existence of an unbiased forecasting rule is guaranteed if the error function  $e_F$  satisfies a global invertibility condition. To be specific, we exploit now the linear structure of (1) and construct such a forecasting rule as follows. Suppose  $B^{(0)}$  is invertible and define

$$(13) \quad \psi(\eta, Y, Z) := -B^{(0)-1} \left[ \sum_{i=1}^{n_1} A^{(i)} Y^{(i)} + \sum_{i=0}^{n_2-1} B^{(n_2-i)} Z^{(i+1)} - Z^{(1)} + \eta \right]$$

with  $Y = (Y^{(1)}, \dots, Y^{(n_1)})$  and  $Z = (Z^{(1)}, \dots, Z^{(n_2)})$ . Then  $e_F(\eta, Y, Z, \psi(\eta, Y, Z)) = 0$  identically on  $\mathbb{R}^{d(n_1+n_2+1)}$  and the function (13) induces a forecasting rule of the functional form (8) by setting

$$(14) \quad \begin{aligned} y_{t,t+1+n_2}^e &:= \psi(\mathbb{E}_t[\xi_{t+1}], Y_t, y_{t-1,t+n_2}^e, \dots, y_{t-1,t+1}^e), \\ y_{t,t+i}^e &:= y_{t-1,t+i}^e, \quad i = 1, \dots, n_2. \end{aligned}$$

The forecasting rule (14) is clearly consistent in the sense of Definition 3.1 and will be referred to as a *no-updating rule*. Inserting (14) into the system (1), the time-one map (10) takes the form

$$(15) \quad \begin{cases} y_{t+1} &= y_{t-1,t+1}^e + \xi_{t+1} - \mathbb{E}_t[\xi_{t+1}], \\ y_{t,t+1+n_2}^e &= -B^{(0)-1} \left[ \sum_{i=1}^{n_1} A^{(i)} y_{t+1-i} + \sum_{i=0}^{n_2-1} B^{(n_2-i)} y_{t-1,t+1+i}^e - y_{t-1,t+1}^e + \mathbb{E}_t[\xi_{t+1}] \right], \\ y_{t,t+i}^e &= y_{t-1,t+i}^e, \quad i = 1, \dots, n_2. \end{cases}$$

It is easy to check that all conditional forecast errors generated by (15) vanish, since by construction  $E_t[y_{t+1}] = y_{t-1,t+1}^e$  and hence all forecasts for  $y_{t+1}$  satisfy

$$(16) \quad \mathbb{E}_t[y_{t+1}] = y_{t-i,t+1}^e \quad \mathbb{P} - \text{a.s.}, \quad i = 0, \dots, n_2,$$

as long as the *no-updating rule* (14) has been applied for the past  $n_2 + 1$  periods. Setting  $\mathcal{U}(\eta) = \mathbb{R}^d \times \mathbb{R}^d$  for all  $\eta \in \mathbb{R}^d$ , this shows that the no-updating rule (14) is unbiased in the sense of Definition 3.2 and all trajectories of (15) correspond to REE in the classical sense.<sup>4</sup>

A remarkable fact in economic systems with expectational leads ( $n_2 > 0$ ) is that for invertible  $B^{(0)}$ , the no-updating forecasting rule (14) generates rational expectations in the sense that **all** forecast errors (16) conditioned on information available at date  $t$  vanish, including those which have been made prior to that date. Thus, an unbiased no-updating rule yields the most precise forecasts in the sense that forecast errors vanish conditioned on information which is not available at the stage in which they have been

---

<sup>4</sup>Notice that an unbiased no-updating rule for nonlinear  $F$  is formally defined in the very same way.

issued. This peculiar phenomenon is caused by the feedback nature of the forecasts and indeed holds true only for forecasts which feed into the system in a non-trivial manner. Since the forecasts are simply inputs which feed back into the system (1), the no-updating rule (14) is closely related to the classical Åström-Wittenmark self-tuning regulator (Åström & Wittenmark 1973), known from the literature on adaptive control and optimal tracking, see e.g. Caines (1988).

The no-updating rule (14) is constructed in such a way that  $\mathbb{E}_t[y_{t+1}] = y_{t,t+1}^e$  for all choices of  $y_{t,t+i}^e$ ,  $i = 1, \dots, n_2$ . This implies a considerable freedom in selecting these first  $n_2$  ( $n_2 > 0$ ) forecasts, because any martingale difference sequence could, in principle, be added to these forecasts without violating the unbiasedness. It constitutes the primary cause for the emergence of multiple unbiased forecasting rules and thus of REE in the presence of expectational leads.

There are well-known examples in which  $B^{(0)}$  is not invertible, see Evans & Honkapohja (2001, Chap. 10). In these cases (14) does not exist and the construction of unbiased forecasting rules may become quite involved. However, the first condition in Definition 3.2 implies local invertibility of the error function and one way out of the problem with singular  $B^{(0)}$  may be as discussed in Appendix A.2. Another approach is presented in the next section.

## 4 MSV predictors

In the literature on rational expectations, the notion of minimal-state-variable solutions is a popular concept to describe situations in which agents have rational expectations along particular trajectories of the system, cf. McCallum (1983, 1998, 1999). We relate this concept to our notion of a consistent forecasting rule by introducing *minimal-state-variable predictors* (MSV predictors) for expectational leads of arbitrary length. For simplicity, we assume throughout this section that  $n_1 = n_3 = 1$  for the lag length of both the endogenous and exogenous variables, we let  $n_2 > 0$  of arbitrary size, and consider the white-noise case  $n_4 = 0$  with an AR(1) process  $\{w_t\}_{t \in \mathbb{N}}$ . Consider a *perceived law of motion* of the form

$$(17) \quad \begin{cases} y_{t+1} = \hat{A}y_t + \hat{D}w_t + \hat{b} + \epsilon_{t+1}, \\ w_{t+1} = \Pi w_t + \eta_{t+1}, \end{cases}$$

on the basis of which a forecasting agency forms expectations.  $y_t$  and  $w_t$  denote vectors of endogenous and exogenous variables, respectively,  $\hat{A}$  and  $\hat{D}$  are non-random matrices of respective dimensions, and  $\hat{b}$  a non-random vector.  $\{\epsilon_t\}_{t \in \mathbb{N}}$  and  $\{\eta_t\}_{t \in \mathbb{N}}$  are martingale difference sequences on  $(\Omega, \mathcal{F}, \mathbb{P})$  adapted to the filtration  $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$  which are assumed to be unobservable. If  $(w_t, y_t)$  describes the state of the economy at date  $t$ , the perceived

state at date  $t + j$  derived from the model (17) is

$$(18) \quad \begin{cases} y_{t+j} = \hat{A}^j y_t + \sum_{i=1}^j \hat{A}^{j-i} [\hat{D} w_{t+i-1} + \hat{b} + \epsilon_{t+i}], \\ w_{t+j} = \Pi^j w_t + \sum_{i=1}^j \Pi^{j-i} \eta_{t+i}. \end{cases}$$

With the notation above,  $\mathbb{E}_t[\xi_{t+1}] = \hat{D}w_t + \hat{b}$ , such that  $\mathbb{E}_t[\xi_{t+1}]$  will be replaced by  $w_t$  in all subsequent expressions. Taking conditional expectations with respect to  $\mathcal{F}_t$  and using the fact that  $\mathbb{E}_t[w_{t+j}] = \Pi^j w_t$ , we obtain a forecasting rule  $\varphi = (\varphi^{(1)}, \dots, \varphi^{(n_2+1)})$  by setting

$$(19) \quad y_{t,t+j}^e = \varphi^{(j)}(w_t, y_t) := \hat{A}^j y_t + \sum_{i=1}^j \hat{A}^{j-i} [\hat{D} \Pi^{i-1} w_t + \hat{b}], \quad j = 1, \dots, n_2 + 1.$$

Since the forecasting rule  $\varphi$  defined in (19) uses a minimal amount of endogenous as well as exogenous variables, it will be referred to as a *minimal-state-variable* predictor (MSV predictor). By construction, one has  $y_{t,t+j}^e = \mathbb{E}_t[y_{t+j}]$ , if  $y_{t+j}$  is thought of being generated from (18), and a straightforward calculation shows that each forecast (19) satisfies

$$\mathbb{E}_{t-1}[y_{t,t+j}^e] = \mathbb{E}_{t-1}[\varphi^{(j)}(w_t, y_t)] = \varphi^{(j+1)}(w_{t-1}, y_{t-1}) = y_{t-1,t+j}^e, \quad j = 1, \dots, n_2 + 1.$$

This implies that a MSV predictor is a consistent forecasting rule in the sense of Definition 3.1. Moreover, for each  $j > 1$ ,

$$(20) \quad y_{t,t+j}^e = \varphi^{(j-1)}(\Pi w_t, \varphi^{(1)}(w_t, y_t)) = \varphi^{(1)}(\Pi^{j-1} w_t, \varphi^{(j-1)}(w_t, y_t)),$$

such that all forecasts can be obtained from a repeated application of the one-period-ahead forecasting rule  $\varphi^{(1)}$ .

In order to construct an MSV predictor  $\varphi = (\varphi^{(1)}, \dots, \varphi^{(n_2+1)})$  which is unbiased in the sense of Definition 3.2, for each  $w \in \mathbb{R}^d$ , let

$$\Gamma_\varphi(w) := \{(y, z^{(1)}, \dots, z^{(n_2+1)}) \in \mathbb{R}^{d(n_2+2)} \mid z^{(j)} = \varphi^{(j)}(w, y), 1 \leq j \leq n_2 + 1\}.$$

Then, replacing  $\mathbb{E}_t[\xi_{t+1}]$  by  $w_t$  as above, each  $\Gamma_\varphi(w)$  contains  $\mathbb{R}^d \times \{0\}$  as a subset and it is easy to see that the family of sets  $\{\Gamma_\varphi(w)\}_{w \in \mathbb{R}^d}$  is invariant under the corresponding time-one map  $F_\varphi$ . Hence  $\varphi$  satisfies Properties (ii) and (iii) of Definition 3.2. A MSV predictor  $\varphi = (\varphi^{(1)}, \dots, \varphi^{(n_2+1)})$  is therefore unbiased in the sense of Definition 3.2, if

$$(21) \quad e_F(w, y, \varphi^{(1)}(w, y), \dots, \varphi^{(n_2+1)}(w, y)) = 0 \quad \forall (w, y) \in \mathbb{R}^d \times \mathbb{R}^d,$$

where  $e_F$  denotes the error function associated with the economic law (1) (with  $n_1 = n_3 = 1$  and  $n_4 = 0$ ).

If  $\varphi_\star$  is an unbiased MSV predictor with coefficients  $\hat{A}_\star$ ,  $\hat{D}_\star$ , and  $\hat{b}_\star$ , it follows from (3.2) that

$$y_{t+1} = \varphi_\star^{(1)}(w_t, y_t) + \epsilon_{t+1} := \hat{A}_\star y_t + \hat{D}_\star w_{t+1} + \hat{b}_\star + \epsilon_{t+1}$$

for all times  $t$ . Thus, unbiased MSV predictors, if they exist, generate minimal-state-variable solutions to (1) in the sense of McCallum (1983) which correspond to the perceived law (17). Notice that all relevant forecasts are obtained from (20) such that all conditional forecast errors vanish along any trajectory of the resulting system and in particular for any initial values  $w_0 \in \mathbb{R}^d$ ,  $y_0 \in \mathbb{R}^d$ .

To compute an unbiased MSV predictor may become a difficult problem. Inserting (19) into (21), a MSV predictor  $\varphi$  is unbiased, if and only if the unknown coefficients  $\hat{A}$ ,  $\hat{D}$ , and  $\hat{b}$  in (19) satisfy

$$(22) \quad A^{(1)} + [B^{(n_2)} - I_d]\hat{A} + \sum_{i=0}^{n_2-1} B^{(n_2-i)}\hat{A}^{i+1} = 0,$$

$$(23) \quad \sum_{i=0}^{n_2-1} \sum_{j=1}^{i+1} B^{(n_2-i)}\hat{A}^{i+1-j}\hat{D}\Pi^{j-1} + [B^{(n_2)} - I]\hat{D} + D^{(1)} = 0,$$

$$(24) \quad \left( \sum_{i=0}^{n_2-1} \sum_{j=1}^{i+1} B^{(n_2-i)}\hat{A}^{i+1-j} + [B^{(n_2)} - I] \right) \hat{b} + b = 0.$$

Eqs. (22)-(24) correspond to Eqs. (10.5)-(10.7) in Evans & Honkapohja (2001, p. 230) for an arbitrary length of leads. It is shown in Lancaster (1969, Thm. 8.4.1, p. 262) that (23) can be transformed to a linear matrix equation. Thus (23) and (24) are linear equations for  $\hat{D}$  and  $\hat{b}$ , once  $\hat{A}$  is known. It is straightforward to verify that  $\hat{A}_\star$  is a solution to (22), if and only if the *linear factor*  $(\lambda I_d - \hat{A}_\star)$ ,  $\lambda \in \mathbb{R}$  is a right divisor of the matrix polynomial

$$L(\lambda) = A^{(1)} + [B^{(n_2)} - I_d]\lambda + \sum_{i=0}^{n_2-1} B^{(n_2-i)}\lambda^{i+1}, \quad \lambda \in \mathbb{R},$$

see Gohberg, Lancaster & Rodman (1982, p. 125). Solutions  $\hat{A}_\star$  with real coefficients need neither exist nor be uniquely determined, because multiple right divisors may, in general, exist. In fact, the full set of solutions may be obtained from a decomposition of  $L(\lambda)$  into linear factors, cf. Gohberg, Lancaster & Rodman (1982, p. 113).

However, explicit solutions for the case  $n_2 = 1$  are found in McCallum (1983, 1998, 1999). The existence of solutions to (22)-(24), in general, do not require an invertible  $B^{(0)}$ . However, it requires full knowledge of the coefficient matrix  $\Pi$ , which is not needed for a no-updating rule. Notice also that all eigenvalues of a solution  $\hat{A}_\star$  must have modulus less than unity, in order to obtain a stable process and thus an economic meaningful solution. There are simple examples for which unbiased MSV predictors do not exist (see McCallum 1983) whereas no-updating rules exist, cf. Böhm & Wenzelburger (2000a).<sup>5</sup> Despite this fact, unbiased MSV predictors are important as they may generate stable long-run behavior under rational expectations in an otherwise unstable situation.

---

<sup>5</sup>The deterministic case treated there carries over to the stochastic case. Take  $d = 1$  and a polynomial with complex roots only.

## 5 Adaptive learning of unbiased forecasting rules

In this section we introduce a learning scheme for unbiased forecasting rules for the general system (1) using methods from control theory, found in Lai & Wei (1986b). For simplicity of exposition, we assume that the constant term in (1) is  $b = 0$ .<sup>6</sup> Inserting (2), the linear law (1) takes the form

$$(25) \quad y_t = \sum_{i=1}^{n_1} A^{(i)} y_{t-i} + \sum_{i=0}^{n_2} B^{(n_2-i)} y_{t-1,t+i}^e + \sum_{i=1}^{n_3} C^{(i)} w_{t-i} + \epsilon_t + \sum_{i=1}^{n_4} D^{(i)} \epsilon_{t-i}.$$

Setting

$$\theta = (A^{(1)}, \dots, A^{(n_1)}, B^{(0)}, \dots, B^{(n_2)}, C^{(1)}, \dots, C^{(n_3)}, D^{(1)}, \dots, D^{(n_4)})$$

and

$$(26) \quad x_{t-1}^\top = (y_{t-1}^\top, \dots, y_{t-n_1}^\top, y_{t-1,t+n_2}^{e\top}, \dots, y_{t-1,t}^{e\top}, w_{t-1}^\top, \dots, w_{t-n_3}^\top, \epsilon_{t-1}^\top, \dots, \epsilon_{t-n_4}^\top),$$

the system (1) may be rewritten as

$$(27) \quad y_t = \theta x_{t-1} + \epsilon_t.$$

We will use the representation (27) to approximate the unbiased no-updating rule  $\psi$  given in (13) from estimates for the coefficients  $\theta$  in two steps. In the first step, we estimate the unknown coefficient matrix  $\theta$  from historical data. In the second step we compute the desired approximation of the forecasting rule. The estimation of  $\theta$  is done by means of the so-called *approximate-maximum-likelihood (AML)* algorithm, see e.g. Caines (1988) and Lai & Wei (1986b). The AML algorithm is a recursive scheme which successively generates estimates for  $\theta$  based on information available at date  $t$ . This algorithm is an extension of the ordinary-least-square (OLS) algorithm known from time-series analysis to a class of more general noise processes defined in Assumption 2.1. Let

$$(28) \quad \hat{\theta}_t = (\hat{A}_t^{(1)}, \dots, \hat{A}_t^{(n_1)}, \hat{B}_t^{(0)}, \dots, \hat{B}_t^{(n_2)}, \hat{C}_t^{(1)}, \dots, \hat{C}_t^{(n_3)}, \hat{D}_t^{(1)}, \dots, \hat{D}_t^{(n_4)})$$

denote the estimate for  $\theta$  based on information available at date  $t$ . Since the noise process  $\{\epsilon_t\}_{t \in \mathbb{N}}$  is assumed to be unobservable, we replace the regressor  $x_{t-1}$  in (27) by

$$(29) \quad \begin{aligned} \hat{x}_{t-1} &:= (y_{t-1}^\top, \dots, y_{t-n_1}^\top, y_{t-1,t+n_2}^{e\top}, \dots, y_{t-1,t}^{e\top}, w_{t-1}^\top, \dots, w_{t-n_3}^\top, \hat{\epsilon}_{t-1}^\top, \dots, \hat{\epsilon}_{t-n_4}^\top)^\top, \\ \hat{\epsilon}_{t-i} &:= y_{t-i} - \hat{\theta}_{t-i} \hat{x}_{t-i-1}, \quad i = 1, \dots, n_4, \end{aligned}$$

<sup>6</sup>This term could be treated as an  $d \times 1$  matrix together with an additional constant scalar input equal to 1. The generalizations of all technical theorems stated below (see Sec. 6) to additional inputs are straightforward.

where the  $\hat{\epsilon}_{t-i}$  are also called *a-posteriori* prediction errors. The AML algorithm<sup>7</sup> is recursively defined by

$$(30) \quad \begin{cases} \hat{\theta}_t &= \hat{\theta}_{t-1} + (y_t - \hat{\theta}_{t-1} \hat{x}_{t-1}) \hat{x}_{t-1}^\top P_{t-1}, \\ P_{t-1} &= P_{t-2} - P_{t-2} (\hat{x}_{t-1} \hat{x}_{t-1}^\top) P_{t-2} (1 + \hat{x}_{t-1}^\top P_{t-2} \hat{x}_{t-1})^{-1}, \end{cases}$$

where by the matrix inversion lemma  $P_t^{-1} = P_{t-1}^{-1} + \hat{x}_t \hat{x}_t^\top$ . The initial conditions  $P_0$  and  $\hat{\theta}_0$  may be chosen arbitrarily with invertible  $P_0$ .

The approximations  $\hat{A}_t^{(i)}$ ,  $\hat{B}_t^{(i)}$ ,  $\hat{C}_t^{(i)}$ , and  $\hat{D}_t^{(i)}$  may now be used to estimate an unbiased no-updating rule (14) as follows. Since  $\{\epsilon_t\}_{t \in \mathbb{N}}$  is unobservable,  $\sum_{i=1}^{n_4} \hat{D}_t^{(i)} \hat{\epsilon}_{t+1-i}$  is an estimate for  $\mathbb{E}[\epsilon_{t+1}] = \sum_{i=1}^{n_4} D_t^{(i)} \epsilon_{t+1-i}$ . Let  $B^\#$  denote the generalized inverse of a matrix  $B$ , also referred to as Moore-Penrose inverse, see Lancaster (1969, p. 303).  $B^\#$  coincides with the inverse  $B^{-1}$ , if  $B$  is invertible. Using (30), an approximation of the unbiased no-updating rule (14) is

$$(31) \quad \begin{aligned} y_{t,t+j}^e &= y_{t-1,t+j}^e, \quad j = 1, \dots, n_2, \\ y_{t,t+1+n_2}^e &= \hat{\psi}_t(\hat{x}_t) \\ &:= -\hat{B}_t^{(0)\#} \left[ \sum_{i=1}^{n_1} \hat{A}_t^{(i)} y_{t+1-i} + \sum_{i=0}^{n_2-1} \hat{B}_t^{(n_2-i)} y_{t,t+1+i}^e - y_{t,t+1}^e \right. \\ &\quad \left. + \sum_{i=1}^{n_3} \hat{C}_t^{(i)} w_{t+1-i} + \sum_{i=1}^{n_4} \hat{D}_t^{(i)} \hat{\epsilon}_{t+1-i} \right]. \end{aligned}$$

We will show next how the approximated no-updating rule (31) together with the AML algorithm (30) can be used to design an AML-based learning scheme.<sup>8</sup> Instead of (31), an approximation of any other unbiased forecasting rule, in particular of a MSV predictor, could be taken. However, since the computation of a MSV predictor is rather involved (the case  $n_2 = 1$  already requires several hidden invertibility conditions which need not be satisfied for approximations) and the focus of this paper is on global convergence of the AML algorithm, we take (31). Notice in this connection, that contrary to (31), MSV predictors require also full knowledge of the actual exogenous observable process. If the design matrix pertaining to the (recursive-least-squares) estimates of the corresponding coefficients does not satisfy a so-called persistent excitation condition, strongly consistent estimates for these coefficients may be unavailable (cf. Lai & Wei 1986a) which in turn may result in a failure of the learning scheme.

<sup>7</sup>This algorithm is also referred to as *recursive extended-least-squares (RELS)* algorithm, see Caines (1988, p. 540 and p. 557).

<sup>8</sup>If  $B^{(0)}$  is known to be invertible, then an approximation of the no-updating rule can directly be obtained from regressing over (14) instead of (25), thereby avoiding the somewhat involved computation of a Moore-Penrose inverse, see Wenzelburger (2001b).

## 6 Convergence of an AML-based learning scheme

We will now investigate under which conditions a learning scheme based on the AML algorithm (30) and the forecasting rule (31) converges to rational expectations globally for all initial conditions. Under a number of stability assumptions which will be stated below, global convergence obtains, if the coefficient estimates  $\hat{\theta}_t$  generated by the AML algorithm are consistent in the sense that  $\hat{\theta}_t$  converges a.s. to  $\theta$ , independently of initial states of the systems and independently of initial parameter estimates. The convergence of the AML algorithm has first been investigated by Ljung (1977) using the stability analysis of a limiting non-random ordinary differential equation and by Solo (1979) using a martingale approach. Ljung (1977) showed that under certain standard assumptions,  $\hat{\theta}_t$  converges to  $\theta$  a.s., if the estimated regressors  $\hat{x}_t$  given in (29) satisfy the *persistent excitation* property

$$(32) \quad \frac{1}{T} \sum_{t=1}^T \hat{x}_t \hat{x}_t^\top \text{ converges a.s. to a positive definite matrix.}$$

The main problem with property (32) is that the regressors  $\hat{x}_t$  cannot be manipulated directly. Moreover, it has long been recognized that in feedback systems such as (25), (32) may be violated such that the parameter estimates may fail to be consistent, see Lai & Wei (1986b, pp. 228-231) for an account of the literature. However, Lai & Wei (1986b) showed that the persistent excitation condition (32) can be translated to conditions on the inputs alone which in our case are the forecasts and the exogenous variables. Following Lai & Wei (1986b), we first provide a technical theorem which establishes strong consistency of the AML algorithm under much weaker assumptions than used by Ljung and Solo.

**Proposition 6.1** (*Lai & Wei 1986b, Thm. 4, p.241*)

*Consider the stochastic regression model*

$$y_t = A\eta_{t-1} + \epsilon_t + D^{(1)}\epsilon_{t-1} + \dots + D^{(n_4)}\epsilon_{t-n_4},$$

*where  $A$  is a  $d \times k$  non-random matrix and where  $D^{(1)}, \dots, D^{(n_4)}$  are non-random  $d \times d$  matrices and  $\{\epsilon_t\}_{t \in \mathbb{N}}$  is a stochastic process which satisfies Assumption 2.1. Let  $\eta_t$  be  $\mathcal{F}_t$  measurable and*

$$\theta = (A, D^{(1)}, \dots, D^{(n_4)}) \quad \text{and} \quad x_{t-1} = (\eta_{t-1}^\top, \epsilon_{t-1}^\top, \dots, \epsilon_{t-n_4}^\top)^\top.$$

*Consider the AML algorithm, given by*

$$\begin{cases} \hat{\theta}_t &= \hat{\theta}_{t-1} + (y_t - \hat{\theta}_{t-1}^\top \hat{x}_{t-1}) \hat{x}_{t-1}^\top P_{t-1}, \\ P_{t-1}^{-1} &= P_{t-2}^{-1} + (\hat{x}_{t-1} \hat{x}_{t-1}^\top)^\top, \end{cases}$$



where

$$\hat{x}_{t-1} = (\eta_{t-1}^\top, \hat{\epsilon}_{t-1}^\top, \dots, \hat{\epsilon}_{t-n_4}^\top)^\top \quad \text{and} \quad \hat{\epsilon}_t := y_t - \hat{\theta}_t^\top \hat{x}_{t-1}.$$

If the (weak) persistent excitation condition

$$(33) \quad \frac{\lambda_{\min} \left( \sum_{t=1}^T x_t x_t^\top \right)}{\log \left( e + \lambda_{\max} \left( \sum_{t=1}^T x_t x_t^\top \right) \right)} \rightarrow \infty \quad \mathbb{P} - a.s. \text{ as } T \rightarrow \infty,$$

holds, then

$$\hat{\theta}_t \rightarrow \theta \quad \mathbb{P} - a.s. \text{ as } t \rightarrow \infty,$$

where  $\lambda_{\min}(B)$  and  $\lambda_{\max}(B)$  denote the minimum and maximum eigenvalues of a symmetric matrix  $B$ , respectively.

**Corollary 6.2** Under the hypotheses of Proposition 6.1,

$$\|\hat{\theta}_t - \theta\|^2 = O \left( \frac{\log \lambda_{\max}(P_t^{-1})}{\lambda_{\min}(P_t^{-1})} \right) \quad \mathbb{P} - a.s.$$

The proof of Corollary 6.2 is found in Lai & Wei (1986a). It provides upper bounds for the convergence of the AML algorithm in terms of observable quantities. The persistent excitation condition (33) in Proposition 6.1 will now be replaced by a more direct assumption concerning the forecasts alone. Let  $A(s)$ ,  $B(s)$ ,  $C(s)$  and  $D(s)$ ,  $s \in \mathbb{C}$  be matrix polynomials of the form

$$(34) \quad \begin{aligned} A(s) &= I_d s^{n_1} - A^{(1)} s^{n_1-1} - \dots - A^{(n_1)}, \\ B(s) &= B^{(0)} s^{n_2} + B^{(1)} s^{n_2-1} + \dots + B^{(n_2)}, \\ C(s) &= C^{(1)} s^{n_3-1} + \dots + C^{(n_3)}, \\ D(s) &= I_d s^{n_4} + D^{(1)} s^{n_4-1} + \dots + D^{(n_4)}, \end{aligned}$$

where  $I_d$  denotes the  $d \times d$  identity matrix and  $A^{(i)}$ ,  $B^{(i)}$ ,  $C^{(i)}$ , and  $D^{(i)}$  are the  $d \times d$  matrices appearing in the representation (25). Letting  $\iota$  denote the unit shift operator (i.e.,  $\iota x_t = x_{t+1}$ ), the system (25) may be rewritten as

$$(35) \quad A(\iota) y_{t-n_1} = B(\iota) u_{t-n_2+1} + C(\iota) w_{t-n_3} + D(\iota) \epsilon_{t-n_4},$$

where we have set  $u_{t-1} = y_{t-1, t+n_2}^e$  for the forecasts.

The matrix polynomials  $A(s)$ ,  $B(s)$ ,  $C(s)$  and  $D(s)$  are said to be *left coprime*, if their greatest common left divisors are unimodular (i.e., with constant determinants  $\neq 0$ ), or equivalently, if there exist matrix polynomials  $K(s)$ ,  $L(s)$ ,  $M(s)$ , and  $N(s)$  such that  $A(s)K(s) + B(s)L(s) + C(s)M(s) + D(s)N(s) = I_d$ , cf. Kailath (1980, p. 399) or Hannan & Deistler (1988, Chap. 2) The left coprime assumption is often referred to as an *identifiability assumption* for an ARMAX system with general colored noise, as defined

in Assumption 2.1. It is automatically satisfied for the white-noise case  $C(s) = I_d$  ( $n_4 = 0$ ). Ljung, Söderström & Gustavsson (1975) provide counter-examples to the strong consistency of the AML algorithm, if the so-called positive real condition, i.e. Assumption 2.1 (iii), is violated.<sup>9</sup>

**Proposition 6.3** *Suppose that the matrix polynomials  $A(s)$ ,  $B(s)$ ,  $C(s)$  and  $D(s)$  defined in (34) are left coprime and that (25) holds for all  $t$ . Let  $\mathcal{G}_t := \sigma(\mathcal{F}_{t-1}, y_t, w_t)$  denote the  $\sigma$ -algebra generated by  $\mathcal{F}_{t-1}$ ,  $y_t$ , and  $w_t$ . Set  $\tilde{u}_t = u_t - \mathbb{E}[u_t | \mathcal{G}_t]$  with  $\mathcal{F}_t$ -measurable  $u_t$  and assume, in addition, that  $\mathbb{P}$ -a.s.:*

$$(36) \quad \sup_{t \geq 1} \mathbb{E}[\|\tilde{u}_t\|^\alpha | \mathcal{G}_t] < \infty \quad \text{for some } \alpha > 2,$$

$$(37) \quad \frac{\lambda_{\min}(\sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top)}{\log T + \|\tilde{u}_T\|^2} \rightarrow \infty \quad \text{as } T \rightarrow \infty,$$

$$(38) \quad \sum_{t=1}^T \|\mathbb{E}[u_t | \mathcal{G}_t]\|^2 = O\left(\lambda_{\min}\left(\sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top\right)\right).$$

Then the following properties hold true  $\mathbb{P}$ -a.s.:

$$(39) \quad (i) \sum_{t=1}^T \|u_t\|^2 = O(T), \quad (ii) \sum_{t=1}^T \|\tilde{u}_t\|^2 = O(T), \quad (iii) \liminf_{T \rightarrow \infty} \frac{\lambda_{\min}(\sum_{t=1}^T x_t x_t^\top)}{\lambda_{\min}(\sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top)} > 0.$$

The proof of Proposition 6.3 is a generalization of Lai & Wei (1986b, Thm. 5 (ii), p. 242) to a system with additional exogenous variables as inputs and given in Appendix A.1. The next theorem is an analogous generalization of Lai & Wei (1986b, Thm. 6 (ii)) and provides conditions under which strong consistency of the AML algorithm obtains.

**Theorem 6.4** *Let the following hypotheses be satisfied.*

- (i) *All eigenvalues of the matrix  $\mathcal{A}$  defined in (4) have modulus less or equal to unity.*
- (ii) *Assumptions 2.1 and 2.2 on the exogenous perturbations are fulfilled.*
- (iii) *Identifiability assumption. The matrix polynomials  $A(s)$ ,  $B(s)$ ,  $C(s)$  and  $D(s)$  given in (34) have no common left factor.*
- (iv) *The inputs  $u_t = y_{t,t+1+n_2}^e$  satisfy Assumptions (36)-(38) of Proposition 6.3.*

The AML-algorithm (30) generates strongly consistent estimates, i.e.,

$$\hat{\theta}_t \rightarrow \theta \quad \mathbb{P} - \text{a.s. as } t \rightarrow \infty.$$

---

<sup>9</sup>This condition seems to be standard in linear systems theory, see e.g. Caines (1988, p. 245) for details.

**Proof.** By Assumptions (iii) and (iv), the hypothesis of Proposition 6.3 are satisfied such that  $\sum_{t=1}^T \|u_t\|^2 = O(T)$ . In view of the assumptions, the prerequisites of Theorem 2.1 are then satisfied as well, implying that there exists  $\alpha > 0$  such that  $\|x_t\|^2 = O(t^\alpha)$   $\mathbb{P}$ -a.s. Hence

$$\log \lambda_{\max} \left( \sum_{t=1}^T x_t x_t^\top \right) = O(\log T) \quad \mathbb{P} - \text{a.s.}$$

By (37) and (39), it is seen that

$$\frac{\lambda_{\min} \left( \sum_{t=1}^T x_t x_t^\top \right)}{\log T} \rightarrow \infty \quad \mathbb{P} - \text{a.s. as } T \rightarrow \infty$$

and the assertion follows from Proposition 6.1. Q.E.D.

In order to arrive at the desired AML-based learning scheme which converges to rational expectations, we need to ensure that the forecasts generated by the approximation of the no-updating rule (31) satisfies Assumption (iv) of Theorem 6.4. According to Proposition 6.3, this conditions entails a boundedness condition of the forecasts which cannot be guaranteed beforehand. We therefore amend the approximation (31) as follows. Using the forecast  $\hat{\psi}_t(\hat{x}_t)$  as given in (31), define

$$(40) \quad \begin{aligned} y_{t,t+j}^e &= y_{t-1,t+j}^e, \quad j = 1, \dots, n_2, \\ y_{t,t+1+n_2}^e &= \begin{cases} \hat{\psi}_t(\hat{x}_t) + v_t & \text{if } \|\hat{\psi}_t(\hat{x}_t)\| \leq c_t, \\ c_t \|\hat{\psi}_t(\hat{x}_t)\|^{-1} \hat{\psi}_t(\hat{x}_t) + v_t & \text{otherwise,} \end{cases} \end{aligned}$$

where  $\{v_t\}_{t \in \mathbb{N}}$  represents an exogenous white-noise process (small in magnitude) satisfying certain assumptions given below and  $\{c_t\}_{t \in \mathbb{N}}$  is a sequence of positive numbers such that  $c_t \rightarrow \infty$  and  $\sum_{t=1}^T c_t^2 = O(T)$ . Such a sequence, for example, is given by  $c_t = \log t$ . The motivation for the forecasting rule (40) is to meet the requirements (36)-(38) for inputs imposed by Proposition 6.3 thereby controlling the growth of the forecasts. We call (40) an AML-based forecast for the realizations  $y_{t+1}, \dots, y_{t+1+n_2}$ . The series of forecasts generated by the AML algorithm (30) and the forecasting rule (40) define an AML-based adaptive learning scheme for the linear system (1).

**Theorem 6.5** *Let Assumptions (i)–(iii) of Theorem 6.4 be satisfied and the forecasts be generated by the AML-based learning scheme defined by (30) and (40). Assume, in addition, that the following hypotheses are fulfilled.*

- (i) *The system (1) under the unbiased no-updating rule (14) is stable, that is, all eigenvalues of the coefficient matrix associated with (15) have modulus less or equal to unity;*
- (ii) *The process  $\{v_t\}_{t \in \mathbb{N}}$  in (40) satisfies*
  - (a)  *$\mathbb{E}[v_t | \mathcal{G}_t] = 0$  for all  $t \in \mathbb{N}$  and  $\sup_t \mathbb{E}[\|v_t\|^\alpha | \mathcal{G}_t] < \infty$   $\mathbb{P}$ -a.s. for some  $\alpha > 2$ , where  $\mathcal{G}_t = \sigma(\mathcal{F}_{t-1}, y_t, w_t)$  as before;*

$$(b) \liminf_{t \rightarrow \infty} \lambda_{\min}(\mathbb{E}[v_t v_t^\top | \mathcal{G}_t]) > 0 \text{ a.s.}$$

Then the AML-based adaptive learning scheme generates strongly consistent estimates, i.e.,  $\hat{\theta}_t \rightarrow \theta$   $\mathbb{P}$ -a.s. as  $t \rightarrow \infty$ . and all trajectories of the AML-based learning scheme converge to trajectories with rational expectations.

**Proof.** Let  $u_t = y_{t,t+1+n_2}^e$  as before. Then  $\tilde{u}_t = u_t - \mathbb{E}[u_t | \mathcal{G}_t] = v_t$  by (40) and (36) is satisfied. As shown in Lai & Wei (1983), Assumption (ii) implies that  $\mathbb{P}$ -a.s.

$$(41) \quad (i) \sum_{t=1}^T \|v_t\|^2 = O(T), \quad (ii) \|v_T\|^2/T \rightarrow 0, \quad (iii) \liminf_{T \rightarrow \infty} T^{-1} \lambda_{\min} \left( \sum_{t=1}^T v_t v_t^\top \right) > 0.$$

Properties (ii) and (iii) in (41) imply Condition (37). Since  $\|\mathbb{E}[u_t | \mathcal{G}_t]\| = \|u_t - v_t\| \leq c_t$  by (40),  $\sum_{t=1}^T \|\mathbb{E}[u_t | \mathcal{G}_t]\|^2 = O(T)$  and using (iii) in (41), this yields Condition (38). Thus Assumption (iv) of Theorem 6.4 is satisfied and we can apply Theorem 6.4 to conclude that  $\hat{\theta}_t \rightarrow \theta$   $\mathbb{P}$ -a.s. Since  $c_t \rightarrow \infty$  and the dynamics under rational expectations is stable by Assumption (i), all trajectories of the AML-based learning scheme converge to trajectories with rational expectations. *Q.E.D.*

The concept of continually disturbed inputs, also referred to as *dither*, is originally due to Caines, see e.g. Lai & Wei (1986b, p. 247) and references therein. The idea is that the system under rational expectations should be stable implying that the forecasts must be bounded with probability one. If the forecasting rule (31) yields large values of  $y_{t,t+1+n_2}^e$ , we can interpret this as an indicator that the precision of the parameter estimates  $\hat{\theta}_t$  involved in (31) is insufficient. The perturbations  $v_t$  which can be taken of small magnitude are designed to satisfy the weak excitation condition (33) which assures convergence of the AML algorithm. As pointed out above, in feedback systems with inputs of the form (31), the weak excitation condition may well be violated. Note in this connection that Corollary 6.2 may serve as an indicator for the quality of the parameter estimates and that the no-updating rule could be replaced by an MSV predictor, whenever the estimates are sufficiently close to their true values.

The idea of perturbing the inputs is generalized in Lai & Wei (1986a, 1987) to the concept of *occasional excitation* of the inputs, where the dither is exerted only if the precision of parameter estimates appears to be insufficient. These techniques are readily available for the univariate version of our model, however, their presentation and extension to the multivariate case are beyond the scope of this paper.

## 7 Conclusions

We introduced a learning scheme based on the recursive-extended-least-squares algorithm which converges globally for all initial conditions under standard assumptions. Although the scheme can be used for linear models only, the analysis reveals that is advantageous to first estimate the whole system including the feedback of any forecast acting on the system itself and then compute an approximation of an unbiased forecasting rule which generates the desired rational expectations equilibria. The main lesson suggests a careful distinction between four separate issues. First, the existence of a desired forecasting rule; second, the dynamic stability of the system under this forecasting rule; third, the dynamic stability of the system under the applied learning scheme and, fourth, between the success of the learning scheme in terms of providing strongly consistent estimates of the system's parameters. As shown in Wenzelburger (2001a, b), this methodology carries over to the non-linear case and the lesson strongly suggest that the mathematical object to be estimated from time series data is an appropriately defined error function associated with the economic law. There is a considerable amount of hope that such a task could, at least in part, be accomplished by combining our approach with the nonlinear estimation methods provided in White, Gallant, Hornik, Stinchcombe & Woolridge (1992) and Chen & White (1998).

## A Appendix

### A.1 Proofs of main theorems

#### Proof of Theorem 2.1.

The proof of (i) is an adaptation of Lemma 4 in Lai & Wei (1982). Let

$$Y_t = (y_t^\top, \dots, y_{t-n_1+1}^\top)^\top \quad \text{and} \quad \Xi_t = \left( \left( \sum_{i=0}^{n_2} B^{(n_2-i)} y_{t-1,t+i}^e + b + \xi_t \right)^\top, 0, \dots, 0 \right)^\top.$$

Using (4), the system (1) and (2) may be rewritten as  $Y_t = \mathcal{A}Y_{t-1} + \Xi_t$  and thus

$$(42) \quad Y_T = \mathcal{A}^T Y_0 + \sum_{t=1}^T \mathcal{A}^{T-t} \Xi_t.$$

Let  $\mathcal{C}$  be nonsingular such that  $\mathcal{A} = \mathcal{C}\mathcal{J}\mathcal{C}^{-1}$ , where  $\mathcal{J}$  is the Jordan normal form of  $\mathcal{A}$ . Since all eigenvalues of  $\mathcal{A}$  lie inside the unit circle, it follows from Varga (1962, Thm. 3.1, p. 65) that

$$(43) \quad \|\mathcal{A}^T\| \leq \|\mathcal{C}\mathcal{J}^T\mathcal{C}^{-1}\| \leq \|\mathcal{C}\| \|\mathcal{C}^{-1}\| \|\mathcal{J}^T\| \leq O(T^{M-1}),$$

where  $M = \max_j m_j$  and  $m_j$  denotes the multiplicity of the  $j$ -th eigenvalue of  $\mathcal{A}$ . In view of Assumptions (ii) and (iii), we have  $\|\Xi_t\| = O(t^\gamma)$  for some suitable  $\gamma > 0$  and therefore by (42) and (43)

$$\|Y_T\| \leq \|\mathcal{A}^T\| \|Y_0\| + \sum_{t=1}^T \|\mathcal{A}^{T-t}\| \|\Xi_t\| \leq O(t^\gamma) \quad \mathbb{P} - \text{a.s.}$$

This proves (i). In order to prove (ii), note first that the Schwarz inequality implies

$$\begin{aligned} \sum_{t=1}^T \left( \sum_{s=1}^t \|\mathcal{A}^{t-s}\| \|\Xi_s\| \right)^2 &\leq \sum_{t=1}^T \left( \sum_{s=1}^t \|\mathcal{A}^{t-s}\| \right) \left( \sum_{s=1}^t \|\mathcal{A}^{t-s}\| \|\Xi_s\|^2 \right) \\ &\leq \left( \sum_{t=0}^{\infty} \|\mathcal{A}^t\| \right) \sum_{s=1}^T \left( \sum_{t=s}^T \|\mathcal{A}^{t-s}\| \right) \|\Xi_s\|^2 \\ &\leq \left( \sum_{t=0}^{\infty} \|\mathcal{A}^t\| \right)^2 \sum_{s=1}^T \|\Xi_s\|^2. \end{aligned}$$

In view of (42), it follows from Assumptions (i) – (iii) that

$$\sum_{t=1}^T \|Y_t\|^2 \leq 2 \left( \sum_{t=1}^T \|\mathcal{A}^t\|^2 \right) \|Y_0\|^2 + 2 \left( \sum_{t=0}^{\infty} \|\mathcal{A}^t\| \right)^2 \sum_{s=1}^T \|\Xi_s\|^2 = O(T) \quad \mathbb{P} - \text{a.s.}$$

This implies (ii). *Q.E.D.*

### Proof of Proposition 6.3.

Property (39, ii) follows from (36) as shown in Lai & Wei (1983). This, in turn, implies  $\lambda_{\min} \left( \sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top \right) = O(T)$  and (39, i) follows from (38).

The last property (39, iii) follows from Proposition A.1 by first showing that the regressors (26) can be reduced regressors (45) consisting only of inputs (see below) and then applying Proposition A.4, noting that  $\lambda_{\min} \left( \sum_{t=1}^T z_t z_t^\top \right) \geq \lambda_{\min} \left( \sum_{t=1}^T Z_t Z_t^\top \right)$  for any sub-vector  $z_T$  of  $Z_t$ , where the latter will be defined in (51). Note in this connection that interchanging any vector entries of a regressor  $z_t$  will not change the eigenvalues of the (design) matrix  $z_t z_t^\top$ , since it amounts to interchanging columns and rows simultaneously. *Q.E.D.*

**Proposition A.1** *Suppose that the matrix polynomials  $A(s)$ ,  $B(s)$ ,  $C(s)$ , and  $D(s)$  given (34) are left coprime and that (35) holds for all times  $t$ , that is,*

$$(44) \quad A(\iota) y_{t-n_1} = B(\iota) u_{t-n_2+1} + C(\iota) w_{t-n_3} + D(\iota) \epsilon_{t-n_4}.$$

Let

$$(45) \quad \begin{aligned} x_{t-1}^\top &= (y_{t-1}^\top, \dots, y_{t-n_1}^\top, u_{t-1}^\top, \dots, u_{t-n_2+1}^\top, w_{t-1}^\top, \dots, w_{t-n_3}^\top, \epsilon_{t-1}^\top, \dots, \epsilon_{t-n_4}^\top), \\ z_{t-1}^\top &= (u_{t-1+dn_1}^\top, \dots, u_{t-n_2+1}^\top, w_{t-1+dn_1}^\top, \dots, w_{t-n_3}^\top, \epsilon_{t-1+dn_1}^\top, \dots, \epsilon_{t-n_4}^\top). \end{aligned}$$

Then there exists  $\rho > 0$  such that for all  $T \geq dn_1$ ,

$$\lambda_{\min} \left( \sum_{t=1}^T x_t x_t^\top \right) \geq \rho \lambda_{\min} \left( \sum_{t=1}^T z_t z_t^\top \right).$$

The proof of Proposition A.1 is a generalization of Theorem 2 in Lai & Wei (1986b, p. 235) for an additional source of inputs and will be prefaced by two technical lemmatas. Let  $\tilde{A}(s)$  denote the adjoint of the matrix  $A(s)$  such that  $\tilde{A}(s)A(s) = a(s)I_d$  with  $a(s) = \det A(s)$ . We have  $a(s) = \sum_{j=0}^{dn_1} a_j s^j$  for suitable real coefficients  $a_j$  with  $a_0 = 1$ .

**Lemma A.2** *Let  $A(s), K_1(s), \dots, K_m(s)$  be left coprime matrix polynomials. Assume that the degree of the matrix polynomial  $A(s)$  is  $\deg A(s) = n_1$  with the identity matrix as leading coefficient and let  $\tilde{A}(s)$  denote the adjoint of  $A(s)$  with  $a(s) = \det A(s)$ . If there exist matrix polynomials  $N_1(s), \dots, N_m(s)$  and vectors  $\alpha_1, \dots, \alpha_{n_1} \in \mathbb{R}^d$  such that*

$$(46) \quad \sum_{i=1}^{n_1} \alpha_i^\top \tilde{A}(s) K_j(s) s^{i-1} = a(s) N_j(s) \quad \text{for all } j = 1, \dots, m,$$

then  $\alpha_1 = \dots = \alpha_{n_1}$ .

**Proof.** By the coprimeness, there exist matrix polynomials  $M_0(s), \dots, M_m(s)$  such that

$$A(s)M_0(s) + K_1(s)M_1(s) + \dots + K_m(s)M_m(s) = I_d.$$

Suppose that (46) holds. Then

$$(47) \quad \begin{aligned} \sum_{i=1}^{n_1} \alpha_i^\top \tilde{A}(s) s^{i-1} &= \sum_{i=1}^{n_1} \alpha_i^\top \tilde{A}(s) \left[ A(s)M_0(s) + \sum_{j=1}^m K_j(s)M_j(s) \right] s^{i-1} \\ &= a(s) \left[ \sum_{i=1}^{n_1} \alpha_i^\top M_0(s) + \sum_{j=1}^m N_j(s)M_j(s) \right]. \end{aligned}$$

Since the degree of  $\tilde{A}(s)$  is  $\deg \tilde{A}(s) = (d-1)n_1$  and the degree of  $a(s)$  is  $\deg a(s) = dn_1$ ,

$$\deg \left( \sum_{i=1}^{n_1} \alpha_i^\top \tilde{A}(s) s^{i-1} \right) \leq n_1 - 1 + \deg \tilde{A}(s) = dn_1 - 1 \leq \deg a(s).$$

It follows from (47) that  $(\sum_{i=1}^{n_1} \alpha_i s^{i-1})^\top \tilde{A}(s) = 0$  for all  $s$ . Since the leading coefficient of  $A(s)$  is the identity matrix,  $A(s)$  and thus  $\tilde{A}(s)$  has full rank for almost all  $s$ . Thus  $\sum_{i=1}^{n_1} \alpha_i s^{i-1} = 0$  for almost all  $s$  and therefore  $\alpha_1 = \dots = \alpha_{n_1} = 0$ . *Q.E.D.*

**Lemma A.3** Let  $A$  and be an  $d_1 \times d_2$  matrix with  $\ker A^\top = \{0\}$  and  $B$  be a symmetric  $d_1 \times d_1$  matrix. Then there exists  $\rho > 0$  with  $\lambda_{\min}(ABA^\top) \geq \rho \lambda_{\min}(B)$ .

**Proof.** By Theorem 3.2.1, Lancaster (1969, p. 109),  $\lambda_{\min}(C) = \min_{\|x\|=1} (x^\top C x)$  for any symmetric matrix  $C$ . Since  $AA^\top$  is nonsingular, for any  $0 \neq x \in \mathbb{R}^{d_2}$ ,

$$\frac{x^\top ABA^\top x}{x^\top x} = \frac{x^\top ABA^\top x}{x^\top AA^\top x} \cdot \frac{x^\top AA^\top x}{x^\top x} \geq \frac{x^\top AA^\top x}{x^\top x} \lambda_{\min}(B).$$

The lemma follows from another application of Theorem 3.2.1 with  $\rho = \lambda_{\min}(AA^\top) > 0$ .

*Q.E.D.*

### Proof of Proposition A.1.

Let  $\tilde{A}(s)$  denote the adjoint of the matrix  $A(s)$  as before. We have  $a(s) = \sum_{j=0}^{dn_1} a_j s^j$ , ( $a_0 = 1$ ) and thus

$$a(\iota)x_{t-1} = \sum_{j=0}^{dn_1} a_j x_{t-1+dn_1-j}.$$

Since  $\tilde{A}(\iota)A(\iota)x_{t-1} = a(\iota)x_{t-1}$ , (44) implies

$$(48) \quad \begin{aligned} a(\iota)y_{t-j} &= \tilde{A}(\iota) [B(\iota)\iota^{n_1-j}u_{t-n_2+1} + D(\iota)\iota^{n_1-j}w_{t-n_3} + C(\iota)\iota^{n_1-j}\epsilon_{t-n_4}], \\ a(\iota)u_{t-j} &= a(\iota)\iota^{n_2+1-j}u_{t-n_2+1}, \\ a(\iota)w_{t-j} &= a(\iota)\iota^{n_4-j}w_{t-n_3}, \\ a(\iota)\epsilon_{t-j} &= a(\iota)\iota^{n_4-j}\epsilon_{t-n_4}. \end{aligned}$$

Now set  $X_t = (X_t^{(1)\top}, \dots, X_t^{(4)\top})^\top = a(\iota)x_{t-1}$  with

$$\begin{aligned} X_t^{(1)} &= ((a(\iota)y_{t-1})^\top, \dots, (a(\iota)y_{t-n_1})^\top)^\top, & X_t^{(2)} &= ((a(\iota)u_{t-1})^\top, \dots, (a(\iota)u_{t-n_2+1})^\top)^\top, \\ X_t^{(3)} &= ((a(\iota)w_{t-1})^\top, \dots, (a(\iota)w_{t-n_3})^\top)^\top, & X_t^{(4)} &= ((a(\iota)\epsilon_{t-1})^\top, \dots, (a(\iota)\epsilon_{t-n_4})^\top)^\top \end{aligned}$$

and

$$\begin{aligned} U_t &= (u_{t-1+dn_1}^\top, \dots, u_{t-n_2+1}^\top)^\top, & W_t &= (w_{t-1+dn_1}^\top, \dots, w_{t-n_3}^\top)^\top, \\ E_t &= (\epsilon_{t-1+dn_1}^\top, \dots, \epsilon_{t-n_4}^\top)^\top. \end{aligned}$$

Then (48) has a block-matrix representation as

$$(49) \quad \begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \\ X_t^{(3)} \\ X_t^{(4)} \end{pmatrix} = \mathcal{D} \begin{pmatrix} U_t \\ W_t \\ E_t \end{pmatrix} \quad \text{where} \quad \mathcal{D} = \begin{pmatrix} \mathcal{D}_1 & \mathcal{D}_2 & \mathcal{D}_3 \\ \mathcal{D}_4 & 0 & 0 \\ 0 & \mathcal{D}_5 & 0 \\ 0 & 0 & \mathcal{D}_6 \end{pmatrix}$$



The block matrix  $\mathcal{D}$  in (49) is a  $d(n_1 + 1 + n_2 + n_3 + n_4) \times d(3dn_1 + 1 + n_2 + n_3 + n_4)$  matrix with constant coefficients. If the rank of  $\mathcal{D}$  were less than  $d(n_1 + 1 + n_2 + n_3 + n_4)$ , then there exist vectors  $\alpha_1, \dots, \alpha_{n_1}, \beta_1, \dots, \beta_{n_2+1}, \gamma_1, \dots, \gamma_{n_3}$ , and  $\delta_1, \dots, \delta_{n_4}$  in  $\mathbb{R}^d$  such that

$$\begin{aligned} \sum_{i=1}^{n_1} \alpha_i^\top \tilde{A}(s)B(s)s^{i-1} &= -a(s) \sum_{j=0}^{n_2} \beta_j^\top I_d s^j, \\ \sum_{i=1}^{n_1} \alpha_i^\top \tilde{A}(s)C(s)s^{i-1} &= -a(s) \sum_{j=0}^{n_3-1} \gamma_j^\top I_d s^j, \\ \sum_{i=1}^{n_1} \alpha_i^\top \tilde{A}(s)D(s)s^{i-1} &= -a(s) \sum_{j=0}^{n_4-1} \delta_j^\top I_d s^j, \end{aligned}$$

which is impossible by Lemma A.2. Hence the matrix in (49) has full rank. Noting that  $z_{t-1} = (U_t^\top, W_t^\top, E_t^\top)^\top$ , we can apply Lemma A.3 to see that there exists  $\rho > 0$  such that

$$\lambda_{\min} \left( \sum_{t=1}^{T-dn_1} X_t X_t^\top \right) \geq \rho \lambda_{\min} \left( \sum_{t=1}^{T-dn_1} z_t z_t^\top \right).$$

To complete the proof, it suffices to show that

$$(50) \quad \lambda_{\min} \left( \sum_{t=1}^{T-dn_1} X_t X_t^\top \right) \leq (dn_1 + 1) \left( \sum_{j=0}^{dn_1} a_j^2 \right) \lambda_{\min} \left( \sum_{t=1}^T x_t x_t^\top \right),$$

Let  $\zeta$  be a unit vector. Then

$$\zeta^\top X_t X_t^\top \zeta = \left( \sum_{j=0}^{dn_1} a_j \zeta^\top x_{t+dn_1-j} \right)^2 \leq \left( \sum_{j=0}^{dn_1} a_j^2 \right) \sum_{j=0}^{dn_1} (\zeta^\top x_{t+dn_1-j})^2$$

implies

$$\sum_{t=1}^{T-dn_1} \zeta^\top X_t X_t^\top \zeta \leq (dn_1 + 1) \left( \sum_{j=0}^{dn_1} a_j^2 \right) \sum_{t=1}^T \zeta^\top x_t x_t^\top \zeta,$$

proving (50). Q.E.D.

**Proposition A.4** *Under the hypotheses of Propositions 6.3 and A.1, for arbitrary  $n > 0$ , let*

$$(51) \quad Z_t = (\epsilon_{t+n}^\top; u_{t+n-1}^\top, w_{t+n-1}^\top, \epsilon_{t+n-1}^\top; \dots; u_t^\top, w_t^\top, \epsilon_t^\top)^\top.$$

Then

$$\liminf_{T \rightarrow \infty} \frac{\lambda_{\min}(\sum_{t=1}^T Z_t Z_t^\top)}{\lambda_{\min}(\sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top)} > 0 \quad \mathbb{P} - \text{a.s.}$$

**Proof.** The proof consists of an induction argument on  $n$  and of a repeated application of Theorem A.5 stated below. To apply Theorem A.5, let  $n = 1$  and proceed in three steps. *Step 1.* Consider the random vector  $\zeta_t^{(1)} = (w_t^\top, \epsilon_t^\top)^\top$ . Set  $\mathcal{H}_{t-1} = \sigma(\mathcal{F}_{t-1}, \epsilon_t)$ ,  $\tau_t = \epsilon_t$ ,  $\xi_t = \mathbb{E}[w_t | \mathcal{H}_{t-1}]$ , and  $\eta_t = w_t - \mathbb{E}[w_t | \mathcal{H}_{t-1}]$ . By Assumptions 2.1 and 2.2, all requirements of Corollary A.6 are fulfilled. In particular, since  $\lambda_{\min} \left( \sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top \right) = O(T)$ ,

$$(52) \quad \liminf_{T \rightarrow \infty} \frac{\lambda_{\min} \left( \sum_{t=1}^T \zeta_t^{(1)} \zeta_t^{(1)\top} \right)}{\lambda_{\min} \left( \sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top \right)} > 0 \quad \mathbb{P} - \text{a.s.}$$

*Step 2.* Consider a second random vector  $\zeta_t^{(2)} = (\zeta_t^{(1)\top}, u_t^\top)^\top$ . Set  $\tau_t = \zeta_t^{(1)}$ ,  $\xi_t = \mathbb{E}[u_t | \mathcal{G}_t]$ ,  $\eta_t = \tilde{u}_t$ , and  $\mathcal{H}_t = \mathcal{G}_{t-1}$ . By Step 1,  $\log \sum_{t=0}^T \|\tau_t\|^2 = O(\log T)$ . In view of (36)-(38),  $\zeta_t^{(2)}$  again meets all requirements of Corollary A.6. Hence, we conclude from  $\lambda_{\min} \left( \sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top \right) = O(T)$  that

$$(53) \quad \liminf_{T \rightarrow \infty} \frac{\lambda_{\min} \left( \sum_{t=1}^T \zeta_t^{(2)} \zeta_t^{(2)\top} \right)}{\lambda_{\min} \left( \sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top \right)} > 0 \quad \mathbb{P} - \text{a.s.}$$

*Step 3.* Consider a third random vector  $\zeta_t^{(3)} = (\zeta_t^{(2)\top}, \epsilon_{t+1}^\top)^\top$ , where  $\tau_t = \zeta_t^{(2)}$ ,  $\xi_t = 0$ ,  $\eta_t = \epsilon_{t+1}$ , and  $\mathcal{H}_t = \mathcal{F}_{t+1}$ . Since again  $\log \sum_{t=0}^T \|\tau_t\|^2 = O(\log T)$ , it follows from Assumption 2.1 (i) and (ii) that  $\zeta_t^{(3)}$  satisfies all hypotheses of Corollary A.6. From an analogous argument used in Step 2, we obtain

$$(54) \quad \liminf_{T \rightarrow \infty} \frac{\lambda_{\min} \left( \sum_{t=1}^T \zeta_t^{(3)} \zeta_t^{(3)\top} \right)}{\lambda_{\min} \left( \sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top \right)} > 0 \quad \mathbb{P} - \text{a.s.}$$

This proves the case  $n = 1$ . By (37), we have<sup>10</sup>  $\lambda_{\min} \left( \sum_{t=1}^T \tilde{u}_t \tilde{u}_t^\top \right) \sim \lambda_{\min} \left( \sum_{t=1}^{T-m} \tilde{u}_t \tilde{u}_t^\top \right)$  for any fixed  $m$  and a repeated application of the above three steps establishes the desired conclusion for arbitrary  $n$ . *Q.E.D.*

**Theorem A.5** (*Lai & Wei 1986b, Thm. 3, p.235*)

Let  $\zeta_t = \begin{pmatrix} \xi_t + \eta_t \\ \tau_t \end{pmatrix}$  be a random vector and  $\{\mathcal{H}_t\}_{t \in \mathbb{N}}$  be an increasing sequence of  $\sigma$ -algebras. Suppose that the following properties hold with probability 1.

(1)  $\{\eta_t\}_{t \in \mathbb{N}}$  is an  $\mathbb{R}^{d_2}$ -valued martingale difference sequence with respect to  $\{\mathcal{H}_t\}_{t \in \mathbb{N}}$  such that

(a)  $\sup_{t \in \mathbb{N}} \mathbb{E}[\|\eta_t\|^\alpha | \mathcal{H}_t] < \infty$  for some  $\alpha > 2$ , (b)  $\lambda_{\min} \left( \sum_{t=1}^T \eta_t \eta_t^\top \right) \rightarrow \infty$  as  $T \rightarrow \infty$ .

---

<sup>10</sup> By  $a_t \sim b_t$   $\mathbb{P}$ -a.s. for two sequences of random numbers  $(a_t)$  and  $(b_t)$ , we mean that  $\lim_{t \rightarrow \infty} \frac{a_t}{b_t} = 1$   $\mathbb{P}$ -a.s. By  $a_t = o(b_t)$   $\mathbb{P}$ -a.s., we mean that  $\lim_{t \rightarrow \infty} \left| \frac{a_t}{b_t} \right| = 0$   $\mathbb{P}$ -a.s.

(2)  $\xi_t \in \mathbb{R}^{d_2}$  is  $\mathcal{H}_{t-1}$  measurable with

$$\sum_{t=1}^T \|\xi_t\|^2 = O\left(\lambda_{\min}\left(\sum_{t=1}^T \eta_t \eta_t^\top\right) + \lambda_{\min}\left(\sum_{t=1}^T \tau_t \tau_t^\top\right)\right).$$

(3)  $\tau_t \in \mathbb{R}^{d_1}$  is  $\mathcal{H}_{t-1}$  measurable with<sup>10</sup>

$$(a) \log \sum_{t=1}^T \|\tau_t\|^2 = o\left(\lambda_{\min}\left(\sum_{t=1}^T \eta_t \eta_t^\top\right)\right), \quad (b) \lambda_{\min}\left(\sum_{t=1}^T \tau_t \tau_t^\top\right) \rightarrow \infty \text{ as } T \rightarrow \infty.$$

Then

$$\liminf_{T \rightarrow \infty} \frac{\lambda_{\min}\left(\sum_{t=1}^T \zeta_t \zeta_t^\top\right)}{\min\{\lambda_{\min}\left(\sum_{t=1}^T \tau_t \tau_t^\top\right), \lambda_{\min}\left(\sum_{t=1}^T \eta_t \eta_t^\top\right)\}} > 0 \quad \text{a.s.}$$

**Corollary A.6** Under the hypotheses of Theorem A.5, if, in addition,

$$(55) \quad \liminf_{T \rightarrow \infty} \frac{\lambda_{\min}\left(\sum_{t=1}^T \tau_t \tau_t^\top\right)}{\lambda_{\min}\left(\sum_{t=1}^T \eta_t \eta_t^\top\right)} > 0 \quad \text{a.s.},$$

then

$$\liminf_{T \rightarrow \infty} \frac{\lambda_{\min}\left(\sum_{t=1}^T \zeta_t \zeta_t^\top\right)}{\lambda_{\min}\left(\sum_{t=1}^T \eta_t \eta_t^\top\right)} > 0 \quad \text{a.s.}$$

## A.2 Example with singular $B^{(0)}$

We present a particular easy example, indicating to what extent an unbiased forecasting rule will exist, if the invertibility condition of the coefficient matrix  $B^{(0)}$  in Sec. 3 is violated. This example provides a complementary approach to the one found in McCallum (1998) and may easily be extended to the general case. With the notation of Sec. 2, consider the simple economic law

$$(56) \quad y_t = B y_{t-1, t+1}^e + \xi_t,$$

where  $B \equiv B^{(0)}$  is a possibly singular  $d \times d$  matrix. A similar example to (56) is found in McCallum (1998), although with different dating. Assume that there exists a coordinate transformation  $z = Qy$  with a nonsingular  $d \times d$  matrix  $Q$  such that

$$\begin{pmatrix} 0 & 0 \\ \tilde{B}^{(21)} & \tilde{B}^{(22)} \end{pmatrix} = Q^{-1} B Q$$

with a nonsingular  $d_1 \times d_1$  matrix  $B^{(11)}$  and a  $d_1 \times d_2$  matrix  $B^{(12)}$ , where  $d_1 + d_2 = d$ . Setting  $z_t = Qy_t$ ,  $z_{t-1, t+1}^e = Qy_{t-1, t+1}^e$ , and  $\eta_t = Q\xi_t$ , (56) takes the form

$$(57) \quad \begin{pmatrix} z_t^{(1)} \\ z_t^{(2)} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \tilde{B}^{(21)} & \tilde{B}^{(22)} \end{pmatrix} \begin{pmatrix} z_{t-1, t+1}^{(e1)} \\ z_{t-1, t+1}^{(e2)} \end{pmatrix} + \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix}.$$

Thus (56) decouples into two (vector) equations (57), the first of which receives no feedback of the forecasts. Since  $z_t^{(1)} = \eta_t^{(1)}$  and  $\tilde{B}^{(21)}$  is invertible, an unbiased forecasting rule for (57) is

$$(58) \quad \begin{pmatrix} z_{t-1,t+1}^{(e1)} \\ z_{t-1,t+1}^{(e2)} \end{pmatrix} = \begin{pmatrix} I_{d_1} & 0 \\ -\tilde{B}^{(22)-1} \tilde{B}^{(21)} & \tilde{B}^{(22)-1} \end{pmatrix} \begin{pmatrix} \mathbb{E}_{t-1}[\eta_{t+1}^{(1)}] \\ z_{t-2,t}^{(e2)} \end{pmatrix} - \begin{pmatrix} 0 \\ B^{(22)-1} \mathbb{E}_{t-1}[\eta_t^{(2)}] \end{pmatrix}.$$

Setting

$$(59) \quad y_{t-1,t+1}^e = Q^{-1} z_{t-1,t+1}^e$$

with  $z_{t-1,t+1}^e$  given by (58), we obtain an unbiased forecasting rule in the sense of Definition 3.2. Notice that (59) is a no-updating rule for the second component, that is,

$$\mathbb{E}_{t-1}[y_t^{(2)} - y_{t-2,t}^{(2e)}] = 0,$$

if  $\mathbb{E}_{t-1}[\eta_t^{(1)} - \eta_{t+1}^{(1)}] = 0$ .

*Note: See [www.wiwi.uni-bielefeld.de/~boehm](http://www.wiwi.uni-bielefeld.de/~boehm) to download the author's discussion papers.*

## References

- ARNOLD, L. (1998): *Random Dynamical Systems*. Springer-Verlag, Berlin a.o.
- ÅSTRÖM, K. J. & B. WITTENMARK (1973): "On Self-Tuning Regulators", *Automatica*, 9, 195–199.
- BÖHM, V. & C. CHIARELLA (2000): "Mean Variance Preferences, Expectations Formations, and the Dynamics of Random Asset Prices", Discussion Paper No. 448, University of Bielefeld.
- BÖHM, V. & J. WENZELBURGER (1999): "Expectations, Forecasting and Perfect Foresight - A Dynamical Systems Approach", *Macroeconomic Dynamics*, 3(2), 167–186.
- (2000a): "Expectational Leads in Economic Dynamical Systems", Revised Version September 2001, Discussion Paper No. 373, University of Bielefeld, forthcoming.
- (2000b): "Perfect Predictions in Economic Dynamical Systems with Random Perturbations", Revised Version September 2000, Discussion Paper No. 356, University of Bielefeld, forthcoming *Macroeconomic Dynamics*.
- BRAY, M. (1982): "Learning, Estimation, and the Stability of Rational Expectations Equilibria", *Journal of Economic Theory*, 26, 318–339.

- BRAY, M. & N. E. SAVIN (1986): “Rational Expectations Equilibria, Learning, and Model Specification”, *Econometrica*, 54, 1129–1160.
- CAINES, P. E. (1988): *Linear Stochastic Systems*. Wiley, New York a.o.
- CHATTERJI, S. & S. CHATTOPADHYAY (2000): “Global Stability in Spite of “Local Instability” with Learning”, *Journal of Mathematical Economics*, 33, 155–165.
- CHEN, X. & H. WHITE (1998): “Nonparametric Adaptive Learning with Feedback”, *Journal of Economic Theory*, 82(1), 199–222.
- EVANS, G. W. & S. HONKAPOHJA (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press.
- GOHBERG, I., P. LANCASTER & L. RODMAN (1982): *Matrix Polynomials*. Academic Press, New York.
- HANNAN, E. J. & M. DEISTLER (1988): *The Statistical Theory of Linear Systems*, Wiley Series in Probability and Mathematical Statistics. John Wiley, New York a.o.
- KAILATH, T. (1980): *Linear Systems*. Prentice-Hall, Inc., Englewood Cliffs, N. J.
- KUAN, C.-M. & H. WHITE (1994): “Adaptive Learning with Nonlinear Dynamics Driven By Dependent Processes”, *Econometrica*, 62(5), 1087–1114.
- LAI, T. L. & C. Z. WEI (1982): “Asymptotic Properties of Projections with Applications to Stochastic Regression Problems”, *Journal Of Multivariate Analysis*, 12, 346–370.
- (1983): “A Note on Martingale Difference Sequences Satisfying the Local Marcinkiewicz-Zygmund Condition”, *Bulletin of the Institute of Mathematics Academia Sinica*, 11, 1–13.
- LAI, T. L. & C.-Z. WEI (1986a): “Extended Least Squares and Their Applications to Adaptive Control and Prediction Linear Systems”, *IEEE Transactions on Automatic Control*, AC-31, 898–905.
- LAI, T. L. & C. Z. WEI (1986b): “On the Concept of Excitation in Least Squares Identification and Adaptive Control”, *Stochastics*, 16, 227–254.
- (1987): “Asymptotically Efficient Self-Tuning Regulators”, *SIAM Journal of Control and Optimization*, 25(2), 466–481.
- LANCASTER, P. (1969): *Theory of Matrices*. Academic Press, New York.
- LJUNG, L. (1977): “On Positive Real Transfer Functions and the Convergence of Some Recursive Schemes”, *IEEE Transactions on Automatic Control*, AC-22, 539–551.

- LJUNG, L., T. SÖDERSTRÖM & I. GUSTAVSSON (1975): “Counterexamples to a General Convergence of a Commonly Used Recursive Identification Method”, *IEEE Transactions on Automatic Control*, AC-20, 643–652.
- MCCALLUM, B. T. (1983): “On Non-Uniqueness in Rational Expectations Models – An Attempt at Perspective”, *Journal of Monetary Economics*, 11, 139–168.
- (1998): “Solutions to Linear Rational Expectations Models: A Compact Exposition”, *Economics Letters*, 61, 143–147.
- (1999): “Role of Minimal State Variable Criterion in Rational Expectations Models”, *International Tax and Public Finance*, 6, 621–639.
- SOLO, V. (1979): “The Convergence of AML”, *IEEE Transactions on Automatic Control*, AC-24, 958–962.
- VARGA, R. S. (1962): *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs.
- WENZELBURGER, J. (2001a): “Global Convergence of Adaptive Learning in Models of Pure Exchange”, *Economic Theory*, forthcoming.
- (2001b): “Learning to Predict Rationally when Beliefs are Heterogeneous”, Discussion Paper No. 477, University of Bielefeld.
- WHITE, H., A. GALLANT, K. HORNIK, M. STINCHCOMBE & J. WOOLRIDGE (1992): *Artificial Neural Networks: Approximation and Learning Theory*. Blackwell, Cambridge (Mass.) a.o.