555 LECTURE NOTES IN ECONOMICS AND MATHEMATICAL SYSTEMS

Jan Wenzelburger

Learning in Economic Systems with Expectations Feedback

Springer

Lecture Notes in Economics and Mathematical Systems 555

Jan Wenzelburger

Learning in Economic Systems with Expectations Feedback

With 4 Figures

Springer

Jan Wenzelburger
Bielefeld University
Department of Business Administration
and Economics
Postfach 100 131
33501 Bielefeld
Germany
jwenzelburger@wiwi.uni-bielefeld.de

ISBN-10 3-540-24322-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-24322-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springer.com

Satz: Camera ready by author
Cover: Erich Kirchner, Heidelberg
Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig

SPIN 11375647 Printed on acid-free paper – 88/3100YL – 5 4 3 2 1 0

Für Frauke

Preface

Recently more and more economists have abandoned the assumption that agents have perfect knowledge about the economic environment in which they live. The focus has been turned to scenarios in which agents' views of the world may be erroneous. The reason for this development is that the task of specifying all necessary capabilities which would allow an agent to attain the required perfect knowledge from observational data has not yet been accomplished. This fact holds particularly true for non-linear models in which forecasts feed back into the evolution of endogenous variables.

These notes introduce the concept of a perfect forecasting rule which provides best least-squares predictions during the evolution of an economic system and thus generates rational expectation equilibria. The framework for adaptive learning schemes which under suitable conditions converge to perfect forecasting rules is developed. It is argued that plausible learning schemes should aim at estimating a perfect forecasting rule taking into account the correct feedback structure of an economy. A link is provided between the traditional rational-expectations view and recent behavioristic approaches.

The material of these notes covers a revised version of my *Habilitationsschrift* which was accepted by the Fakultät für Wirtschaftswissenschaften at Bielefeld University. My primary scientific indebtedness is to Volker Böhm whose profound concern with dynamic models in economics inspired me to write these notes. I am grateful for the fruitful and creative atmosphere at the Lehrstuhl Böhm which all members created throughout the past years. Marten Hillebrand, Jochen Jungeilges, Thorsten Pampel, Olaf Schmitz, and George Vachadze helped to improve the manuscript at various stages and provided numerous helpful and critical comments. I am indebted to Hans Gersbach and Ulrich Horst for pointing out to me important new aspects of the theory.

Bielefeld, July 2006 *Jan Wenzelburger*

Contents

1

Introduction

The interplay between realizations of economic variables and agents' beliefs concerning their future development is a fundamental feature of evolving markets. In many economic situations, agents' subjective assessments of the future have a significant impact on the evolution of an economy. Adaptive learning schemes are a widely accepted concept to describe the way in which agents form and update beliefs on the basis of past observations. The original intention of introducing such learning schemes was to justify the rational expectations hypothesis as a long-run concept in which agents successfully eliminated all systematic forecast errors, cf. Blume & Easley (1982). This justification was first supported by Bray (1982), Bray & Savin (1986), Fourgeaud, Gourieroux & Pradel (1986), Marcet & Sargent (1989a), and others, who provide conditions for linear models under which adaptive learning schemes find rational expectations equilibria. These techniques are based on the ordinary least-squares (OLS) method. The current state of these techniques including their convergence properties are presented in the book by Evans & Honkapohja (2001).

While many economists use OLS-based learning schemes when modelling learning agents, there are four open issues which have not been resolved yet. First, most of the convergence results are only of a local nature, so that initial guesses of the system's parameters may have to be quite close to their true values in order to assure convergence. The available criteria that assure global convergence are often difficult or even impossible to verify. Second, in view of policy applications an open issue is how to select between multiple rational-expectations solutions criteria during the estimation process using economic criteria. Third, a systematic treatment of models which involve forecasts for more than two periods ahead of the current state has not yet been accomplished. With regard to overlapping generation models with multiperiod life cycles, this is unsatisfactory.

Finally, and most importantly, OLS-based techniques are restricted to linear and linearized models. Indeed, in economies whose evolution is driven by non-linear maps, the problem of learning schemes that generate forecasts converging to rational expectations equilibria remains to a large extent unresolved. For the standard OLG model of pure exchange, for example, Bullard (1994) and Marcet & Sargent (1989b) showed that ordinary least-squares learning may generate forecasts which do not converge to perfect foresight equilibria of the system. Convergence is obtained only if initial parameter guesses are sufficiently close to their true values. Similar observations have been made for adaptive schemes such as Bayesian learning, where agents update beliefs according to Bayes's rule (cf. Blume & Easley (1993, 1998) and references therein), and schemes based on parametric estimation techniques as in Kuan & White (1994).

The fact that adaptive learning schemes may often fail to localize the desired equilibria is one reason why many economists have abandoned the classical rational expectations hypothesis and replaced it by a behavioral assumption on how agents update beliefs. The focus has been directed towards the long-run outcomes of learning behavior, sometimes referred to as *learning equilibria*. The economic justification for this *'agent-based'* behavioral approach is that agents with limited statistical tools may be unable to identify systematic forecast errors and hence see no reason to revise their behavior. For the example in Bullard (1994), Schönhofer (1999, 2001) demonstrates that the long-run outcome of the system may be described by *consistent expectations equilibria* in the sense of Hommes & Sorger (1998) in which the associated forecast errors have vanishing sample means and confirm the hypothesized autocorrelation structure.

The plausibility of Schönhofer's results is questioned by Jungeilges (2003) who shows that the underlying linearity hypothesis for applying an OLS-based learning scheme in a non-linear environment may be rejected within finite time already by means of techniques from linear statistics. Hence, applying OLS-based learning schemes in non-linear models may be inappropriate. So far only a few contributions, as Chen & White (1998), Chatterji & Chattopadhyay (2000), or Wenzelburger (2002) succeeded in providing global results for non-linear models supporting the original intention of the learning literature. In contrast to Bullard's findings, Wenzelburger (2002) shows how a boundedly rational forecasting agency in a stationary exchange economy may always find the perfect foresight equilibria of the economy. However, all these convergence results are confined to a rather narrow class of non-linear 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.

These notes are motivated by the four open issues mentioned at the beginning of this introduction. We seek a method for non-linear models for which convergence results can be obtained and which allows to treat multiple rational-expectations solutions in a systematic manner. The focus is on multivariate models with expectational leads of arbitrarily finite length. These 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. They occur naturally in overlapping generations with multi-period life cycles, as in Auerbach & Kotlikoff (1987) or Bullard & Duffy (2001), Kübler & Polemarchakis (2004) or Geanakopolos, Magill & Quinzii (2004) as well as in in financial markets models with long-lived assets, e.g., see Hillebrand (2003) or Hillebrand & Wenzelburger (2006a). In macroeconomics, expectational leads 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 of an economy.

Following previous work by Böhm & Wenzelburger (1999, 2002, 2004), 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 together with a model for the exogenous perturbations may be seen as a *deterministic dynamical system* in a *random environment* which is defined explicitly and globally on the whole state space. The concept of an unbiased forecasting rule is generalized to non-linear stochastic models with expectational leads. Unbiased forecasting rules by definition provide best least-squares predictions conditional on the available information and thus rational expectations in the classical sense. We interpret unbiased forecasting rules as the generators for rational expectations equilibria and use them to characterize these equilibria. This result complements the literature on rational expectations which often presumes extraordinary capabilities of agents without making explicit how agents can obtain the desired degree of rationality. In fact, the focus of the traditional literature is on the existence of rational-expectations solutions for particular models rather than on the forecasting rules pertaining to the forecasts, thus taking a solutions approach rather than a dynamical-systems approach, e.g., see Evans & Honkapohja (2001). The concept of a forecasting rule together with the notion of unbiasedness provides a link between the traditional rational-expectations view and recent behavioristic approaches.

A key feature of systems with expectational leads is the fact that several forecasts, each formed at a different time, refer to the same realization of an endogenous variable. This property turns out to be the primary cause for an inherent multiplicity of unbiased forecasting rules and thus for the multiplicity of rational expectations equilibria. Extending Böhm & Wenzelburger (2002), we shed new light on this phenomenon by showing that an invertibility condition of the economic law is responsible for the existence and uniqueness of an unbiased forecasting rule. A special role is played by *no-updating* forecasting rules which are characterized by the property that forecasts formed at previous dates are never updated. We will show that unbiased *no-updating* forecasting rules yield the most precise forecasts in the sense that all forecasts are best least-squares predictions conditional on information available at a particular date. This includes those forecasts which were formed at a stage when this information was not yet available. This phenomenon is caused by the feedback property of the forecasts and constitutes a major difference between models with expectational leads and those without. In the linear case, the existence of an unbiased no-updating rule depends on the non-singularity of a particular parameter matrix, so that linear unbiased no-updating forecasting rules are analytically tractable whenever they exist.

Having established the concept of an unbiased forecasting rule, the question is how agents can learn such a rule from historical data. The main methodological innovation of these notes is the systematic treatment of forecasts as exogenous inputs in the sense of systems and control theory (see, e.g., Caines 1988). At each point in time, we distinguish between the estimation of the system with an emphasis on the feedback of all relevant forecasts and the approximation of a desired forecasting rule. Four separate issues should thereby be distinguished carefully. First, the existence of a desired forecasting rule; second, the dynamic stability of the system under this particular forecasting rule; third, the dynamic stability of the system under the applied learning scheme and, fourth, the success of the learning scheme in terms of providing strongly consistent estimates of a forecasting rule. It will turn out that it is essentially the economic law which has to be estimated from time series data. The approach proposed in these notes opens up the possibility to select an approximation of a preferred forecasting rule on the basis of economic reasoning. It also allows, in principle, to keep a system dynamically stable at all stages of the learning scheme.

The core of our adaptive learning scheme consists of nonparametric estimation techniques developed by Chen & White (1996, 1998, 2002) which are based on stochastic approximation methods of Yin & Zhu (1990) and Yin (1992). Nonparametric estimation techniques generalize the recursive ordinary least-squares algorithm which is well established in the learning literature. These techniques are indispensable when treating non-linear models without linearizing. Already Kuan & White (1994) and Chen & White (1998) incorporated the expectations feedback into the estimation of an economic system. However, it remains an open issue how to deal with the multiplicity of rational-expectations solutions in a more systematic manner. Since these multiplicities will occur generically, it is important to have a method that allows to select between multiple solutions on economic grounds rather than letting an algorithm decide which solution to select.

In these notes the issue of specifying plausible learning schemes will, in essence, be reduced to a pure estimation and control problem. Roughly speaking, if the estimation technique succeeds in providing consistent estimators of the economic law, then the learning scheme will converge to the desired forecasting rule. Clearly, situations in which agents attain rational expectations are likely to remain special cases, as long as model specifications for which no consistent estimators exist are conceivable. A main conclusion, however, is that the proper incorporation of the expectations feedback is an indispensable methodological feature for the design of plausible learning schemes, as expectations, in general, will have a non-trivial feedback effect on an economic system. It is left for future research to apply the methodology advocated in these notes to game theoretic settings and combine it with the theory of learning in games developed by Fudenberg & Levine (1998). An application to duopoly games, inspired by Kirman (1975) and Chen & White (1998), can be found in Wenzelburger (2004a) as a first step in this direction.

Most of the material of these notes does not have to be read in a chronological order. The general setup for non-linear economic models with expectational leads is introduced in Chapter 2. The basic methodological concepts are essentially contained in Chapter 3 which contains a nearly self-contained treatment of linear models and can be read without any knowledge about nonparametric methods. Chapter 4, which may be omitted in a first reading, is primarily concerned with the issue of existence and uniqueness of unbiased forecasting rules and contains refinements of Chapter 2. The nonparametric estimation techniques required for non-linear models are presented in Chapter 5. Those readers who are primarily interested in applications may start with either Chapter 6 or Chapter 7. As an example for a non-linear model, a standard version of a stochastic exchange economy is discussed in Chapter 6, a more involved financial market model is presented in Chapter 7.

2

Economic Systems With Expectations Feedback

A fundamental feature of many economic systems is the interaction of a market mechanism with expectations of agents who try to predict the future evolution of the economy. As a way of formalizing this feedback effect of expectations, we distinguish between an *economic law* describing the basic market mechanism of an economy and a *forecasting rule* which models the way in which an agent or a forecasting agency forms expectations. The combination of both ingredients with a model for exogenous perturbations constitutes a *deterministic dynamical system* in a *random environment* which describes the evolution of endogenous variables of an economy. Many well-known economic models can be cast in such a form. Powerful mathematical methods are available with the help of which their dynamic and stochastic behavior can be analyzed. The mathematical framework with many intuitive examples including iterated function systems (see Barnsley 1988) is, for example, provided in Lasota & Mackey (1994). More advanced methods are developed by Arnold (1998), Borovkov (1998) and others.

In this chapter the *non-linear* setup introduced in Böhm & Wenzelburger (1999, 2002, 2004) is extended to stochastic systems which involve forecasts for an arbitrarily finite period ahead of the current state of the system. The aim of these notes is to introduce a learning scheme which will find the best least-squares predictions in such models. To this end we will first introduce and classify those forecasting rules which provide the best least-squares predictions for a given economic scenario. We will show that these so-called *unbiased forecasting rules* together with the economic law are the generators of rational expectations equilibria. In Chapter 5 we will develop nonparametric methods to estimate unbiased forecasting rules from historical data. The distinction between an economic law, a forecasting rule, and a model for exogenous perturbations leads us to treating expectations systematically as exogenous inputs of an economic law. This concept will turn out to be important when establishing convergence results for learning schemes.

Before introducing the general setup of these notes, let us start with a basic introductory example which stylizes a standard model in macroeconomics.

2.1 An Introductory Example

Consider a univariate linear model of the form

$$y_{t+1} = Ay_t + B^{(1)}y^e_{t,t+1} + B^{(0)}y^e_{t,t+2} + \xi_{t+1}, \tag{2.1}$$

where $y_t \in \mathbb{R}$ denotes the endogenous variable describing the state of the economy at date t, A, $B^{(0)}$, and $B^{(1)}$ are non-random constants, and $\{\xi_t\}_{t\in\mathbb{N}}$ is a sequence of exogenously given iid shocks. The variables $y^e_{t,t+1}, y^e_{t,t+2} \in \mathbb{R}$ are forecasts for future realizations y_{t+1} and y_{t+2}, respectively, which are based upon information available at date t. Up to an additive constant which is omitted for simplicity, Evans & Honkapohja (2001, Chaps. 8 and 9) provide many well-founded economic examples which are of the functional form (2.1). In general, the model (2.1) admits multiple rational-expectations solutions, whereas the so-called minimal-state-variable solutions (MSV solutions) play a prominent role.

An MSV solution, introduced by McCallum (1983), is an AR(1) process of the form

$$y_{t+1} = \beta_\star y_t + \xi_{t+1}, \tag{2.2}$$

where, according to the method of undetermined coefficients, $\beta_\star$ is a solution of the quadratic equation

$$0 = A + (B^{(1)} - 1)\beta + B^{(0)}\beta^2. \tag{2.3}$$

Clearly, MSV solutions are economically meaningful only for real $\beta_\star$, i.e., for $(1 - B^{(1)})^2 \geq 4AB^{(0)}$ and, as is well known, they are generically non-unique. We will show in Example 2.2 below, any real solution $\beta_\star$ of (2.3) defines a forecasting rule

$$y^e_{t,t+1} = \beta_\star y_t, \quad y^e_{t,t+2} = \beta_\star^2 y_t \tag{2.4}$$

which is unbiased in the sense that it provides best least-squares predictions for the system (2.1) when the forecasts (2.4) are inserted. That is to say, the process (2.2) generates rational expectations equilibria of the system (2.1).

Imagine now a forecasting agency in charge of issuing forecasts without knowing the true model (2.1). Their first task will be to learn the rational-expectations solutions of the model by means of an adaptive learning scheme. The second task is to select between multiple such solutions. For example, if (2.1) admits two real MSV solutions $\beta_\star^{(1)}$ and $\beta_\star^{(2)}$ with $|\beta_\star^{(1)}| < 1 < |\beta_\star^{(2)}|$, it may be desirable to select the non-explosive one $\beta_\star^{(1)}$. Depending on the economic context, other solutions such as ARMA solutions might also be of particular interest, cf. Evans & Honkapohja (1986). One way of selecting a rational expectations solution is given by the *expectational-stability criterion* first introduced by Evans (1985, 1986) and further developed in Evans (1989) and Evans & Honkapohja (1992). This criterion provides conditions under which a learning scheme converges to a particular rational-expectations solution. As has been noted by McCallum (1999), the problem with this approach is that it is beyond the control of a forecasting agency to select a particular solution.

In terms of the linear model (2.1), the methodological difference of the approach advocated in these notes in comparison to the traditional literature can be outlined as follows. The classical approach starts with a so-called *perceived law of motion* which, in view of possible MSV solutions, is of the form

$$y_{t+1} = \beta y_t + \xi_{t+1}. \tag{2.5}$$

Then the method of ordinary least squares (OLS) is used to estimate the unknown coefficient β in (2.5) from historical observations. This learning scheme is successful if the sequence of subsequent OLS-estimates $\{\hat{\beta}_t\}_{t\in\mathbb{N}}$ for β converges to a MSV solution. Notice that there is no way to select between the two possible solutions $\beta_\star^{(1)}$ and $\beta_\star^{(2)}$ on the basis of economic reasoning. Instead, the convergence properties of the OLS scheme determine the outcome.

Contrary to this traditional approach, the basic idea of our approach is to successively estimate the parameters $(A, B^{(0)}, B^{(1)})$ of (2.1) from a perceived economic law

$$y_{t+1} = \hat{A} y_t + \hat{B}^{(1)} y^e_{t,t+1} + \hat{B}^{(0)} y^e_{t,t+2} + \xi_{t+1} \tag{2.6}$$

which is of the very same form as (2.1), thereby treating the forecasts $y^e_{t,t+1}$ and $y^e_{t,t+2}$ as exogenous inputs. These may be thought of as being obtained from some previously applied forecasting rule. Let $(\hat{A}_t, \hat{B}_t^{(0)}, \hat{B}_t^{(1)})$ denote the period-t OLS-estimates of the unknown parameters $(A, B^{(0)}, B^{(1)})$ obtained from estimating the perceived law (2.6). Then the equation

$$0 = \hat{A}_t + (\hat{B}_t^{(1)} - 1)\beta + \hat{B}_t^{(0)}\beta^2 \tag{2.7}$$

approximates equation (2.3) which defines MSV solutions. Whenever (2.7) admits positive real solutions $\hat{\beta}_t^{(1)}$ and $\hat{\beta}_t^{(2)}$, these approximate the two possible MSV solutions. A major advantage of such an approach is that it allows to select an approximation of the preferred solution for each parameter estimate using economic criteria. As soon as the estimates are sufficiently precise, a forecasting agency, for instance, would be in a position to choose a non-explosive solution to (2.1) whenever such a solution exists. Approximations of other solutions could be computed from $(\hat{A}_t, \hat{B}_t^{(0)}, \hat{B}_t^{(1)})$ as well. Moreover, a forecasting agency may want to alter the forecasts in order to keep the system dynamically stable.

This methodology of estimating linear models with expectational leads will be treated in Chapter 3. In subsequent chapters, we will show that this methodology carries over to general non-linear models with expectational leads of arbitrary finite length. Such a generalization is necessary, for example, to treat overlapping generations models with multi-period life cycles, cf. Auerbach & Kotlikoff (1987). The remainder of this chapter will therefore be devoted to the general non-linear setup used throughout these notes.

2.2 The General Setup

Let $\mathbb{X} \subset \mathbb{R}^d$ denote the space of all possible realizations of the endogenous variables and $x_t \in \mathbb{X}$ be the vector of *endogenous* variables describing the state of the economy at date t. The vector of *exogenous variables* is denoted by $\xi_t \in \Xi$ and assumed to be driven by an exogenous stochastic process. To fix notation, we make the following assumption regarding this noise process.

Assumption 2.1. *The exogenous noise is driven by a sequence of random variables $\{\xi_t\}_{t\in\mathbb{N}}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a subset $\Xi \subset \mathbb{R}^{d_\xi}$, such that for each $t \in \mathbb{N}$, $\xi_t : \Omega \to \Xi$.*

It is often convenient to assume that each state $x_t \in \mathbb{X}$ can be subdivided into $x_t = (\overline{x}_t, y_t) \in \overline{\mathbb{X}} \times \mathbb{Y} = \mathbb{X}$, where $y_t \in \mathbb{Y} \subset \mathbb{R}^{d_y}$ describes the endogenous variables for which expectations are formed, and $\overline{x}_t \in \overline{\mathbb{X}} \subset \mathbb{R}^{\bar{d}_x}$ is the vector of the remaining variables with $d_x = \bar{d}_x + d_y$. Let

$$\mathbb{Y}^m = \underbrace{\mathbb{Y} \times \ldots \times \mathbb{Y}}_{m - \text{times}}.$$

denote the m-fold product of $\mathbb{Y}$. An *economic law* is a map

$$G : \Xi \times \mathbb{X} \times \mathbb{Y}^m \longrightarrow \mathbb{X} \tag{2.8}$$

with the interpretation that

$$x_{t+1} = G(\xi_{t+1}, x_t, y^e_{t,t+1}, \ldots, y^e_{t,t+m})$$

describes the state of the economy at time $t + 1$, given the predicted values

$$y^e_{t,t+1}, \ldots, y^e_{t,t+m} \in \mathbb{Y}$$

formed at time t for $y_{t+1}, \ldots, y_{t+m}$, respectively, the current state x_t, and a realization $\xi_{t+1} \in \Xi$ of the stochastic noise variable. The state x_t itself may be a vector of lagged endogenous variables including past realizations ξ_s, $s \leq t$ of the exogenous noise process and past forecasts. We say that G has an *expectational lead* if it contains forecasts for more than one period ahead of the current state, that is, if $m > 1$. Throughout the remainder of these notes we assume $m > 1$ without further mentioning. The case $m = 1$ is extensively treated in Böhm & Wenzelburger (2002).

The deterministic map G represents the true economic system and determines the state of an economy one time step ahead of the current state, given subjective expectations and given the realization of the exogenous noise variable. From a stochastic viewpoint, each map

$$G(\xi_{t+1}(\cdot), x, y^e) : \Omega \to \mathbb{X}, \quad (x, y^e) \in \mathbb{X} \times \mathbb{Y}^m, \quad t \in \mathbb{N} \tag{2.9}$$

is a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In particular, given some perturbation $\omega \in \Omega$ and some $(x_t, y^e_t) \in \mathbb{X} \times \mathbb{Y}^m$, where

$$y_t^e := (y_{t,t+1}^e, \dots, y_{t,t+m}^e) \in \mathbb{Y}^m$$

denotes the vector of current forecasts, the realization

$$x_{t+1}(\omega) = G(\xi_{t+1}(\omega), x_t, y_t^e)$$

of the random variable (2.9) describes the state of the economy in period $t+1$.

It will be convenient to split the economic law (2.8) into two maps

$$\overline{G} : \Xi \times \mathbb{X} \times \mathbb{Y}^m \to \overline{\mathbb{X}} \quad \text{and} \quad g : \Xi \times \mathbb{X} \times \mathbb{Y}^m \to \mathbb{Y}$$

yielding

$$\begin{cases} \overline{x}_{t+1} = \overline{G}(\xi_{t+1}, \overline{x}_t, y_t, y_t^e) \\ y_{t+1} = g(\xi_{t+1}, \overline{x}_t, y_t, y_t^e) \end{cases} . \tag{2.10}$$

In many applications it is assumed that agents form expectations for all relevant variables of the economic process under consideration, in which case $\mathbb{Y} = \mathbb{X}$ and $G \equiv g : \Xi \times \mathbb{X}^{m+1} \longrightarrow \mathbb{X}$.

Remark 2.1. *In many economic models the point forecasts* $y_{t,t+1}^e, \dots, y_{t,t+m}^e \in \mathbb{Y}$ *will be based on a subjective joint probability distribution of all future realizations* $y_{t+1}, \dots, y_{t+m}$ *and thus may be interpreted as subjective mean values for these realizations. An economic law of the form (2.8) could then be defined as a function depending on subjective probability distributions rather than on point forecasts. For example, the overlapping generations model with multi-period life cycles and stochastic endowments fits into such a framework, e.g., see Kübler & Polemarchakis (2004) or Geanakopolos, Magill & Quinzii (2004). For the case of OLG models with two-period lives (*$m = 1$*), this has been done in Böhm & Wenzelburger (2002). In focusing on expectational leads, we assume that subjective probability distributions are parameterized in subjective means and choose the functional form (2.8) for simplicity of exposition. The examples of Chapters 6 and 7 indicate how to generalize the setup of this chapter to economic laws which depend on subjective distributions. The full-fledged analysis of stochastic OLG models with multiperiod lives must be left for future research, cf. Hillebrand (2006). A linear OLG-model with expectational leads is analyzed in Hillebrand & Wenzelburger (2006a).*

A complete description of the economy requires to specify a procedure of how the forecasts $y_{t,t+1}^e, \dots, y_{t,t+m}^e$ are determined in each period t. This specification will depend on the type of information that agents have available at time t. Unless stated otherwise, we assume throughout these notes that a forecasting agency which is boundedly rational in the sense of Sargent (1993) is responsible for issuing all relevant forecasts. The forecasts are determined according to a *forecasting rule* (also referred to as a predictor) which at date t is assumed to be a function depending on all available information.

Two cases will be treated. First, the case in which the economic system is governed by a Markov process with stationary transition probabilities. This case, for example, arises naturally, if the exogenous noise $\{\xi_t\}_{t\in\mathbb{N}}$ is an iid process. Second, the case in which the exogenous noise is driven by a stationary ergodic process. The first case will be treated next, the second one is postponed to Chapter 4.

2.3 Forecasting Rules With Finite Memory

For the remainder of this chapter, assume that the economic system is a homogeneous Markov process. In order to justify this assumption we need to specify the rules according to which forecasts are formed. At date t, the time series information of such a forecasting rule is comprised in the actual history of the economy $\{x_s, y^e_{s-1}\}_{s\leq t}$, where

$$y^e_{s-1} = (y^e_{s-1,s}, \dots, y^e_{s-1,s-1+m}) \in \mathbb{Y}^m, \quad s \in \mathbb{N}. \tag{2.11}$$

To fix the notation, let $r \in \mathbb{N}$ be arbitrary but fixed and set

$$\Sigma' := \underbrace{(\mathbb{Y}^m \times \mathbb{X}) \times \cdots \times (\mathbb{Y}^m \times \mathbb{X})}_{(r-1)\,-\,\text{times}}$$

and

$$\Sigma := \Sigma' \times (\mathbb{Y}^m \times \mathbb{X})$$

for the set of finite histories, relevant for the economy. Apart from the current state of the economy x_t, let us assume that the relevant information in period t is given by an information vector of past observed states and forecasts

$$Z_t = (y^e_{t-1}, x_{t-1}, \dots, y^e_{t-r}, x_{t-r}) \in \Sigma. \tag{2.12}$$

Formally, a forecasting rule for an economic law could be an arbitrary function depending on the whole history of the economy. In view of a Markovian structure and in order to classify those forecasting rules which are capable of providing best least-squares predictions, however, we focus on a special class of functions that depend on a finite amount of past realizations given by (2.12). The information vector (2.12) could, in principle, include other quantities such as sunspot variables. However, in order to obtain best least-squares predictions along a possible time path, it will turn out that this information vector contains precisely the required time-series information at date t, provided that the length of the memory r is long enough.

Letting (2.12) be the relevant information vector, a *forecasting rule* $\Psi = (\Psi^{(1)}, \dots, \Psi^{(m)})$ is formally defined as a function

$$\Psi : \mathbb{X} \times \Sigma \to \mathbb{Y}^m, \quad (x, Z) \mapsto \Psi(x, Z), \tag{2.13}$$

such that

$$y^e_{t,t+i} = \Psi^{(i)}(x_t, Z_t), \quad i = 1, \dots, m$$

describe the vector of forecasts $y^e_t = (y^e_{t,t+1}, \dots, y^e_{t,t+m})$ in period t. If $(x_t, Z_t, y^e_t) \in \mathbb{X} \times \Sigma \times \mathbb{Y}^m$ denotes the state of the system at date t, then the state of the system in period $t+1$ is given by

$$\begin{cases} x_{t+1} &= G\big(\xi_{t+1}, x_t, y^e_t\big) \\ Z_{t+1} &= \big(y^e_t, x_t, \mathrm{pr}^{\Sigma}_{-r} Z_t\big) \\ y^e_{t+1} &= \Psi(x_{t+1}, Z_{t+1}) \end{cases} \tag{2.14}$$

with

$$\mathrm{pr}^{\Sigma}_{-r} : \Sigma \to \Sigma', \quad (Z^{(1)}, \dots, Z^{(r)}) \mapsto (Z^{(1)}, \dots, Z^{(r-1)})$$

denoting the projection onto the first $r-1$ components. Replacing y^e_t with $\Psi(x_t, Z_t)$, the map (2.14) may be rewritten in the more compact form

$$G_\Psi : \Xi \times \mathbb{X} \times \Sigma \to \mathbb{X} \times \Sigma,$$

where

$$\begin{cases} x_{t+1} &= G\big(\xi_{t+1}, x_t, \Psi(x_t, Z_t)\big) \\ Z_{t+1} &= \big(\Psi(x_t, Z_t), x_t, \mathrm{pr}^{\Sigma}_{-r} Z_t\big) \end{cases} . \tag{2.15}$$

Assuming that the forecasting agency uses the forecasting rule Ψ, the evolution of the economy is driven by the economic law G together with Ψ. Given a sequence of realizations of the exogenous noise process $\{\xi_t\}_{t\in\mathbb{N}}$ and initial conditions $(x_0, Z_0) \in \mathbb{X} \times \Sigma$, the sequence of states $\gamma(x_0, Z_0) := \{(x_t, Z_t)\}_{t\in\mathbb{N}}$ generated by (2.15) describes how the economy evolves over time. In the sequel, we will refer to $\gamma(x_0, Z_0)$ as an *orbit* of the system G_Ψ.

To see when (2.15) is indeed a Markov process, define for each period $t \in \mathbb{N}$ the sub-σ-algebra

$$\mathcal{F}_{t,r} := \sigma\big((x_s, y^e_s)\big|\ t - r - 1 < s \le t\big) \tag{2.16}$$

of $\mathcal{F}$ which describes the relevant information of period t generated by the information vector (2.12). The whole history of the process at time t is represented by the sub-σ-algebra $\mathcal{F}_t := \mathcal{F}_{t,\infty}$. Let $\mathbb{E}[\cdot|\mathcal{F}_{t,r}]$ denote the conditional expectations operator with respect to $\mathcal{F}_{t,r}$ and $\mathbb{E}[\cdot|\mathcal{F}_t]$ denote the conditional expectations operator with respect to $\mathcal{F}_t$. Then the assumption that (2.15) is a Markov process is formally stated as follows.

Assumption 2.2. *The length r of the information vector (2.12) is chosen sufficiently large so that (2.15) is a homogeneous Markov process in the sense that*

$$\mathbb{E}[x_{t+1}|\mathcal{F}_{t,r}] = \mathbb{E}[x_{t+1}|\mathcal{F}_t] \quad \mathbb{P} - a.s.$$

for all $t \in \mathbb{N}$.

The factorization lemma (Bauer 1992, p. 71) and the choice of the forecasting rule Ψ in (2.14) imply the existence of a measurable map $G_{\mathbb{E}}$ such that

$$\mathbb{E}[x_{t+1}|\mathcal{F}_{t,r}] = \mathbb{E}\big[G(\xi_{t+1}, x_t, y_t^e)|\mathcal{F}_{t,r}\big] = G_{\mathbb{E}}(x_t, Z_t) \quad \mathbb{P}-\text{a.s.} \tag{2.17}$$

for each $t \in \mathbb{N}$. Notice that each forecast y_t^e is based on information available in period t and hence must be $\mathcal{F}_t$ measurable. Assumption 2.2 together with (2.17) thus justify the functional form of the forecasting rule (2.13).

A special case arises when $\{\xi_t\}_{t\in\mathbb{N}}$ is an iid process.

Remark 2.2. *If $\{\xi_t\}_{t\in\mathbb{N}}$ is an iid process, then it is well known that*

$$\mathbb{E}[x_{t+1}|\mathcal{F}_t] = \mathbb{E}\big[G(\xi_{t+1}, x_t, y_t^e)\big] = G_{\mathbb{E}}(x_t, y_t^e) \quad \mathbb{P}-a.s.$$

for all $t \in \mathbb{N}$, cf. Bauer (1991, p. 134). In this case one may set $\Sigma = \mathbb{Y}^m$ for the set of finite histories, $r = 0$ for the memory length, and $\mathcal{F}_{t,0} = \mathcal{F}$ for all $t \in \mathbb{N}$. Choosing the information vector of period t to be $Z_t = y_{t-1}^e$, one obtains

$$\mathbb{E}[y_{t+1}^e|\mathcal{F}_t] = \mathbb{E}\big[\Psi\big(G(\xi_{t+1}, x_t, y_t^e), y_t^e\big)\big] = \Psi_{\mathbb{E}}(x_t, y_t^e) \quad \mathbb{P}-a.s.$$

for some suitable function $\Psi_{\mathbb{E}}$, so that (2.15) becomes the Markov process

$$\begin{cases} x_{t+1} &= G\big(\xi_{t+1}, x_t, \Psi(x_t, Z_t)\big) \\ Z_{t+1} &= \Psi(x_t, Z_t) \end{cases}.$$

The justification for this choice of Z_t will become clear in Remark 2.4 below, after unbiased forecasting rules have been introduced.

We are now ready to introduce the notion of an *unbiased forecasting rule* which is designed to provide best least-squares predictions for a given economic law for the case in which the system (2.14) is a homogeneous Markov process.

2.4 Consistent Forecasting Rules

Before unbiased forecasting rules can be introduced, the notion of a consistent forecasting rule will be established in this section. A fundamental property of systems with expectational leads is that in each period t there are m forecasts for y_{t+1} of which $m-1$ forecasts have been set prior to date t. This implies that the current forecast will in some sense have to be consistent with the previously formed forecasts. Otherwise not all forecasts can be correct. Consistent forecasting rules will now be designed to form consistent expectations in the sense that the law of iterated expectations is respected.

As before, denote by $\mathbb{E}[\cdot|\mathcal{F}_{t-i}]$ the conditional expectations operator with respect to $\mathcal{F}_{t-i}$. Each forecast $y^e_{t-i,t+1}$ is based on information available in period $t-i$ and must be $\mathcal{F}_{t-i}$ measurable. The forecast errors $y_{t+1} - y^e_{t-i,t+1}$, $i = 0, \ldots, m-1$, vanish on average conditional on the respective σ-algebras $\mathcal{F}_{t-i}$ if and only if

$$\mathbb{E}\big[(y_{t+1} - y^e_{t-i,t+1})\big|\mathcal{F}_{t-i}\big] = 0 \quad \mathbb{P} - \text{a.s.} \tag{2.18}$$

for all $i = 0, \ldots, m-1$. In other words, each of the forecasts satisfying (2.18) is a best least-squares prediction for y_{t+1} conditional on available information. Hence, a forecasting rule which satisfies condition (2.18) in all periods t is unbiased conditional on available information in the sense that all forecasts have the smallest possible forecast error.

The law of iterated expectations states that

$$\mathbb{E}\big[\mathbb{E}[y_{t+1}|\mathcal{F}_t]\big|\mathcal{F}_{t-i}\big] = \mathbb{E}[y_{t+1}|\mathcal{F}_{t-i}] \quad \mathbb{P} - \text{a.s.} \tag{2.19}$$

for all times $t \in \mathbb{N}$ and all $i \geq 0$. This implies that the forecast $y^e_{t,t+1}$ for y_{t+1} chosen in period t has to satisfy the *consistency conditions*

$$\mathbb{E}[y^e_{t,t+1}|\mathcal{F}_{t-i}] = y^e_{t-i,t+1} \quad \mathbb{P} - \text{a.s.} \tag{2.20}$$

for all $i = 1, \ldots, m-1$. Otherwise one of the forecasts in (2.20) cannot be a best least-squares prediction. Similarly, the law of iterated expectations implies for the forecasts $y^e_{t,t+i}$, $i = 1, \ldots, m-1$, that

$$\mathbb{E}[y^e_{t,t+i}|\mathcal{F}_{t-1}] = y^e_{t-1,t+i} \quad \mathbb{P} - \text{a.s.} \tag{2.21}$$

Clearly, if (2.21) holds for all forecasts in all periods prior to time t, then (2.20) holds as well. Hence (2.18) is automatically satisfied provided that the most recent forecast $y^e_{t,t+1}$ is a best least-squares prediction, that is, if

$$\mathbb{E}\big[(y_{t+1} - y^e_{t,t+1})\big|\mathcal{F}_t\big] = 0 \quad \mathbb{P} - \text{a.s.}$$

This shows that the notion of consistency is a basic prerequisite for an unbiased forecasting rule.

The consistency conditions (2.21) can now be expressed in terms of forecasting rules $\Psi = (\Psi^{(1)}, \ldots, \Psi^{(m)})$ of the form (2.13) by making use of the Markov property of (2.15). Let $\mathbb{E}_j[\cdot] \equiv \mathbb{E}[\cdot|\mathcal{F}_{j,r}]$ denote the conditional expectations operator with respect to the σ-algebra $\mathcal{F}_{j,r}$, with $r \in \mathbb{N}$ as above. Since (2.15) generates a homogeneous Markov process, we have

$$\mathbb{E}[y^e_{t,t+i}|\mathcal{F}_{t-1}] = \mathbb{E}_{t-1}\big[\Psi^{(i)}(x_t, Z_t)\big] \quad \mathbb{P} - \text{a.s.}$$

for each $i = 1, \ldots, m-1$ and each $t \in \mathbb{N}$. By the factorization lemma, there exists a measurable map $\Psi^{(i)}_{\mathbb{E}}$ such that

$$\mathbb{E}[y^e_{t,t+i}|\mathcal{F}_{t-1}] = \Psi^{(i)}_{\mathbb{E}}(x_{t-1}, Z_{t-1}) \quad \mathbb{P}-\text{a.s.} \tag{2.22}$$

for all $i = 1,\dots,m-1$. Using (2.22), the consistency conditions (2.21) then take the form

$$\Psi^{(i)}_{\mathbb{E}}(x_{t-1}, Z_{t-1}) = y^e_{t-1,t+i} \quad \mathbb{P}-\text{a.s.} \tag{2.23}$$

for all $i = 1,\dots,m-1$. The requirement (2.23) states that for each $i = 1,\dots,m-1$, the expected value of the new forecast $y^e_{t,t+i}$ conditional on $\mathcal{F}_{t-1}$ has to coincide with the respective old forecast $y^e_{t-1,t+i} = \Psi^{(i+1)}(x_{t-1}, Z_{t-1})$. This motivates the following notion of a consistent forecasting rule.

Definition 2.1. *A forecasting rule* $\Psi = (\Psi^{(1)},\dots,\Psi^{(m)})$ *is called* locally consistent *if there exists a subset* $\mathbb{U} \subset \mathbb{X} \times \Sigma$*, such that for each* $i = 1,\dots,m-1$*,*

$$\Psi^{(i+1)}(x, Z) = \Psi^{(i)}_{\mathbb{E}}(x, Z) \quad \textit{for all} \ \ (x, Z) \in \mathbb{U},$$

with $\Psi^{(i)}_{\mathbb{E}}$ *being defined in (2.22). If* $\mathbb{U} = \mathbb{X} \times \Sigma$*, then* Ψ *is called globally consistent.*

Definition 2.1 provides a minimum requirement for a forecasting rule of the form (2.13) in order to provide best least-squares predictions. A consistent forecasting rule is designed to satisfy (2.21) by construction on a possibly restricted subset $\mathbb{U} \subset \mathbb{X} \times \Sigma$. It follows from the law of iterated expectations that otherwise (2.20) and hence (2.18) cannot be satisfied for all forecasts.

Example 2.1. *The simplest example of a consistent forecasting rule is a forecasting rule that never updates previous forecasts. Such a* no-updating *forecasting rule* $\Psi = (\Psi^{(1)},\dots,\Psi^{(m)})$ *is formally defined by setting*

$$\Psi^{(i)}\big(x, (z, Z')\big) := z^{(i+1)}, \quad (x, z, Z') \in \mathbb{X} \times \mathbb{Y}^m \times (\mathbb{X} \times \Sigma') \tag{2.24}$$

for each $i = 1,\dots,m-1$*, where* $z = (z^{(1)},\dots,z^{(m)}) \in \mathbb{Y}^m$*.* $\Psi^{(m)}$ *may be an arbitrary function depending on the state* x *and the history* $Z = (z, Z')$*. If* $x_t \in \mathbb{X}$ *is the current state of the economy at date* t *and* $y^e_{t-1} = (y^e_{t-1,t},\dots,y^e_{t-1,t-1+m})$ *the vector of forecasts formed in the previous period* $t-1$*, then the current vector of forecasts* $y^e_t = (y^e_{t,t+1},\dots,y^e_{t,t+m})$ *is*

$$\begin{cases} y^e_{t,t+i} &= y^e_{t-1,t+i}, \qquad i = 1,\dots,m-1, \\ y^e_{t,t+m} &= \Psi^{(m)}(x_t, Z_t). \end{cases}$$

Thus the forecasting rule (2.24) never updates forecasts formed at previous dates.

2.5 Unbiased Forecasting Rules

We are now ready to develop the notion of an unbiased forecasting rule for the case in which the system (2.14) is a homogeneous Markov process. Unbiased forecasting rules are designed to provide best least-squares predictions, given the history of the process. They will also be referred to as *perfect forecasting rules for first moments*. If applied, these forecasting rules provide correct first moments along orbits of the system. In this sense, an unbiased forecasting rule $\Psi_\star$ together with the economic law G generates rational expectations equilibria (REE).

Notice first that the Markov property of the process (2.15) and the factorization lemma (Bauer 1992, p. 71) imply

$$\mathbb{E}_t[y_{t+1}] = \mathbb{E}_t\big[g(\xi_{t+1}, x_t, y_t^e)\big] = g_{\mathbb{E}}(x_t, y_t^e, Z_t) \quad \mathbb{P} - \text{a.s.} \tag{2.25}$$

for some suitable measurable map $g_{\mathbb{E}}$. Given the state of the economy $x_t \in \mathbb{X}$ and a vector of forecasts $y_t^e \in \mathbb{Y}^m$, the forecast error for y_{t+1} is

$$y_{t+1} - y_{t,t+1}^e = g(\xi_{t+1}, x_t, y_t^e) - y_{t,t+1}^e,$$

while the forecast error conditional on information available at date t is

$$\mathbb{E}_t\big[y_{t+1} - y_{t,t+1}^e\big] = g_{\mathbb{E}}(x_t, y_t^e, Z_t) - y_{t,t+1}^e \quad \mathbb{P} - \text{a.s.} \tag{2.26}$$

Using (2.26), this gives rise to the definition of a *(conditional) mean error function* associated with the economic law G.

Definition 2.2. *Let* $G = (\overline{G}, g)$ *be an economic law. The* (conditional) mean error function *of* G *is a function* $\mathcal{E}_G : \mathbb{Y}^m \times \mathbb{X} \times \Sigma \to \mathbb{R}^{d_y}$*, defined by*

$$\mathcal{E}_G(z, x, Z) := g_{\mathbb{E}}\big(x, z, Z\big) - z^{(1)}, \tag{2.27}$$

where $z = (z^{(1)}, \ldots, z^{(m)})$ *and* $g_{\mathbb{E}}$ *is defined in (2.25).*

The error function takes a particularly simple form if the exogenous noise is iid.

Remark 2.3. *If* $\{\xi_t\}_{t \in \mathbb{N}}$ *is an iid process, then*

$$\mathbb{E}_t[y_{t+1}] = \mathbb{E}\big[g(\xi_{t+1}, x_t, y_t^e)\big] = g_{\mathbb{E}}(x_t, y_t^e) \quad \mathbb{P} - a.s.$$

for all $t \in \mathbb{N}$*. In this case the error function takes the form*

$$\mathcal{E}_G(z, x) := g_{\mathbb{E}}(x, z) - z^{(1)}, \quad (x, z) \in \mathbb{X} \times \mathbb{Y}^m,$$

where $z = (z^{(1)}, \ldots, z^{(m)})$*.*

Given an arbitrary state of the economy $(x_t, Z_t) \in \mathbb{X} \times \Sigma$ in some period t, the mean error function describes all possible mean forecast errors conditional on all information available at date t, regardless of which forecasting or learning rule has been applied. Given the state (x_t, Z_t), the conditional forecast error (2.26) vanishes if and only if

$$\mathbb{E}_t[y_{t+1} - y^e_{t,t+1}] = \mathcal{E}_G(y^e_t, x_t, Z_t) = 0 \quad \mathbb{P} - \text{a.s.} \tag{2.28}$$

The idea of an unbiased forecasting rule is now to find a locally consistent forecasting rule $\Psi : \mathbb{U} \subset \mathbb{X} \times \Sigma \to \mathbb{Y}^m$ with $y^e_t = \Psi(x_t, Z_t)$ which satisfies

$$\mathcal{E}_G\big(\Psi(x_t, Z_t), x_t, Z_t\big) = 0 \tag{2.29}$$

whenever $(x_t, Z_t) \in \mathbb{U}$. The existence of an unbiased forecasting rule is thus reduced to solving the implicit equation (2.28). More formally, the desired criterion for an unbiased forecasting rule is defined as follows.

Definition 2.3. *Given an economic law $G = (\overline{G}, g)$, a consistent forecasting rule Ψ is called* locally unbiased *for G, if there exists an open subset $\mathbb{U} \subset X \times \Sigma$ such that*

$$\mathcal{E}_G\big(\Psi(x, Z), x, Z\big) = 0 \quad \textit{for all} \;\; (x, Z) \in \mathbb{U}.$$

A major problem with rational/unbiased expectations in non-linear systems will be that an orbit $\gamma(x_0, Z_0)$ of G_Ψ starting with unbiased forecasts in some $(x_0, Z_0) \in \mathbb{U}$ may lose this property over time. Thus, if Ψ is a forecasting rule which is locally unbiased in a neighborhood $\mathbb{U} \subset \mathbb{X} \times \Sigma$ of (x_0, Z_0), then an additional requirement that guarantees that the orbit stays in $\mathbb{U}$ is needed. Formally, a sufficient requirement for the existence of such *unbiased orbits* is that there exists a subset $\mathbb{U}' \subset \mathbb{U}$ which is forward invariant under G_Ψ, that is, for which

$$G_\Psi(\xi, x, Z) \in \mathbb{U}' \quad \text{for all} \quad (x, Z) \in \mathbb{U}', \; \xi \in \Xi. \tag{2.30}$$

In this case, any orbit $\gamma(x_0, Z_0)$ that starts in $(x_0, Z_0) \in \mathbb{U}'$ will forever remain in $\mathbb{U}'$. Summarizing, an unbiased forecasting rule is defined as follows.

Definition 2.4. *Given an economic law G, a function $\Psi_\star = (\Psi^{(1)}_\star, \ldots, \Psi^{(m)}_\star)$ is called an* unbiased forecasting rule *for G if there exists a subset $\mathbb{U} \subset \mathbb{X} \times \Sigma$ such that the following holds:*

(i) Consistency. For $i = 1, \ldots, m-1$,

$$\Psi^{(i+1)}_\star(x, Z) = \Psi^{(i)}_{\star\mathbb{E}}(x, Z) \quad \textit{for all} \;\; (x, Z) \in \mathbb{U}.$$

(ii) Local unbiasedness. $\Psi_\star$ satisfies

$$\mathcal{E}_G\big(\Psi_\star(x, Z), x, Z\big) = 0 \quad \textit{for all} \;\; (x, Z) \in \mathbb{U}.$$

(iii) Invariance. $\mathbb{U}$ *is forward-invariant under* $G_{\Psi_\star}$*, i.e.,*

$$G_{\Psi_\star}(\xi, x, Z) \in \mathbb{U} \quad \text{for all} \quad (x, Z) \in \mathbb{U},\ \xi \in \Xi.$$

The three conditions of Definition 2.4 imply that a forecasting rule $\Psi_\star$ is unbiased if and only if *all* orbits $\gamma(x_0, Z_0)$, $(x_0, Z_0) \in \mathbb{U}$ of $G_{\Psi_\star}$ are unbiased orbits. That is, for each $t \in \mathbb{N}$, the forecasts $y^e_{t,t+1}, \ldots, y^e_{t,t+m} \in \mathbb{Y}$ provide best least-squares predictions for $y_{t+1}, \ldots, y_{t+m}$, respectively. In this sense, unbiased forecasting rules generate rational expectations equilibria (REE).

Example 2.2. *Reconsider the linear economic law (2.1) of Section 2.1 as given by*

$$y_{t+1} = A y_t + B^{(1)} y^e_{t,t+1} + B^{(0)} y^e_{t,t+2} + \xi_{t+1}, \tag{2.31}$$

where $\{\xi_t\}_{t\in\mathbb{N}}$ *is a sequence of iid perturbations with zero means. The error function associated with (2.31) is*

$$\mathcal{E}_G(y^e_{t,t+1}, y^e_{t,t+2}, y_t) = A y_t + (B^{(1)} - 1) y^e_{t,t+1} + B^{(0)} y^e_{t,t+2}. \tag{2.32}$$

Suppose $B^{(0)} \neq 0$. *Solving the condition*

$$\mathcal{E}_G(y^e_{t,t+1}, y^e_{t,t+2}, y_t) \stackrel{!}{=} 0$$

for $y^e_{t,t+2}$*, we obtain an unbiased no-updating forecasting rule which takes the form (compare with Example 2.1)*

$$\begin{cases} y^e_{t,t+1} = \Psi^{(1)}_\star(y_t, y^e_{t-1,t+1}) := y^e_{t-1,t+1}, \\ y^e_{t,t+2} = \Psi^{(2)}_\star(y_t, y^e_{t-1,t+1}) := \frac{1}{B^{(0)}}\left[(1 - B^{(1)}) y^e_{t-1,t+1} - A y_t\right]. \end{cases} \tag{2.33}$$

As an alternative, consider a forecasting rule of the form

$$\begin{cases} y^e_{t,t+1} = \Phi^{(1)}(y_t) := \beta y_t, \\ y^e_{t,t+2} = \Phi^{(2)}(y_t) := \beta^2 y_t, \end{cases} \tag{2.34}$$

which has already been introduced in Section 2.1. Now, the unknown coefficient β *may be obtained by inserting the forecasting rule (2.34) into the error function (2.32) and solving the equation*

$$\mathcal{E}_G(\beta y, \beta^2 y, y) \stackrel{!}{=} 0 \quad \text{for all} \quad y \in \mathbb{R}$$

for the unknown parameter β. *This equation is equivalent to the quadratic equation*

$$A + (B^{(1)} - 1)\beta + B^{(0)} \beta^2 \stackrel{!}{=} 0. \tag{2.35}$$

The solutions to (2.35) are given by

$$\beta_\star^{(i)} = \frac{1-B^{(1)}}{2B^{(0)}} \pm \sqrt{\left(\frac{1-B^{(1)}}{2B^{(0)}}\right)^2 - \frac{A}{B^{(0)}}}, \quad i=1,2. \tag{2.36}$$

Clearly, only real solutions are economically meaningful which is the case for $(1-B^{(1)})^2 \geq 4AB^{(0)}$*. As an alternative to the unbiased no-updating rule (2.33), we obtain an unbiased forecasting rule of the form (2.34). This rule is defined by*

$$\begin{cases} y^e_{t,t+1} = \beta_\star^{(i)} y_t, \\ y^e_{t,t+2} = \beta_\star^{(i)2} y_t, \end{cases} \tag{2.37}$$

with $\beta_\star^{(i)}$, $i=1,2$ *given in (2.36). It is straightforward to verify that (2.37) is consistent in the sense of Definition 2.1. Indeed the forecasting rule (2.37) satisfies*

$$\mathbb{E}_{t-1}[y^e_{t,t+1}] = \beta_\star^{(i)} \mathbb{E}_{t-1}[y_t] = \beta_\star^{(i)} y^e_{t-1,t} = \beta_\star^{(i)2} y_{t-1} = y^e_{t-1,t+1}$$

for all times t*. A forecasting rule of the form (2.34) is referred to as a minimal-state-variable predictor (MSV predictor) as it uses a minimum amount of variables.*

We see from this example that already linear models may admit multiple unbiased forecasting rules. The existence and uniqueness of unbiased forecasting rules for general linear models and methods to estimate these from historical data will be addressed in Chapter 3. See in particular Theorem 3.1.

We conclude this section with discussing the problem of existence of unbiased no-updating forecasting rules. Since all forecasts $y^e_{t,t+i}$, $i=1,\dots,m-1$ have to satisfy the above consistency conditions, the only forecast which is not constrained by the consistency conditions is the forecast $y^e_{t,t+m}$ for the most leading variable y_{t+m}. In the case with no-updating the first $m-1$ forecasts will be set to the values of the previous period such that

$$y^e_{t,t+i} = y^e_{t-1,t+i}, \quad i=1,\dots,m-1. \tag{2.38}$$

Given the information vector Z_t, the idea now is to find a forecasting rule

$$y^e_{t,t+m} = \Psi_\star^{(m)}(x_t, Z_t)$$

and a subset $\mathbb{U} \subset \mathbb{X} \times \Sigma$ so that

$$\mathcal{E}_G\big(y^e_{t-1,t+1}, \dots, y^e_{t-1,t+m-1}, \Psi_\star^{(m)}(x_t, Z_t), x_t, Z_t\big) = 0 \tag{2.39}$$

whenever $(x_t, Z_t) \in \mathbb{U}$. The economic implication for such an unbiased no-updating forecasting rule is that the newest forecast $y^e_{t,t+m}$ has to be chosen in such a way that the previous forecast $y^e_{t-1,t+1} = y^e_{t,t+1}$ becomes correct. Under no updating, then all previous forecasts $y^e_{t-i,t+1}$, $i=1,\dots,m-1$ become best least-squares predictions for y_{t+1} conditional on information available at date t. Indeed, since by construction (2.38) all forecasts for one variable coincide, (2.39) implies

$$\mathbb{E}_t[y_{t+1} - y^e_{t-i,t+1}] = 0 \quad \mathbb{P} - \text{a.s.} \tag{2.40}$$

for all $i = 0, \ldots, m-1$. Thus all previous forecasts inherit the precision of the most recent forecast $y^e_{t,t+1}$.

Remark 2.4. *If $\{\xi_t\}_{t\in\mathbb{N}}$ is an iid process, then the unbiased forecasting rule takes a special form. From Remark 2.3 we know that the error function reduces to*

$$\mathcal{E}_G(z, x) = g_{\mathbb{E}}(x, z) - z^{(1)}, \quad (x, z) \in \mathbb{X} \times \mathbb{Y}^m,$$

where $z = (z^{(1)}, \ldots, z^{(m)})$. The information vector of period t is $Z_t = y^e_{t-1}$ and the m-th component function $\Psi^{(m)}_\star$ of an unbiased no-updating forecasting rule $\Psi_\star$ is formally defined by $\Psi^{(m)}_\star : \mathbb{X} \times \mathbb{Y}^m \to \mathbb{Y}$. Condition (2.39) for the unbiasedness reads

$$\mathcal{E}_G\big(y^e_{t-1,t+1}, \ldots, y^e_{t-1,t-1+m}, \Psi^{(m)}_\star(x_t, y^e_{t-1,t+1}, \ldots, y^e_{t-1,t-1+m}), x_t\big) = 0$$

whenever $(x_t, y^e_{t-1}) \in \mathbb{U}$. This demonstrates that the functional form of $\Psi^{(m)}_\star$ is prescribed by the Implicit Function Theorem. In particular, $\Psi^{(m)}_\star$ must be independent of $y^e_{t-1,t}$ so that $\Psi^{(m)}_\star : \mathbb{X} \times \mathbb{Y}^{m-1} \to \mathbb{Y}$. This justifies the choices made in Remark 2.2.

We see from (2.39) that the existence problem of unbiased forecasting rules is, in essence, reduced to the Implicit Function Theorem which determines the functional form of an unbiased forecasting rule. In the Markovian case of this section, the existence theorem for the deterministic case with expectational leads can readily be applied, cf. Böhm & Wenzelburger (2004). Existence and uniqueness of unbiased no-updating forecasting rules will be established in Theorem 4.2 of Chapter 4 in which we generalize and refine our framework to economic models which are perturbed by stationary ergodic noise. Assuming that an unbiased no-updating forecasting rule exists and exploiting the presumed Markov structure, the last finding of this chapter is summarized in the following theorem.

Theorem 2.1. *Assume that under the hypothesis of this chapter, an unbiased no-updating forecasting rule for the economic law (2.8) in the sense of Definition 2.4 exists. Then for each $t \in \mathbb{N}$, all forecasts $y^e_{t-i,t+1}$, $i = 0, \ldots, m-1$ for y_{t+1} are best least-squares prediction conditional on the history $\mathcal{F}_t$, that is,*

$$\mathbb{E}\big[(y_{t+1} - y^e_{t-i,t+1})\big|\mathcal{F}_t\big] = 0 \quad \mathbb{P} - a.s.$$

for all $i = 0, \ldots, m-1$.

A remarkable fact in economic systems with expectational leads is that unbiased no-updating forecasting rules, if they exist, generate rational expectations in the sense that *all* forecasts $y^e_{t-m+1,t+1}, \dots, y^e_{t,t+1}$ are best least-squares predictions for y_{t+1} conditional on information available at date t, 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 conditional on information which is not available at the stage in which they were issued. This peculiar phenomenon is caused by the feedback property of forecasts and indeed holds true only for forecasts which feed back non-trivially into the economic law.

Concluding Remarks

The careful distinction between an economic law and a forecasting rule was the necessary step to obtain a proper description of an economic dynamical system in a Markovian environment. The results of this chapter demonstrate that the property of rational expectations along orbits of an economic dynamical system imposes structural restrictions on the economic fundamentals which are embodied in an economic law. A remarkable phenomenon in economic systems with expectational leads is the notion of an unbiased no-updating forecasting rule. Within the class of possible unbiased forecasting rules, these yield the most precise forecasts as forecasts are best least-squares predictions conditional on information which is not available at the stage in which they were issued. An economic implication of this chapter is that agents who know an unbiased forecasting rule also know how their forecasts influence the economy. In principle, these agents could try to strategically exploit this information. How this could be achieved is an interesting avenue for future research.

3

Adaptive Learning in Linear Models

In many economic applications linear models are popular either as exact formulations or as linear approximations of originally nonlinear specifications. For this reason it is worthwhile to devote a chapter to the study of the linear case. The advantage of linear models is that they do not require nonparametric estimation techniques, so that we can focus entirely on the methodology. In providing a benchmark situation for the general nonlinear case, it will be seen in Chapter 5 that the methodology for linear models carries over to the nonlinear situation.

The main difference between the adaptive learning scheme developed in this chapter and the traditional approach of the learning literature (e.g., see Kirman & Salmon (1995) or Evans & Honkapohja 2001) is that forecasts are systematically treated as exogenous inputs in the sense of the systems and control literature. At each point in time, we distinguish between the estimation of the economic law and the application of an approximated unbiased forecasting rule to the system. The main advantage of this approach is that the system can be kept dynamically stable at all stages of the learning procedure. It opens up the possibility to select a particular preferred forecasting rule on the basis of an adequate economic reasoning, as soon as parameter estimates are sufficiently precise. This, in essence, resolves an issue put forward in McCallum (1999).

Our adaptive learning scheme for linear stochastic models is based on the recursive extended least-squares algorithm which is well established in the engineering literature, cf. Caines (1988). This algorithm is an extension of the popular ordinary least-squares (OLS) algorithm to a class of more general noise processes. Using ideas from the literature on adaptive control and optimal tracking, results of several papers by Lai and Wei, notably their 1986b paper, will be applied here. These authors pursue a martingale approach paralleling and extending the original approach of Ljung (1977) who uses a limiting non-random ordinary differential equation (ODE) for his convergence analysis.

This chapter reports on refinements of Wenzelburger (2006). We relax standard assumptions of the traditional learning literature as given in Evans & Honkapohja (2001) in several respects. First, we admit expectational 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). Finally, we show convergence of our learning scheme globally for all initial conditions under simple and standard identifiability conditions. These conditions are generically satisfied in the white-noise case. Thus well-known convergence results of the learning literature are strengthened because the conditions require relatively mild technical prerequisites and are much easier to verify and interprete than those based on Ljung's ODE approach.

3.1 Linear Models

Consider a linear economic law of the form

$$y_t = \sum_{i=1}^{n_1} A^{(i)}\, y_{t-i} + \sum_{i=0}^{n_2} B^{(n_2-i)}\, y^e_{t-1,t+i} + \sum_{i=1}^{n_3} C^{(i)} w_{t-i} + \epsilon_t + \sum_{i=1}^{n_4} D^{(i)}\, \epsilon_{t-i} + b, \tag{3.1}$$

where $y_t \in \mathbb{R}^d$ denotes a vector of endogenous variables describing the state of the economy at date t. $A^{(i)}$, $B^{(i)}$, $C^{(i)}$, and $D^{(i)}$ are non-random $d \times d$ matrices and $b \in \mathbb{R}^d$ is a constant non-random vector. The exogenous perturbations are prescribed by means of two stochastic processes $\{w_t\}_{t\in\mathbb{N}}$ and $\{\epsilon_t\}_{t\in\mathbb{N}}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which are adapted to a filtration $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$. $\{w_t\}_{t\in\mathbb{N}}$ describes the process of $\mathbb{R}^d$-valued exogenous variables, which is assumed to be observable. The process $\{\epsilon_t\}_{t\in\mathbb{N}}$ is an $\mathbb{R}^d$-valued martingale difference sequence, i.e. $\mathbb{E}[\epsilon_{t+1}|\epsilon_t, \dots, \epsilon_0] = 0$, $t \in \mathbb{N}$, describing the unobservable perturbations which affect the system (3.1). $y^e_{t-1,t+i} \in \mathbb{R}^d$, $i = 0, \dots, n_2$, are forecasts for future realizations y_{t+i}, $i = 0, \dots, n_2$, respectively.

Remark 3.1. *A complete treatment of models of the type (3.1) seems as yet unaccomplished, cf. Evans & Honkapohja (2001, p. 173). Most models of the literature have $n_2 \leq 1$. The financial market model discussed in Chapter 7, Section 7.4 may serve as an example for $n_2 = 1$. This model is generalized in Hillebrand & Wenzelburger (2006a) to a financial market model with traders whose planning horizon is longer than two periods and which is described by an economic law of the form (3.1) with arbitrarily large $n_2 > 1$.*

The model (3.1) provides a quite general specification of a linear economic system, where for the sake of simplicity we abstract from lagged forecasts made at previous dates as well as from contemporaneous forecasts. For the purposes of this chapter, it is convenient to reformulate the model (3.1) as follows. To this end we adjust the notation of Chapter 2 to the particular needs of this chapter and use '$\top$' to distinguish between row and column vectors. This also facilitates further reading of the related literature. For each $t \in \mathbb{N}$, set

$$\begin{aligned} Y_t &:= (y_t^\top, \ldots, y_{t-n_1+1}^\top)^\top, & Y_t^e &:= (y_{t,t+1+n_2}^{e\top}, \ldots, y_{t,t+1}^{e\top})^\top \\ W_t &:= (w_t^\top, \ldots, w_{t-n_3+1}^\top)^\top, & E_t &:= (\epsilon_t^\top, \ldots, \epsilon_{t-n_4+1}^\top)^\top \end{aligned}$$

and

$$x_t = (Y_t^\top, W_t^\top, E_t^\top)^\top. \tag{3.2}$$

Let I_d denote the identity matrix on $\mathbb{R}^d$ and define block matrices associated with (3.1) as

$$\begin{aligned} \mathbf{A} &:= \begin{pmatrix} A^{(1)} & \cdots & \cdots & A^{(n_1)} \\ I_d & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I_d & 0 \end{pmatrix}, & \mathbf{B} &:= \begin{pmatrix} B^{(0)} & \cdots & B^{(n_2)} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}, \\ \mathbf{C} &:= \begin{pmatrix} C^{(1)} & \cdots & C^{(n_3)} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}, & \mathbf{D} &:= \begin{pmatrix} D^{(1)} & \cdots & D^{(n_4)} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}. \end{aligned} \tag{3.3}$$

Let $\mathrm{pr}_{-1}^n : \mathbb{R}^{dn} \to \mathbb{R}^{dn}$, $n \in \mathbb{N}$, be a projection mapping, given by

$$pr_{-1}^n := \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ I_d & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I_d & 0 \end{pmatrix}, \quad n \in \mathbb{N}.$$

A linear economic law with expectational leads is an (affine-)linear map $G : \mathbb{R}^{2d} \times \mathbb{R}^{d(n_1+n_3+n_4)} \times \mathbb{R}^{d(n_2+1)} \longrightarrow \mathbb{R}^{d(n_1+n_3+n_4)}$, defined by

$$\begin{aligned} x_{t+1} &= G(\xi_{t+1}, x_t, Y_t^e) \\ &:= \begin{pmatrix} \mathbf{A} & \mathbf{C} & \mathbf{D} \\ 0 & \mathrm{pr}_{-1}^{n_3} & 0 \\ 0 & 0 & \mathrm{pr}_{-1}^{n_4} \end{pmatrix} x_t + \begin{pmatrix} \mathbf{B}Y_t^e \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} (\epsilon_{t+1}^\top + b^\top, 0, \ldots, 0)^\top \\ (w_{t+1}^\top, 0, \ldots, 0)^\top \\ (\epsilon_{t+1}^\top, 0, \ldots, 0)^\top \end{pmatrix}, \end{aligned} \tag{3.4}$$

with $\xi_{t+1} = (\epsilon_{t+1}^\top, w_{t+1}^\top)^\top \in \mathbb{R}^{2d}$.

The linear specification of the economic law (3.1) suggests to restrict the analysis of forecasting rules to (affine-) linear functions $\Psi : \mathbb{R}^{d(n_1+n_3+n_4)} \times \mathbb{R}^{d(n_2+1)} \longrightarrow \mathbb{R}^{d(n_2+1)}$, given by

$$Y_t^e := \Psi(x_t, Y_{t-1}^e), \tag{3.5}$$

such that Y_t^e is the vector forecasts formed at date $t \in \mathbb{N}$. It will turn out that forecasting rules which generate rational expectations equilibria (REE) of the model (3.1) are of the functional form (3.5). In order to actually apply (3.5), we will assume for the moment that, contrary to the assumptions stated at the outset, $\{\epsilon_s\}_{s=0}^t$ are observable in period t. When introducing our learning scheme in Section 3.3 below, we will replace the $\{\epsilon_t\}_{t\in\mathbb{N}}$ with appropriate estimates.

Given the economic law G and a forecasting rule Ψ, the future state of the economy is determined by the *time-one map*

$$G_\Psi(\xi_{t+1}, \cdot) : \mathbb{R}^{d(n_1+n_3+n_4)} \times \mathbb{R}^{d(n_2+1)} \longrightarrow \mathbb{R}^{d(n_1+n_3+n_4)} \times \mathbb{R}^{d(n_2+1)}, \tag{3.6}$$

$\xi_{t+1} \in \mathbb{R}^{2d}$, defined by

$$\begin{pmatrix} x_t \\ Y_{t-1}^e \end{pmatrix} \mapsto \begin{pmatrix} G\big(\xi_{t+1}, x_t, \Psi(x_t, Y_{t-1}^e)\big) \\ \Psi\big(x_t, Y_{t-1}^e\big) \end{pmatrix}.$$

In other words, if x_t is the (lagged) state of the economy at date t, Y_{t-1}^e are the previous forecasts and ξ_{t+1} is the future realization of the exogenous shock,

$$\begin{pmatrix} x_{t+1} \\ Y_t^e \end{pmatrix} = G_\Psi(\xi_{t+1}, x_t, Y_{t-1}^e)$$

describes the state x_{t+1} of the economy at date $t+1$ together with all relevant forecasts Y_t^e. The map (3.6) thus drives the evolution of the economy. The system (3.6) can also be seen as a state-space representation of an ARMAX process, cf. Hannan & Deistler (1988). Other forecasting rules, in particular those which depend explicitly on time, can be implemented in a similar manner. This, in fact, will be done when introducing our learning scheme.

Before we address the problem of learning REE from historical data, we state conditions under which the system (3.4) is stable in the sense that the growth of the realizations y_t is $\mathbb{P}$-a.s. bounded. This will become important when introducing our learning scheme because it guarantees dynamic stability of the economic system. Recall the following notation. 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. By $a_t = o(b_t)$ $\mathbb{P}$-a.s., we mean that $\lim_{t\to\infty} \left|\frac{a_t}{b_t}\right| = 0$ $\mathbb{P}$-a.s.

Proposition 3.1. *Consider the economic law (3.1) and let*

$$X_t = (Y_t^\top, Y_t^{e\top}, W_t^\top, E_t^\top)^\top \tag{3.7}$$

Assume that the following hypotheses are satisfied:

(i) All eigenvalues λ_i of $\mathbf{A}$ given in (3.3) lie on or within the unit circle, that is, $|\lambda_i| \leq 1$.

(ii) The two exogenous sequences $\{w_t\}_{t\in\mathbb{N}}$ *and* $\{\epsilon\}_{t\in\mathbb{N}}$ *satisfy* $\|w_t\| = o(t^\beta)$ *and* $\|\epsilon\| = o(t^\beta)$ $\mathbb{P}$*-a.s. for some real number* β.

If, in addition, the forecasts fulfill $\mathbb{P}$*-a.s.*

$$\|y^e_{t,t+i}\| = o(t^\beta) \quad \textit{for all} \;\; j = 1, \dots, n_2 + 1, \tag{3.8}$$

then the system (3.1) is stable in the sense that there exists $\alpha > 0$ *such that*

$$\|X_t\| = O(t^\alpha) \quad \mathbb{P} - a.s.$$

Corollary 3.1. *Under the hypotheses of Proposition 3.1, if all eigenvalues* λ_i *of* $\mathbf{A}$ *given in (3.3) lie within the unit circle, i.e.* $|\lambda_i| < 1$*, then*

$$\|X_t\| = o(t^\beta) \quad \mathbb{P} - a.s.$$

The proofs of Proposition 3.1 and Corollary 3.1 are given in Section 3.5.1 below. Observe that Proposition 3.1 makes no assumption on how the forecasts have been obtained. In particular, the assumptions of Proposition 3.1 do not rule out that (3.1) admits explosive REE. The major implication of Proposition 3.1 is that a forecasting agency controlling the forecasts $y^e_{t,t+j}$, which are actually inserted into the economic law (3.4), has to ensure that these are bounded in the above sense in order to keep the economic system dynamically stable. If Condition (i) were violated, stability cannot be ensured without a-priori knowledge on the parameter matrices $\mathbf{A}$ and $\mathbf{B}$.

Remark 3.2. *We briefly comment on the case when Condition (i) of Proposition 3.1 on the eigenvalues of* $\mathbf{A}$ *is violated. Using the notation introduced above, an affine-linear forecasting rule (3.5) is of the form*

$$Y^e_t = \Psi(Y_t, W_t, E_t, Y^e_{t-1}) := \mathbf{K}\,Y_t + \mathbf{L}\,Y^e_{t-1} + \mathbf{M}\,W_t + \mathbf{N}\,E_t + c,$$

where $\mathbf{K}$, $\mathbf{L}$, $\mathbf{M}$, *and* $\mathbf{N}$ *are non-random matrices of respective dimensions and* $c \in \mathbb{R}^{d(n_2+1)}$ *is a non-random vector. Summarizing the exogenous processes acting on the system in*

$$\begin{aligned} \eta_{t+1} &:= \mathbf{C}W_t + \mathbf{D}E_t + \epsilon_{t+1} + b, \\ \zeta_t &:= \mathbf{M}W_t + \mathbf{N}E_t + c, \end{aligned}$$

the time-one map (3.6) takes the form

$$\begin{cases} Y_{t+1} := \mathbf{A}Y_t + \mathbf{B}Y^e_t + \eta_{t+1} \\ Y^e_t \;\;\; := \mathbf{K}Y_t + \mathbf{L}Y^e_{t-1} + \zeta_t \end{cases}$$

or, in block-matrix representation,

$$\begin{pmatrix} Y_{t+1} \\ Y^e_t \end{pmatrix} = \begin{pmatrix} \mathbf{A} + \mathbf{BK} & \mathbf{BL} \\ \mathbf{K} & \mathbf{L} \end{pmatrix} \begin{pmatrix} Y_t \\ Y^e_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{B}\zeta_t + \eta_{t+1} \\ \zeta_t \end{pmatrix}. \tag{3.9}$$

In view of Proposition 3.1, the major requirement for the stability of (3.9) is that all eigenvalues of the coefficient matrix

$$\begin{pmatrix} \mathbf{A}+\mathbf{BK} & \mathbf{BL} \\ \mathbf{K} & \mathbf{L} \end{pmatrix} \tag{3.10}$$

lie within the unit circle. Given $\mathbf{A}$ *and* $\mathbf{B}$*, it is well known from control theory that* $\mathbf{K}$ *exists such that all eigenvalues of* $\mathbf{A}+\mathbf{BK}$ *lie within the unit circle if the pair* $(\mathbf{A},\mathbf{B})$ *is controllable, cf. Simon & Mitter (1968, Thm. 5.3, p. 325). Choosing* $\mathbf{L}=0$*, we see that in this case the realizations* Y_t *can be made bounded, whereas the forecasts* Y_t^e *are possibly unbounded. Except for the scalar case, choosing* $\mathbf{K}$ *and* $\mathbf{L}$ *such that (3.10) satisfies the eigenvalue condition is significantly more difficult. In any case, however, without a-priori knowledge on* $\mathbf{A}$ *and* $\mathbf{B}$*, it is hardly possible to choose* $\mathbf{K}$ *and* $\mathbf{L}$ *so that the system (3.9) is stable. For this reason the conditions of Proposition 3.1 seem to be indispensable for adaptive learning schemes.*

Remark 3.3. *In view of Chapter 4 we could assume that, in addition, the two processes* $\{w_t\}_{t\in\mathbb{N}}$ *and* $\{\epsilon_t\}_{t\in\mathbb{N}}$ *admit a representation by a so-called* ergodic metric dynamical system *which we will introduce there. Then the time-one map (3.6) becomes an affine random dynamical system in the sense of Arnold (1998).*

3.2 Unbiased Forecasting Rules

In this section we will illustrate the results of Chapter 2 by introducing an (affine-)linear unbiased forecasting rule designed to generate rational expectations along all orbits of the linear system (3.1) or, in other words, rational expectations equilibria (REE). We will begin with linear no-updating rules and then discuss linear MSV predictors.

To derive an unbiased linear forecasting rule, we briefly review results from Chapter 2. Recall that in each period t there are n_2 forecasts $y^e_{t-i,t+1}$, $i=1,\dots,n_2$, for y_{t+1} which have been set at dates prior to t. Denote by $\mathbb{E}_j[\cdot]=\mathbb{E}[\cdot|\mathcal{F}_j]$ the conditional expectations operator. The forecast errors for y_{t+1} conditional on information available at the respective dates vanish if and only if

$$\mathbb{E}_{t-i}[y_{t+1}-y^e_{t-i,t+1}]=0 \quad \mathbb{P}-\text{a.s.}, \quad i=0,\dots,n_2. \tag{3.11}$$

Condition (3.11) and the law of iterated expectations, i.e. $\mathbb{E}_{t-i}[\mathbb{E}_t[y_{t+1}]]=\mathbb{E}_{t-i}[y_{t+1}]$, $i\geq 0$, implies that all forecasts $y^e_{t-n_2,t+1},\dots,y^e_{t,t+1}$ for y_{t+1} have to satisfy the *consistency conditions*

$$\mathbb{E}_{t-i}[y^e_{t,t+1}]=y^e_{t-i,t+1} \quad \mathbb{P}-\text{a.s.}, \quad i=1,\dots,n_2. \tag{3.12}$$

On the other hand, if all forecasts $y^e_{t-n_2,t+1},\dots,y^e_{t,t+1}$ obey (3.12), then the law of iterated expectations implies that (3.11) is automatically satisfied, provided that the most recent forecast $y^e_{t,t+1}$ has a vanishing conditional forecast error, that is, $\mathbb{E}_t[y_{t+1}] - y^e_{t,t+1} = 0$ $\mathbb{P}$-a.s. Since (3.12) has to hold for all times t and analogously for all forecasts $y^e_{t,t+j}$, we can transform (3.12) to consistency conditions for the first n_2 forecasts $y^e_{t,t+1}, \dots, y^e_{t,t+n_2}$ formed at date t. These take the form

$$\mathbb{E}_{t-1}[y^e_{t,t+i}] - y^e_{t-1,t+i} = 0 \quad \mathbb{P} - \text{a.s.}, \quad i = 1, \dots, n_2, \tag{3.13}$$

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

In terms of linear forecasting rules $\Psi = (\Psi^{(1)}, \dots, \Psi^{(n_2+1)})$ of the form (3.5), we have $\Psi^{(i)} : \mathbb{R}^{d(n_1+n_3+n_4)} \times \mathbb{R}^{d(n_2+1)} \to \mathbb{R}^d$ with

$$y^e_{t,t+i} := \Psi^{(i)}(x_t, Y^e_{t-1}) \tag{3.14}$$

for each $i = 1, \dots, n_2+1$. In view of Definition 2.1, Chapter 2, the consistency conditions (3.13) in terms of the linear forecasting rules (3.14) state that for each $i = 1, \dots, n_2$,

$$\mathbb{E}_{t-1}[\Psi^{(i)}(x_t, Y^e_{t-1})] = y^e_{t-1,t+i} \quad \mathbb{P} - \text{a.s.} \tag{3.15}$$

for all (x_t, Y^e_{t-1}) and all periods t. As before, Condition (3.15) provides a minimum requirement a forecasting rule has to meet in order to generate rational expectations, since otherwise (3.11) cannot be satisfied for all forecasts.

To see which forecasting rules generate rational expectations, we first compute the expected value of the future state y_t conditional on information available at date $t-1$ which is

$$\mathbb{E}_{t-1}[y_t] = \sum_{i=1}^{n_1} A^{(i)}\, y_{t-i} + \sum_{i=0}^{n_2} B^{(n_2-i)}\, y^e_{t-1,t+i} + \sum_{i=1}^{n_3} C^{(i)} w_{t-i} + \sum_{i=1}^{n_4} D^{(i)} \epsilon_{t-i} + b.$$

The forecast error $\mathbb{E}_{t-1}[y_t] - y^e_{t-1,t}$ conditional on information available at date $t-1$ is then obtained from the *(mean) error function* associated with the economic law (3.1) which is given by a function

$$\mathcal{E}_G : \mathbb{R}^{d(n_1+n_3+n_4)} \times \mathbb{R}^{d(n_2+1)} \to \mathbb{R}^d,$$

defined by

$$\begin{aligned} \mathcal{E}_G(x_{t-1}, Y^e_{t-1}) := \sum_{i=1}^{n_1} A^{(i)}\, y_{t-i} + \sum_{i=0}^{n_2} B^{(n_2-i)}\, y^e_{t-1,t+i} - y^e_{t-1,t} \\ + \sum_{i=1}^{n_3} C^{(i)} w_{t-i} + \sum_{i=1}^{n_4} D^{(i)}\, \epsilon_{t-i} + b, \end{aligned} \tag{3.16}$$

where

$$x_{t-1} = (y^\top_{t-1}, \dots, y^\top_{t-n_1},\, w^\top_{t-1}, \dots, w^\top_{t-n_3},\, \epsilon^\top_{t-1}, \dots, \epsilon^\top_{t-n_4})^\top$$

and

$$Y^e_{t-1} = (y^{e\top}_{t-1,t+n_2}, \ldots, y^{e\top}_{t-1,t})^\top.$$

Given an arbitrary state of the economy x_{t-1}, the error function describes all possible mean forecast errors, regardless of which forecasting rule has been applied. Geometrically, the graph of the error function (3.16) is a hyper-plane. Given arbitrary forecasts Y^e_{t-1}, the conditional forecast error for $y^e_{t-1,t}$ vanishes if and only if

$$\mathbb{E}_{t-1}[y_t] - y^e_{t-1,t} = \mathcal{E}_G(x_{t-1}, Y^e_{t-1}) = 0 \quad \mathbb{P} - \text{a.s.} \tag{3.17}$$

In view of Definition 2.3, Chapter 2, a forecasting rule Ψ which for given (x_{t-1}, Y^e_{t-2}) sets $Y^e_{t-1} = \Psi(x_{t-1}, Y^e_{t-2})$ such that (3.17) holds is called unbiased at (x_{t-1}, Y^e_{t-2}). A forecasting rule which is unbiased along orbits can now be redefined as follows. (Compare with Definition 2.4, Chapter 2.)

Definition 3.1. *A consistent forecasting rule* $\Psi_\star = (\Psi_\star^{(1)}, \ldots, \Psi_\star^{(n_2+1)})$ *is called* unbiased *if there exists a non-empty subset* $\mathbb{U} \subset \mathbb{R}^{d(n_1+n_2+n_3+n_4+1)}$ *such that the following conditions are satisfied:*

(i) Unbiasedness. $\Psi_\star$ *satisfies*

$$\mathcal{E}_G\big(x, \Psi_\star(x, Y^e)\big) = 0 \quad \textit{for all } (x, Y^e) \in \mathbb{U}.$$

(ii) Invariance. $\mathbb{U}$ *is invariant under the map* $G_{\Psi_\star}$ *defined in (3.6), that is, for each* $\xi \in \mathbb{R}^{2d}$,

$$G_{\Psi_\star}(\xi, x, Y^e) \in \mathbb{U} \quad \textit{for all } (x, Y^e) \in \mathbb{U}.$$

Condition (i) describes states (x, Y^e) in which the forecast error (3.17) vanishes and Condition (ii) implies that (3.17) can be satisfied for all times t. If $\Psi_\star$ is unbiased in the sense of Definition 3.1, then all orbits $\gamma(x_1, Y^e_0) := \{x_t, Y^e_{t-1}\}_{t\in\mathbb{N}}$ generated by the time-one map $G_{\Psi_\star}$, where initial conditions $(x_1, Y^e_0) \in \mathbb{U}$ may be arbitrary, correspond precisely to rational expectations equilibria (REE).

3.2.1 No-Updating Rules

Observe that the error function (3.16) is globally invertible if $B^{(0)}$ is invertible. We can now exploit the linear structure of (3.16) and construct a linear unbiased forecasting rule as follows. Suppose $B^{(0)}$ is invertible. We may then solve (3.17) by setting

$$y^e_{t-1,t-1+i} = y^e_{t-2,t-1+i}, \quad i = 1, \ldots, n_2$$

and solving for $y^e_{t-1,t+n_2}$ to get

$$
\begin{aligned}
y^e_{t-1,n_2+1} &= \Psi_\star^{(n_2+1)}(x_{t-1}, Y^e_{t-2}) \\
&:= -(B^{(0)})^{-1}\Bigg[\sum_{i=1}^{n_1} A^{(i)} y_{t-i} + \sum_{i=0}^{n_2-1} B^{(n_2-i)} y^e_{t-2,t+i} - y^e_{t-2,t} \\
&\qquad + \sum_{i=1}^{n_3} C^{(i)} w_{t-i} + \sum_{i=1}^{n_4} D^{(i)}\,\epsilon_{t-i} + b\Bigg],
\end{aligned}
\tag{3.18}
$$

where

$$
x_{t-1} = (y_{t-1}^\top, \ldots, y_{t-n_1}^\top,\, w_{t-1}^\top, \ldots, w_{t-n_3}^\top,\, \epsilon_{t-1}^\top, \ldots, \epsilon_{t-n_4}^\top)^\top
$$

as before.The function (3.18) induces a linear forecasting rule

$$
\Psi_\star = (\Psi_\star^{(1)}, \ldots, \Psi_\star^{(n_2+1)})
$$

of the functional form (3.5) by setting

$$
\begin{cases}
y^e_{t-1,t-1+i} = \Psi_\star^{(i)}(x_{t-1}, Y^e_{t-2}) := y^e_{t-2,t-1+i}, \quad i = 1, \ldots, n_2, \\
y^e_{t-1,t+n_2} \;\;= \Psi_\star^{(n_2+1)}(x_{t-1}, Y^e_{t-2})
\end{cases}
\tag{3.19}
$$

for an arbitrary state x_{t-1} and arbitrary forecasts Y^e_{t-2}. As (3.19) never updates forecasts made in the previous period, it will be referred to as a (linear) *no-updating forecasting rule*. By construction,

$$
\mathcal{E}_G\big(x, \Psi_\star(x, Y^e)\big) = 0 \quad \text{identically on} \quad \mathbb{R}^{d(n_1+n_2+n_3+n_4+1)}
$$

and since the forecasting rule (3.19) clearly fulfills the consistency conditions (3.15), Condition (i) of Definition 3.1 is satisfied. Inserting (3.19) into the economic law (3.1), the associated time-one map (3.6) takes the form

$$
\begin{cases}
x_t \;= \begin{pmatrix} \mathrm{pr}_{-1}^{n_1} & 0 & 0 \\ 0 & \mathrm{pr}_{-1}^{n_3} & 0 \\ 0 & 0 & \mathrm{pr}_{-1}^{n_4} \end{pmatrix} x_{t-1} + \begin{pmatrix} (\mathrm{pr}_1 Y^e_{t-2})^\top \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \big((\epsilon_t + b)^\top, 0\big)^\top \\ (w_t^\top, 0)^\top \\ (\epsilon_t^\top, 0)^\top \end{pmatrix} \\
\\
Y^e_{t-1} = \Psi_\star(x_{t-1}, Y^e_{t-2})
\end{cases}
\tag{3.20}
$$

with $\mathrm{pr}_1 Y^e_{t-2} = y^e_{t-2,t}$. Since (3.20) is defined globally on $\mathbb{R}^{d(n_1+n_2+n_3+n_4+1)}$, the forecasting rule (3.19) fulfills Condition (ii) of Definition 4.4, so that it is unbiased. It is easy to verify that all conditional forecast errors generated by (3.20) vanish on average, since by construction $\mathbb{E}_{t-1}[y_t] = y^e_{t-1,t}$ and hence all forecasts for y_t satisfy

$$
\mathbb{E}_{t-1}[y_t] = y^e_{t-i,t} \quad \mathbb{P}-\text{a.s.}, \quad i = 1, \ldots, n_2 + 1,
\tag{3.21}
$$

as long as the no-updating rule (3.19) has been applied for the past $n_2 + 1$ periods. Since (3.19) is defined globally on $\mathbb{R}^{d(n_1+n_2+n_3+n_4+1)}$, it is globally unbiased in the sense of Definition 3.1 implying that all orbits of (3.20) correspond to REE in the classical sense. The function (3.19) will henceforth be referred to as an *unbiased no-updating* forecasting rule.

The forecasting rule (3.19) is constructed in such a way that $\mathbb{E}_t[y_{t+1}] = y^e_{t,t+1}$ for all choices of $y^e_{t,t+i}$, $i = 1, \ldots, n_2$. 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. This freedom constitutes a primary cause for the emergence of multiple unbiased forecasting rules and thus of REE in the presence of expectational leads.

Remark 3.4. *Confirming Theorem 2.1 of Chapter 2, the remarkable phenomenon in economic systems with expectational leads ($n_2 > 0$) and invertible $B^{(0)}$ is that the unbiased no-updating forecasting rule (3.19) generates rational expectations in the sense that* all *forecast errors (3.17) conditional on information available at date t vanish, including those which were made prior to that date. It yields the most precise forecasts in the sense that forecast errors vanish conditional on information which is not available at the stage in which they were issued. This fact holds only for forecasts which feed back into the system in a non-trivial manner.*

We note in passing that the no-updating rule (3.19) is closely related to the classical Åström-Wittenmark self-tuning regulator (Åström & Wittenmark 1973), known from the literature on adaptive control and optimal tracking, cf. Caines (1988).

There are well-known examples in which $B^{(0)}$ is not invertible, see Evans & Honkapohja (2001, Chap. 10). In these cases (3.19) does not exist and the construction of an unbiased forecasting rule may become quite involved. One way out of this problem is presented in Section 3.2.2 below. There we discuss linear *minimal-state-variable predictors* (MSV predictors) which generate minimal-state-variable solutions in the sense of McCallum (1983). Another way out of the problem with singular $B^{(0)}$ is outlined in Example 3.1. Here we exploit the fact that the first condition in Definition 3.1 is a local invertibility property of the error function.

Example 3.1. *We present a particularly easy example from Wenzelburger (2006) in which $B^{(0)}$ is singular, indicating to what extent an unbiased forecasting rule will exist if the invertibility condition of the coefficient matrix $B^{(0)}$ is violated. This example provides a complementary approach to the one of McCallum (1998) and may easily be extended to the general case. With the notation of Section 3.1, consider the simple economic law*

$$y_t = By^e_{t-1,t+1} + \epsilon_t, \tag{3.22}$$

where $B \equiv B^{(0)}$ is a singular $d \times d$ matrix such that the inverse of B does not exist. An example similar to (3.22) but with different dating is found in

McCallum (1998). We may find a ‘coordinate transformation’ $z = Qy$ *with a nonsingular* $d \times d$ *matrix* Q*, such that*

$$Q \cdot B \cdot Q^{-1} = \begin{pmatrix} \widetilde{B}^{(11)} & \widetilde{B}^{(12)} \\ 0 & 0 \end{pmatrix}$$

with a non-singular $d_1 \times d_1$ *matrix* $\widetilde{B}^{(11)}$ *and some* $d_1 \times d_2$ *matrix* $\widetilde{B}^{(12)}$*, where* $d_1 + d_2 = d$*. Such a coordinate transformation* Q *would exist, for example, if* B *were diagonalizable. Set*

$$z_t = Qy_t, \quad z^e_{t-1,t+1} = Qy^e_{t-1,t+1}, \text{ and } \eta_t = Q\epsilon_t. \tag{3.23}$$

Then (3.22) can be transformed into

$$\begin{pmatrix} z_t^{(1)} \\ z_t^{(2)} \end{pmatrix} = \begin{pmatrix} \widetilde{B}^{(11)} & \widetilde{B}^{(12)} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z^{(e1)}_{t-1,t+1} \\ z^{(e2)}_{t-1,t+1} \end{pmatrix} + \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix}, \tag{3.24}$$

where

$$z^e_{t-1,t+1} = \begin{pmatrix} z^{(e1)}_{t-1,t+1} \\ z^{(e2)}_{t-1,t+1} \end{pmatrix}.$$

Thus the economic law (3.22) decouples into two (vector) equations (3.24), the second of which receives no feedback from the forecasts. Since $z_t^{(2)} = \eta_t^{(2)}$*, we set* $z^{(e2)}_{t-1,t+1} = \mathbb{E}_{t-1}[\eta^{(2)}_{t+1}]$ *to obtain an unbiased forecast for* $z^{(2)}_{t+1}$*. Since* $\widetilde{B}^{(11)}$ *is invertible, we can construct an unbiased no-updating forecasting rule for the first component analogously to (3.19). An unbiased forecasting rule for (3.24) then is*

$$\begin{cases} z^{(e1)}_{t-1,t+1} = (\widetilde{B}^{(11)})^{-1}\left[z^{(e1)}_{t-2,t} - \widetilde{B}^{(12)}\mathbb{E}_{t-1}[\eta^{(2)}_{t+1}] - \mathbb{E}_{t-1}[\eta^{(1)}_t]\right] \\ z^{(e2)}_{t-1,t+1} = \mathbb{E}_{t-1}[\eta^{(2)}_{t+1}] \end{cases}. \tag{3.25}$$

The unbiasedness is verified by inserting (3.25) into (3.24) and taking conditional expectations, so that

$$\mathbb{E}_{t-1}[z_t^{(1)} - z^{(e1)}_{t-2,t}] = 0 \quad \text{and} \quad \mathbb{E}_{t-2}[z_t - z^{(2e)}_{t-2,t}] = 0 \quad \mathbb{P}-a.s.$$

Observe that $z^{(e1)}_{t-2,t}$ *is a best least-squares prediction for* $z_t^{(1)}$ *with respect to* $\mathcal{F}_{t-1}$ *and hence with respect to* $\mathcal{F}_{t-2}$*, whereas, in general,* $\mathbb{E}_{t-1}[z_t - z^{(2e)}_{t-2,t}] \neq 0$*. Setting*

$$y^e_{t-1,t+1} = Q^{-1}z^e_{t-1,t+1}, \tag{3.26}$$

we thus obtain an unbiased forecasting rule in the sense of Definition 3.1 for the original economic law (3.22). Indeed, it follows from (3.23) that the forecast errors for (3.22) are

$$y_t - y^e_{t-2,t} = Q^{-1}[z_t - z^e_{t-2,t}] \quad \mathbb{P}-a.s.,$$

implying that the forecast errors conditional on $\mathcal{F}_{t-2}$ vanish. Taking conditional expectations with respect to $\mathcal{F}_{t-1}$, we see in addition that even

$$\mathbb{E}_{t-1}[y_t^{(1)} - y_{t-2,t}^{(1e)}] = 0, \tag{3.27}$$

provided that $\mathbb{E}_{t-1}[\eta_t^{(2)}] - \mathbb{E}_{t-2}[\eta_t^{(2)}] = 0$. The higher precision (3.27) holds, for instance, if $\{\eta_t\}_{t\in\mathbb{N}}$ is a martingale difference sequence.

3.2.2 Linear MSV Predictors

In the literature on rational expectations, the notion of minimal-state-variable solutions is a well-established concept to describe situations in which agents have rational expectations along particular orbits of the system. For linear models with expectational leads of order 2 this concept was developed in McCallum (1983, 1998, 1999). A univariate linear example with leads of order 3 is found in Evans & Honkapohja (2001). We relate this concept to our notion of a consistent forecasting rule by introducing *minimal-state-variable predictors* (MSV predictors) for linear models with 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 $n_2 > 0$ be of arbitrary size, and restrict the following analysis to the white-noise case $n_4 = 0$ with an AR(1) process $\{w_t\}_{t\in\mathbb{N}}$ and a martingale difference sequence $\{\epsilon_t\}_{t\in\mathbb{N}}$. Consider a *perceived law of motion* of the form

$$\begin{cases} y_{t+1} = & \hat{A}y_t + \hat{C}w_t + \hat{b} + \epsilon_{t+1}, \\ w_{t+1} = & \Pi w_t + \eta_{t+1}, \end{cases} \tag{3.28}$$

on the basis of which a forecasting agency forms expectations. $y_t \in \mathbb{R}^d$ and $w_t \in \mathbb{R}^d$ denote vectors of endogenous and exogenous variables, respectively, $\hat{A}$, $\hat{C}$, and Π are non-random matrices of respective dimensions, and $\hat{b} \in \mathbb{R}^d$ 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 $(y_t, w_t) \in \mathbb{R}^d \times \mathbb{R}^d$ describes the state of the economy at date t, the perceived state at date $t + j$ derived from the model (3.28) is

$$\begin{cases} y_{t+j} = & \hat{A}^j y_t + \sum_{i=1}^{j} \hat{A}^{j-i}[\hat{C}w_{t+i-1} + \hat{b} + \epsilon_{t+i}], \\ w_{t+j} = & \Pi^j w_t + \sum_{i=1}^{j} \Pi^{j-i}\eta_{t+i}. \end{cases} \tag{3.29}$$

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 $\Phi = (\Phi^{(1)}, \dots, \Phi^{(n_2+1)})$ by setting

$$y_{t,t+j}^e = \Phi^{(j)}(y_t, w_t) := \hat{A}^j y_t + \sum_{i=1}^{j} \hat{A}^{j-i}[\hat{C}\Pi^{i-1}w_t + \hat{b}] \tag{3.30}$$

for all $j = 1, \ldots, n_2 + 1$. Since the forecasting rule Φ defined in (3.30) uses a minimal amount of endogenous and exogenous variables, it will be referred to as *minimal-state-variable* predictor (MSV predictor). By construction, one has $y^e_{t,t+j} = \mathbb{E}_t[y_{t+j}]$ if y_{t+j} is thought of being generated from (3.29), and a straightforward calculation shows that each forecast (3.30) satisfies

$$\mathbb{E}_{t-1}[y^e_{t,t+j}] = \mathbb{E}_{t-1}[\Phi^{(j)}(y_t, w_t)] = \Phi^{(j+1)}(w_{t-1}, y_{t-1}) = y^e_{t-1,t+j}$$

for all $j = 1, \ldots, n_2$. This implies that a MSV predictor is a consistent forecasting rule, since it satisfies Condition (3.15). Observe that the MSV predictor associated with (3.28) can be represented by the quadruple $(\hat{A}, \hat{C}, \hat{b}, \Pi)$.

In order to construct a MSV predictor $\Phi = (\Phi^{(1)}, \ldots, \Phi^{(n_2+1)})$ which is unbiased in the sense of Definition 3.1, let

$$\Gamma_\Phi := \left\{ (y, w, z^{(1)}, \ldots, z^{(n_2+1)}) \in \mathbb{R}^{d(n_2+3)} \middle| z^{(j)} = \Phi^{(j)}(y, w),\ 1 \leq j \leq n_2 + 1 \right\}$$

be the graph of the MSV predictor $\Phi = (\Phi^{(1)}, \ldots, \Phi^{(n_2+1)})$. It is easy to see that Γ_Φ is invariant under the corresponding time-one map G_Φ. Hence Φ satisfies Property (ii) of Definition 3.1. A MSV predictor Φ is unbiased in the sense of Definition 3.1 if

$$\mathcal{E}_G\big(y, w, \Phi^{(1)}(y, w), \ldots, \Phi^{(n_2+1)}(y, w)\big) = 0 \tag{3.31}$$

for all $(y, w) \in \mathbb{R}^d \times \mathbb{R}^d$, where $\mathcal{E}_G$ denotes the error function associated with the economic law (3.1) (with $n_1 = n_3 = 1$ and $n_4 = 0$).

Inserting (3.30) into (3.31), a MSV predictor Φ is unbiased if and only if the unknown $d \times d$ coefficient matrices $\hat{A}$ and $\hat{C}$, and $\hat{b} \in \mathbb{R}^d$ in (3.30) satisfy

$$0 = A^{(1)} + [B^{(n_2)} - I_d]\hat{A} + \sum_{i=0}^{n_2-1} B^{(n_2-i)} \hat{A}^{i+1}, \tag{3.32}$$

$$0 = \sum_{i=0}^{n_2-1} \sum_{j=1}^{i+1} B^{(n_2-i)} \hat{A}^{i+1-j} \hat{C} \Pi^{j-1} + [B^{(n_2)} - I_d]\hat{C} + C^{(1)}, \tag{3.33}$$

$$0 = \left(\sum_{i=0}^{n_2-1} \sum_{j=1}^{i+1} B^{(n_2-i)} \hat{A}^{i+1-j} + [B^{(n_2)} - I_d] \right) \hat{b} + b. \tag{3.34}$$

Summarizing, this yields the following existence theorem for unbiased MSV predictors.

Theorem 3.1. *Suppose that $n_1 = n_3 = 1$, $n_4 = 0$, and that the exogenous observable noise $\{w_t\}_{t \in \mathbb{N}}$ is an AR(1) process of the form*

$$w_t = \Pi w_{t-1} + \eta_t,$$

where Π is a non-random $d \times d$ matrix whose eigenvalues lie within the unit circle and $\{\eta_t\}_{\in\mathbb{N}}$ is a martingale difference sequence. Let

$$\Phi_\star = (\Phi_\star^{(1)}, \dots, \Phi_\star^{(n_2+1)})$$

be a MSV predictor pertaining to the perceived law of motion (3.28) which is represented by the quadruple $(\hat{A}_\star, \hat{C}_\star, \hat{b}_\star, \Pi)$, where $\hat{A}_\star$ and $\hat{C}_\star$ are $d \times d$ matrices with real coefficients and $\hat{b}_\star \in \mathbb{R}^d$. Then $\Phi_\star$ is unbiased for the economic law (3.1) in the sense of Definition 3.1 if and only if the triplet $(\hat{A}_\star, \hat{C}_\star, \hat{b}_\star)$ is a solution of eqs. (3.32)-(3.34).

Eqs. (3.32)-(3.34) correspond to Eqs. (10.5)-(10.7) in Evans & Honkapohja (2001, p. 230) for expectational leads of arbitrary length. It is shown in Lancaster (1969, Thm. 8.4.1, p. 262) that (3.33) can be transformed into a linear matrix equation. Thus (3.33) and (3.34) are linear equations for $\hat{C}$ and $\hat{b}$, once $\hat{A}$ is known. It is straightforward to verify that $\hat{A}_\star$ is a solution to (3.32) 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, in general, there are multiple right divisors. 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).

If $\Phi_\star$ is an unbiased MSV predictor with coefficients, it follows from Theorem 3.1 together with (3.31) that

$$y_t = \Phi_\star^{(1)}(y_{t-1}, w_{t-1}) + \epsilon_t = G\big(\xi_t, y_{t-1}, w_{t-1}, \Phi_\star(y_{t-1}, w_{t-1})\big) \tag{3.35}$$

for all $t \in \mathbb{N}$, where G denotes the economic law with $n_1 = n_3 = 1$, $n_4 = 0$ as defined in (3.4). In other words, unbiased MSV predictors, if they exist, generate minimal-state-variable solutions to (3.1) in the sense of McCallum (1983). These correspond to the perceived law of motion (3.28) with coefficients $(\hat{A}_\star, \hat{C}_\star, \hat{b}_\star, \Pi)$. All forecasts obtained from $\Phi_\star$ are best least-squares predictions with respect to the corresponding information along any orbit of the resulting time-one map (3.6).

To compute an unbiased MSV predictor may become a difficult problem. Explicit solutions to (3.32)-(3.34) for the case $n_2 = 1$ are found in McCallum (1983, 1998, 1999). The existence of solutions to (3.32)-(3.34), in general, do not require an invertible $B^{(0)}$. However, it does requires full knowledge of the coefficient matrix Π 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 economically meaningful solution. In Example 2.2 of Chapter 2 we have already seen that unbiased MSV predictors may not exist, while unbiased no-updating rules exist. Despite this fact, however, unbiased MSV predictors are important as they may generate stable long-run behavior under rational expectations in an otherwise unstable situation.

Remark 3.5. *In some cases linear models may admit unbiased MSV predictors which are non-linear functions, cf. Froot & Obstfeld (1991). These correspond to non-linear rational-expectations solutions and will be treated in Chapter 4.*

3.3 Approximating Unbiased Forecasting Rules

We are now in a position to introduce a learning scheme for unbiased forecasting rules for the general linear system (3.1) using methods from control theory which are developed in Lai & Wei (1986b). For simplicity of exposition, we assume that the constant term in (3.1) is zero, i.e. $b = 0$. 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

$$X_{t-1}^{\top} = (Y_{t-1}^{\top}, Y_{t-1}^{e\top}, W_{t-1}^{\top}, E_{t-1}^{\top})^{\top}, \tag{3.36}$$

the linear system (3.1) may be rewritten as

$$y_t = \theta\, X_{t-1} + \epsilon_t. \tag{3.37}$$

We will use the representation (3.37) to approximate the unbiased no-updating rule $\Phi_\star$ given in (3.18) from estimates for the coefficients θ in two steps. In the first step, we estimate the unknown coefficient matrix θ from historical data. In the second step we compute an approximation of an unbiased forecasting rule. The estimates of the unknown coefficient matrix θ in (3.37) are obtained from the so-called *approximate-maximum-likelihood (AML)* algorithm, cf. Lai & Wei (1986b). This method provides a recursive scheme which generates successive estimates

$$\hat{\theta}_t = (\hat{A}_t^{(1)}, \dots, \hat{A}_t^{(n_1)},\, \hat{B}_t^{(0)}, \dots, \hat{B}_t^{(n_2)},\, \hat{C}_t^{(1)}, \dots, \hat{C}_t^{(n_3)},\, \hat{D}_t^{(1)}, \dots, \hat{D}_t^{(n_4)}) \tag{3.38}$$

for θ 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 (3.37) by

$$\begin{aligned} \hat{X}_{t-1} &:= (Y_{t-1}^{\top}, Y_{t-1}^{e\top}, W_{t-1}^{\top}, \hat{E}_{t-1}^{\top})^{\top}, \\ \hat{E}_{t-1} &:= (\hat{\epsilon}_{t-1}^{\top}, \dots, \hat{\epsilon}_{t-n_4}^{\top})^{\top}, \end{aligned} \tag{3.39}$$

where

$$\hat{\epsilon}_{t-i} := y_{t-i} - \hat{\theta}_{t-i}\hat{X}_{t-i-1}, \quad i = 1, \ldots, n_4 \tag{3.40}$$

are also called *a-posteriori* prediction errors. The AML algorithm is recursively defined by

$$\begin{cases} \hat{\theta}_t &= \hat{\theta}_{t-1} + (y_t - \hat{\theta}_{t-1}\hat{X}_{t-1})\hat{X}_{t-1}^\top P_{t-1}, \\ P_{t-1} &= P_{t-2} - P_{t-2}(\hat{X}_{t-1}\hat{X}_{t-1}^\top)P_{t-2}\,(1 + \hat{X}_{t-1}^\top P_{t-2}\hat{X}_{t-1})^{-1}, \end{cases} \tag{3.41}$$

where the initial conditions P_0 and $\hat{\theta}_0$ may be chosen arbitrarily with P_0 invertible.[1]

The parameter estimate $\hat{\theta}_t$ as given in (3.38) can now be used to estimate an unbiased no-updating rule (3.19) as follows. 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. An approximation of the unbiased no-updating rule (3.19) is given by the map

$$\begin{aligned} y^e_{t,t+j} &= y^e_{t-1,t+j}, \quad j = 1, \ldots, n_2, \\ y^e_{t,t+1+n_2} &= \hat{\psi}(\hat{x}_t, Y^e_{t-1}; \hat{\theta}_t) \\ &:= -\hat{B}_t^{(0)\#}\Big[\sum_{i=1}^{n_1} \hat{A}_t^{(i)}\, y_{t+1-i} + \sum_{i=0}^{n_2-1} \hat{B}_t^{(n_2-i)}\, y^e_{t-1,t+1+i} \\ &\qquad -y^e_{t-1,t+1} + \sum_{i=1}^{n_3} \hat{C}_t^{(i)}\, w_{t+1-i} + \sum_{i=1}^{n_4} \hat{D}_t^{(i)}\, \hat{\epsilon}_{t+1-i}\Big], \end{aligned} \tag{3.42}$$

where

$$\hat{x}_t := (Y_t^\top, W_t^\top, \hat{E}_t^\top)^\top$$

is an approximation of the vector of endogenous variables x_t defined in (3.2). Using (3.41), the forecasting rule (3.42) will then be updated over time with the updates of the parameter estimates $\hat{\theta}_t$.

Remark 3.6. *(i) By Theorem 2 of Lancaster (1969, p. 306), $x_0 := B^\# b$ minimizes $\|Bx - b\|$ in the sense that any $x \neq x_0 \in \mathbb{R}^d$ satisfies either $\|Bx - b\| > \|Bx_0 - b\|$ or $\|Bx - b\| = \|Bx_0 - b\|$ and $\|x\| > \|x_0\|$.*

(ii) If $B^{(0)}$ is known to be invertible, then an approximation of the unbiased no-updating rule could directly be obtained from regressing over (3.18) instead of (3.1), thereby avoiding the somewhat involved computation of a Moore-Penrose inverse.

(iii) Instead of (3.42), an approximation of any other unbiased forecasting rule could be taken. Observe, however, that taking an MSV predictor as introduced in Section 3.2.2 may be computationally demanding. Eqs. (3.32)-(3.34) may not admit solutions with real coefficients for approximated parameters. With the focus on global convergence of the AML algorithm, we therefore choose (3.42).

[1] By the matrix inversion lemma, $P_t^{-1} = P_{t-1}^{-1} + \hat{X}_t\hat{X}_t^\top$, cf. Caines (1988).

(iv) *A constant term $b \neq 0$ could be incorporated into the AML algorithm as follows. The simplest way is to assume that $\bar{\epsilon}$ with $\left(I_d + \sum_{i=1}^{n_4} D^{(i)}\right)\bar{\epsilon} = b$ exists. We may then set $\tilde{\epsilon}_t := \epsilon_t + \bar{\epsilon}$ and replace ϵ_t with $\tilde{\epsilon}_t$ in the definition of the a-posteriori prediction errors (3.40). Then $\hat{b}_t := \frac{1}{t+1}\sum_{i=0}^{t}\hat{\epsilon}_i$ is a period-t estimate for b. By a reasoning analogous to the proof of Theorem 3.3 below, $\hat{\epsilon}_t$ will converge $\mathbb{P}$-a.s. to ϵ_t in either case, so that $\hat{b}_t$ converges to b.*

3.4 Convergence of an AML-Based Learning Scheme

We will now investigate under which conditions a learning scheme based on the AML algorithm (3.41) and the forecasting rule (3.42) converges to rational expectations. Under a number of assumptions discussed in this section, 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 θ, independently of the initial state of the system and independently of initial parameter estimates. Lai & Wei (1986b) show that under certain standard assumptions (see Proposition 3.3 below) $\hat{\theta}_t$ converges to θ a.s. if the regressors X_t given in (3.7) satisfy the *weak persistent excitation condition*

$$\frac{\lambda_{\min}\left(\sum_{t=1}^{T} X_t X_t^{\top}\right)}{\log\left(e + \lambda_{\max}\left(\sum_{t=1}^{T} X_t X_t^{\top}\right)\right)} \to \infty \quad \mathbb{P}-\text{a.s. as } T \to \infty, \tag{3.43}$$

where $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ denote the minimum and maximum eigenvalues of a symmetric matrix B, respectively. It has long been recognized that in feedback systems of the form (3.1), (3.43) may be violated such that the parameter estimates may fail to be consistent. The main problem with (3.43) is that the regressors X_t cannot be manipulated directly. However, Lai & Wei (1986b) showed that the *weak persistent excitation condition* (3.43) can be translated into conditions on the inputs alone which in our case are the forecasts and the two exogenous variables. Following Lai & Wei (1986b), Condition (3.43) will now be replaced by more direct assumptions concerning these inputs alone. This requires a modification of the forecasting rule (3.42). The corresponding assumptions are stipulated as follows.

Assumption 3.1. *The unobservable 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 conditions.*

(i) $\sup_{t\geq 1}\mathbb{E}\left[\|\epsilon_t\|^{\alpha}\middle|\mathcal{F}_{t-1}\right] < \infty \quad \mathbb{P}-\text{a.s.}$ *for some* $\alpha > 2$.

(ii) $\|\epsilon_T\| = o(T^{\frac{1}{2}})$, $\sum_{t=1}^{T}\|\epsilon_t\|^2 = O(T)$, *and*

$$\liminf_{T\to\infty}\frac{1}{T}\lambda_{\min}\left(\sum_{t=1}^{T}\epsilon_t\epsilon_t^{\top}\right) > 0 \;\; \mathbb{P}-\text{a.s.}$$

(iii) Let $D^{(0)}, \ldots, D^{(n_4)}$, $n_4 \geq 0$, *be non-random* $d \times d$ *matrices, where* $D^{(0)} = I_d$ *is the* $d \times d$ *identity matrix. Assume that the matrix polynomial*

$$\Gamma(s) = D^{(0)} + D^{(1)}s + \cdots + D^{(n_4)}s^{n_4}, \quad s \in \mathbb{C}$$

is strictly positive real, *i.e.,*

(a) $\det \Gamma(s) \neq 0$ *for all* $s \in \mathbb{C}$ *with* $|s| \leq 1$, *and*

(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,*

$$\zeta^\top \big(\Gamma(s) + \Gamma^\top(\bar{s})\big)\zeta > 0 \quad \textit{for all} \;\; 0 \neq \zeta \in \mathbb{R}^d.$$

Assumption 3.2. *The observable 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* σ*-algebra generated by* $\mathcal{F}_{t-1}$ *and* ϵ_t, *where* ϵ_t *is defined in Assumption 3.1. 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 conditions.*

(i) $\sup_{t\geq 1} \mathbb{E}[\|\tilde{w}_t\|^\beta|\mathcal{F}'_t] < \infty \quad \mathbb{P}-\text{a.s.} \;\; \textit{for some} \;\; \beta > 2.$

(ii) $\sum_{t=1}^T \|\tilde{w}_t\|^2 = O(T)$ *and* $\liminf_{T\to\infty} \frac{1}{T}\lambda_{\min}\left(\sum_{t=1}^T \tilde{w}_t\tilde{w}_t^\top\right) > 0 \quad \mathbb{P}-\text{a.s.}$

(iii) $\|w_T\| = o(T^{\frac{1}{2}})$ *and* $\sum_{t=1}^T \|\mathbb{E}[w_t|\mathcal{F}'_t]\|^2 = O\left(\lambda_{\min}\left(\sum_{t=1}^T \tilde{w}_t\tilde{w}_t^\top\right)\right) \quad \mathbb{P}-\text{a.s.}$

The process defined in Assumption 3.1 is referred to as *colored noise.* The white-noise case in which $\Gamma(s) = I_d$ and $n_4 = 0$ assumed in most economic models satisfies Assumption 3.1. The case in which the observable exogenous process $\{w_t\}_{t\in\mathbb{N}}$ is generated by a stable AR(1) process $w_t = \Pi w_{t-1} + \eta_t$, where Π is a non-random $d \times d$ matrix whose eigenvalues lie within the unit circle and $\{\eta_t\}_{t\in\mathbb{N}}$ is white noise and hence stochastically independent of the unobservable noise $\{\epsilon_t\}_{t\in\mathbb{N}}$, is an example for a process that fulfills Assumption 3.2. Assumptions 3.1 and 3.2 are thus generalizations of popular parameterizations in the learning literature, cf. Evans & Honkapohja (2001).

Remark 3.7. *(i) Since Ljung (1977) Assumption 3.1 appears to be fundamental in the systems and control literature, cf. Caines (1988). 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 3.1 (iii) is violated.*

(ii) Assumption 3.1 contains conditions which are redundant. Condition (i) implies

$$\sum_{t=1}^{\infty} \mathbb{P}\left(\|\epsilon_t\|^\alpha > \delta t^{\alpha\alpha'} \mid \mathcal{F}_{t-1}\right) < \infty$$

for every $\alpha' > \frac{1}{\alpha}$ and each $\delta > 0$. By the conditional Borel-Cantelli Lemma (Stout 1974, Thm. 2.8.5, p. 55), $\|\epsilon_T\| = o(T^{\alpha'})$ $\mathbb{P}$-a.s. for each $\alpha' > \frac{1}{\alpha}$ and thus in particular for $\alpha' = \frac{1}{2}$. It is shown in Lai & Wei (1983) that Condition (ii) follows from Condition (i) and the weaker condition

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

Analogous remarks apply to Assumption 3.2 and to Assumption 3.3 below.

Our last assumption concerns the controllable inputs which are the forecasts. The key idea is that the forecasts are *censored* whenever the persistent excitation condition (3.43) is violated or the forecasts itself become unbounded.

Assumption 3.3. *A forecasting agency issues forecasts based on the AML algorithm (3.41) and the forecasting rule $\hat{\psi}$ defined in (3.42) as follows. Given two sequences of real numbers $\{c_t\}_{t\in\mathbb{N}}$ and $\{v_t\}_{t\in\mathbb{N}}$, the forecasts in period t are based on the current parameter estimate $\hat{\theta}_t$ and given by a map*

$$\hat{\Psi}_t = (\hat{\Psi}_t^{(1)}, \ldots, \hat{\Psi}_t^{(n_2+1)}), \tag{3.44}$$

defined by

$$\begin{aligned} y^e_{t,t+j} &= \hat{\Psi}_t^{(j)}(\hat{x}_t, Y^e_{t-1}; \hat{\theta}_t) := y^e_{t-1,t+j}, \quad j = 1, \ldots, n_2, \\ y^e_{t,t+1+n_2} &= \hat{\Psi}_t^{(n_2+1)}(\hat{x}_t, Y^e_{t-1}; \hat{\theta}_t) \\ &:= \begin{cases} \hat{\psi}(\hat{x}_t, Y^e_{t-1}; \hat{\theta}_t) + v_t & \text{if } \|\hat{\psi}(\hat{x}_t, Y^e_{t-1}; \hat{\theta}_t)\| \le c_t, \\ \dfrac{c_t}{\|\hat{\psi}(\hat{x}_t, Y^e_{t-1}, \hat{\theta}_t)\|}\, \hat{\psi}(\hat{x}_t, Y^e_{t-1}; \hat{\theta}_t) + v_t & \text{otherwise.} \end{cases} \end{aligned}$$

Here for a given tolerance level $c_{\text{tol}} > 0$, the bounds $\{c_t\}_{t\in\mathbb{N}}$ consists of a sequence of positive numbers satisfying $c_t \ge c_{\text{tol}}$ for all $t \in \mathbb{N}$, $\|c_t\| = o(t^{\frac{1}{2}})$, and $\sum_{t=1}^T c_t^2 = O(T)$. The dither $\{v_t\}_{t\in\mathbb{N}}$ is an exogenous white-noise process satisfying the following conditions.

(a) For each $t \in \mathbb{N}$, let $\mathcal{G}_t := \sigma(\mathcal{F}_{t-1}, y_t, w_t)$ denote the σ-algebra generated by $\mathcal{F}_{t-1}$, y_t, and w_t. Assume $\mathbb{E}[v_t|\mathcal{G}_t] = 0$ for all $t \in \mathbb{N}$ and $\sup_{t\ge1} \mathbb{E}[\|v_t\|^\gamma|\mathcal{G}_t] < \infty$ $\mathbb{P}$-a.s. for some $\gamma > 2$, and

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

In view of Corollary 3.1, the sequence $\{c_t\}_{t\in\mathbb{N}}$ ceils the growth of the forecasts and in this sense censors the forecasts as to ensure stability. The dither v_t in (3.44) may be thought of being deliberately applied by a forecasting agency in order to satisfy the excitation condition (3.43). It will turn out Proposition

3.3 below that this guarantees convergence of the AML algorithm. Observe that no requirement on the magnitude of the dither is made in Assumption 3.3 so that the size of this perturbation may be very small.

The series of forecasts generated by the AML algorithm (3.41) and the forecasting rule (3.44) define an AML-based adaptive learning scheme for the linear system (3.1). With the notation of Section 3.1, the system under AML-based adaptive learning evolves over time according to the set of equations

$$\begin{cases} \hat{\theta}_t &= \hat{\theta}_{t-1} + (y_t - \hat{\theta}_{t-1}\hat{X}_{t-1})\hat{X}_{t-1}^\top P_{t-1}, \\ P_{t-1} &= P_{t-2} - P_{t-2}(\hat{X}_{t-1}\hat{X}_{t-1}^\top)P_{t-2}\,(1 + \hat{X}_{t-1}^\top P_{t-2}\hat{X}_{t-1})^{-1}, \\ \hat{E}_t &= \left((y_t - \hat{\theta}_t\hat{X}_{t-1})^\top, (\mathrm{pr}_{-1}^{n_4}\hat{E}_{t-1})^\top\right)^\top, \\ Y_t^e &= \hat{\Psi}_t(Y_t, W_t, \hat{E}_t, Y_{t-1}^e; \hat{\theta}_t), \\ x_{t+1} &= G(\xi_{t+1}, x_t, Y_t^e). \end{cases} \tag{3.45}$$

The last condition required for consistent estimates is an identifiability condition. Let $A(s)$, $B(s)$, $C(s)$, and $D(s)$, $s \in \mathbb{C}$, be matrix polynomials of the form

$$\begin{aligned} A(s) &= I_d s^{n_1} - A^{(1)}s^{n_1-1} - \cdots - A^{(n_1)}, \\ B(s) &= B^{(0)}s^{n_2} + B^{(1)}s^{n_2-1} + \cdots + B^{(n_2)}, \\ C(s) &= C^{(1)}s^{n_3-1} + \cdots + C^{(n_3)}, \\ D(s) &= I_d s^{n_4} + D^{(1)}s^{n_4-1} + \cdots + D^{(n_4)}, \end{aligned} \tag{3.46}$$

where I_d denotes the $d \times d$ identity matrix and $A^{(i)}$, $B^{(i)}$, $C^{(i)}$, and $D^{(i)}$ are the $d \times d$ parameter matrices of the economic law (3.1).

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 $\neq 0$). Equivalently, $A(s)$, $B(s)$, $C(s)$ and $D(s)$ are left coprime, 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 colored noise. It is automatically satisfied for the white-noise case. We are now ready to state our main theorems.

Theorem 3.2. *Consider the system (3.1). Let the following hypotheses be satisfied:*

(i) All eigenvalues λ_i of the matrix $\mathbf{A}$ *defined in (3.3) lie on or within the unit circle, that is, $|\lambda_i| \leq 1$.*
(ii) Identifiability assumption. The matrix polynomials $A(s)$, $B(s)$, $C(s)$, and $D(s)$ given in (3.46) are left coprime.
(iii) The exogenous perturbations satisfy Assumptions 3.1 and 3.2.
(iv) Forecasts are generated by the AML-based learning scheme as defined in Assumption 3.3.

Then the system (3.45) is stable in the sense of Proposition 3.1 and the AML-based adaptive learning scheme generates strongly consistent estimates, i.e., $\hat{\theta}_t \to \theta$ $\mathbb{P}$*-a.s. as* $t \to \infty$*, globally for all initial conditions.*

The proof of Theorem 3.2 is quite involved and will be carried out in Section 3.5.2 below. Conditions (i), (iii), and (iv) of Theorem 3.2 will ensure that the system (3.45) is stable in the sense of Proposition 3.1. Condition (ii) is needed to ensure strongly consistent parameter estimates. It should be emphasized that only the dither and the bounds for the forecasts in Assumption (iv) matter, but not the particular functional form of the forecasting rule.

Theorem 3.2 makes no assumption on the stability of the system (3.1) under rational expectations. So by the following Theorem 3.3, a forecasting agency will apply the approximated unbiased forecasting rule only if the system is stable under that forecasting rule.

Theorem 3.3. *Let the hypotheses of Theorem 3.2 be satisfied and assume, in addition, that the following holds:*

(i) All eigenvalues λ_i *of the matrix* $\mathbf{A}$ *defined in (3.3) lie within the unit circle, i.e.,* $|\lambda_i| < 1$.
(ii) The system (3.1) under the unbiased no-updating rule (3.19) is stable in the sense that all eigenvalues of the coefficient matrix of the system (3.20) lie within the unit circle.
(iii) The parameter estimates generated by (3.20) converge fast enough so that

$$\|\theta - \hat{\theta}_t\| \, \|\hat{X}_{t-1}\| \to 0 \quad \mathbb{P} - a.s. \ as \ \ t \to \infty.$$

Then the tolerance level $c_{\rm tol}$ *may be chosen large enough so that all orbits of the AML-based learning scheme converge to orbits with rational expectations, globally on* $\mathbb{R}^{d(n_1+n_2+n_3+n_4+1)}$*. Moreover,* $\|\hat{\epsilon}_t - \epsilon_t\| \to 0$ $\mathbb{P}$*-a.s. as* $t \to \infty$*.*

Observe that Condition (iii) is automatically satisfied if all exogenous noise is bounded, because in this case $\{\hat{X}_t\}_{t\in\mathbb{N}}$ is bounded. This observation will become clear in below's proof of Theorem 3.3.

Corollary 3.3 below provides convergence rates for the AML algorithm in terms of the observable design matrices P_t defined in (3.41), which are

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

where$P_t^{-1} = P_{t-1}^{-1} + \hat{X}_t \hat{X}_t^{\top}$. These bounds indicate how precise the actual parameter estimates are, so that the dither could be removed as soon as a certain threshold is reached. Thus, the AML-based learning scheme allows a forecasting agency to select an approximation of the desired forecasting rule from actual parameter estimates at any point in time. In particular, the no-updating rule used in Assumption 3.3 could be replaced by any forecasting rule which generates minimal-state-variable solutions provided this rule exists. As soon as parameter estimates are sufficiently close to their true values, this enables an agency to select a stable minimal-state-variable solution if it exists.

Remark 3.8. *(i) The convergence of the AML algorithm has first been investigated by Ljung (1977). His stability analysis is based on a limiting non-random ordinary differential equation. We follow a martingale approach initiated by Solo (1979) and later extended by Lai & Wei (1986b). The weak persistent excitation condition is considerably less restrictive than the assumptions used by Ljung (1977) and Solo (1979). See Lai & Wei (1986b, pp. 228-231) for an account of the literature.*

(ii) The AML algorithm is an extension of the ordinary-least-squares (OLS) algorithm as usually applied in the learning literature to a class of more general noise processes. It is also referred to as recursive extended-least-squares (RELS) *algorithm (Caines 1988, p. 540 and p. 557) or* recursive prediction-error (RPE) *algorithm (Ljung & Söderström 1983, Chap. 3.7).*

(iii) The concept of continually disturbed inputs, referred to as dither*, is originally due to Caines, e.g., see Lai & Wei (1986b, p. 247). It is generalized in Lai & Wei (1986a, 1987) to the concept of* occasional excitation *of the inputs, where the dither is exerted only if the excitation condition is violated. For an application of these techniques, we refer to their work.*

(iv) Section 3.2.2 showed that contrary to (3.42), MSV predictors require full knowledge of the process that governs the exogenous observables. Insufficient estimates for that process may result in a failure of the learning scheme. For example, if the exogenous process $\{w_t\}_{t\in\mathbb{N}}$ *is an AR(1) process, then the design matrix pertaining to the AML estimates of the corresponding coefficients may violate the persistent excitation condition (3.43). As a consequence, strongly consistent estimates for these coefficients may be unavailable, cf. Lai & Wei (1986a).*

Summary and Conclusions

We introduced a learning scheme based on the AML 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 it 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. Resolving an issue put forward in McCallum (1999), this approach allows to select an approximation of a preferred forecasting rule on the basis of an adequate economic reasoning, as soon as parameter estimates are sufficiently precise. The mathematical object to be estimated from time series data was the error function associated with the economic law.

Methodologically, we distinguished between four separate issues. First, the existence of a desired unbiased forecasting rule; second, the dynamic stability of the system under an unbiased forecasting rule; third, the dynamic stability of the system under the applied learning scheme and, finally, the success of the learning scheme in terms of providing strongly consistent estimates of the system's parameters.

3.5 Mathematical Appendix

3.5.1 On the Stability of Linear Systems

We begin with the proofs of Proposition 3.1 and Corollary 3.1.

Proof of Proposition 3.1. The proof is an adaptation of Lai & Wei (1982, Thm. 2 (i), p. 161). Let

$$Z_t = \Big((\sum_{i=0}^{n_2} B^{(n_2-i)}\, y^e_{t-1,t+i} + \sum_{i=1}^{n_3} C^{(i)} w_{t-i} + \epsilon_t + \sum_{i=1}^{n_4} D^{(i)}\, \epsilon_{t-i} + b)^\top, 0, \ldots, 0 \Big)^\top .$$

Using (3.3) and the notation introduced above, the system (3.4) may be rewritten as $Y_t = \mathbf{A} Y_{t-1} + Z_t$ and thus

$$Y_T = \mathbf{A}^T Y_0 + \sum_{t=1}^{T} \mathbf{A}^{T-t}\, Z_t. \tag{3.47}$$

Let $\overline{\lambda}$ denote the maximum modulus of the eigenvalues of $\mathbf{A}$ and $M = \max_j m_j$ with m_j denoting the multiplicity of the j-th eigenvalue of $\mathbf{A}$. By Varga (1962, Thm. 3.1, p. 65), there exists a positive constant $c > 0$ such that[2]

$$\|\mathbf{A}^T\| \sim c T^{M-1} \overline{\lambda}^T \quad \text{as} \quad T \to \infty. \tag{3.48}$$

Since by assumption $\overline{\lambda} \leq 1$, $\|\mathbf{A}^T\| = O(T^{M-1})$. In view of Assumptions (ii) and (3.8), we have $\|Z_t\| = o(t^\beta)$ and therefore by (3.47) and (3.48), there exists some $\alpha > 0$ so that

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

This proves the assertion. □

Proof of Corollary 3.1. In view of the assumptions, $\|Z_t\| = o(t^\beta)$. Since $\overline{\lambda} < 1$, (3.48) implies $\|\mathbf{A}^\tau\| \leq \mu < 1$ for some $\tau > 0$ so that

[2] By $a_t \sim b_t$ $\mathbb{P}$-a.s. for two sequences of random numbers $\{a_t\}$ and $\{b_t\}$, we mean that $\lim_{t\to\infty} \frac{a_t}{b_t} = 1$ $\mathbb{P}$-a.s.

$$\|Y_T\| \leq \|\mathbf{A}^T\| \|Y_0\| + \sum_{t=1}^{T} \|\mathbf{A}^{T-t}\| \|Z_t\| = o(T^\beta) \quad \mathbb{P} - \text{a.s.}$$

□

For sake of completeness, we state the following corollary.

Corollary 3.2. *Consider the linear system,*

$$Y_t = \mathbf{A}Y_{t-1} + Z_t, \quad t \in \mathbb{N}.$$

Assume that all eigenvalues of $\mathbf{A}$ *lie within the unit circle and that*

$$\sum_{t=1}^{T} \|Z_t\|^2 = O(T).$$

Then

$$\sum_{t=1}^{T} \|Y_t\|^2 = O(T).$$

Proof. Note first that the Schwarz inequality implies

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

In view of (3.47), it follows that

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

□

The following approximation result will be needed in the convergence proof of Theorem 3.3 below.

Proposition 3.2. *Consider two linear systems, one given by*

$$Y_t = \mathbf{A}Y_{t-1} + Z_t, \quad t \in \mathbb{N},$$

and another one defined by

$$\hat{Y}_t = \hat{\mathbf{A}}_t \hat{Y}_{t-1} + \hat{Z}_t, \quad t \in \mathbb{N},$$

where $\mathbf{A}$ *is a non-random matrix,* $\{\hat{\mathbf{A}}_t\}_{t\in\mathbb{N}}$ *is a sequence of random matrices, and* $\{Z_t\}_{t\in\mathbb{N}}$ *and* $\{\hat{Z}_t\}_{t\in\mathbb{N}}$ *are two sequences of random vectors which are bounded* $\mathbb{P}$*-a.s. Let all eigenvalues of* $\mathbf{A}$ *lie within the unit circle and assume that* $\mathbb{P}$*-a.s.*

$$\lim_{T\to\infty} \|\mathbf{A} - \hat{\mathbf{A}}_T\| = 0 \quad \textit{and} \quad \lim_{T\to\infty} \|Z_T - \hat{Z}_T\| = 0$$

Then the sequence $\lim_{T\to\infty} \|Y_T - \hat{Y}_T\|$ *is bounded* $\mathbb{P}$*-a.s.*

Proof. For arbitrary $t \in \mathbb{N}$, we have

$$Y_{t+j} = \mathbf{A}^j Y_t + \sum_{i=1}^{j} \mathbf{A}^{j-i} Z_{t+i}$$

and

$$\hat{Y}_{t+j} = \prod_{i=1}^{j} \hat{\mathbf{A}}_{t+i} \hat{Y}_t + \sum_{i=1}^{j-1} \Big(\prod_{k=i+1}^{j} \hat{\mathbf{A}}_{t+k} \Big) \hat{Z}_{t+i} + Z_{t+j}.$$

Thus

$$\|Y_{t+j} - \hat{Y}_{t+j}\| \leq \|\mathbf{A}^j\| \|Y_t - \hat{Y}_t\| + \sum_{i=1}^{j} \|\mathbf{A}^{j-i}\| \|Z_{t+i} - \hat{Z}_{t+i}\|$$
$$+ \Big\| \mathbf{A}^j - \prod_{i=1}^{j} \hat{\mathbf{A}}_{t+i} \Big\| \|\hat{Y}_t\| + \sum_{i=1}^{j-1} \Big\| \mathbf{A}^{j-i} - \prod_{k=i+1}^{j} \hat{\mathbf{A}}_{t+k} \Big\| \|\hat{Z}_{t+i}\|.$$

For each $\delta > 0$ there exists $t \in \mathbb{N}$ such that $\mathbb{P}$-a.s.

$$\|\mathbf{A} - \hat{\mathbf{A}}_{t+i}\| < \delta \quad \text{and} \quad \|Z_{t+i} - \hat{Z}_{t+i}\| < \delta \text{ for all } \ i \geq 1$$

The proposition then follows from the eigenvalue condition and the boundedness of $\{\hat{Z}_t\}_{t\in\mathbb{N}}$. □

3.5.2 Proofs of Theorem 3.2 and Theorem 3.3

The proof of Theorem 3.2 requires a series of Propositions and Theorems which will be developed next. Following Lai & Wei (1986b), we first provide a general theorem which establishes strong consistency of the AML algorithm if the *weak persistent excitation condition* holds. To this end we denote by $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ the minimum and maximum eigenvalues of a symmetric matrix B, respectively.

Proposition 3.3. *(Lai & Wei 1986b, Thm. 4, p. 241)*
Consider the stochastic regression model

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

where A is a $d \times k$ non-random matrix, $D^{(1)}, \ldots, D^{(n_4)}$ are non-random $d \times d$ matrices, and $\{\epsilon_t\}_{t\in\mathbb{N}}$ is a stochastic process which satisfies Assumption 3.1. Let η_t be $\mathcal{F}_t$ measurable and

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

Consider the AML algorithm, given by

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

where

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

If the weak persistent excitation condition

$$\frac{\lambda_{\min}\left(\sum_{t=1}^T X_t X_t^\top\right)}{\log\left(e + \lambda_{\max}\left(\sum_{t=1}^T X_t X_t^\top\right)\right)} \to \infty \quad \mathbb{P} - a.s. \text{ as } T \to \infty \tag{3.49}$$

holds, then

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

Corollary 3.3. *Under the hypotheses of Proposition 3.3,*

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

The proof of Corollary 3.3 is found in Lai & Wei (1986a). Let $\widetilde{A}(s)$ denote the adjoint of the matrix $A(s)$ such that $\widetilde{A}(s)A(s) = a(s)I_d$ 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 3.1. *Let $A(s), K_1(s), \ldots, K_m(s)$ 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 $N_1(s), \ldots, N_m(s)$ and vectors $\alpha_1, \ldots, \alpha_{n_1} \in \mathbb{R}^d$ such that*

$$\sum_{i=1}^{n_1} \alpha_i^\top \tilde{A}(s)K_j(s)s^{i-1} = a(s)N_j(s) \quad \text{for all } j = 1, \ldots, m, \tag{3.50}$$

then $\alpha_1 = \cdots = \alpha_{n_1} = 0$.

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

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

Suppose that (3.50) holds. Then

$$\begin{aligned}\sum_{i=1}^{n_1} \alpha_i^\top \widetilde{A}(s)s^{i-1} &= \sum_{i=1}^{n_1} \alpha_i^\top \widetilde{A}(s)\big[A(s)M_0(s) + \sum_{j=1}^{m} K_j(s)M_j(s)\big]s^{i-1} \\ &= a(s)\left[\sum_{i=1}^{n_1} \alpha_i^\top M_0(s) + \sum_{j=1}^{m} N_j(s)M_j(s)\right]. \qquad (3.51)\end{aligned}$$

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

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

It follows from (3.51) that $\left(\sum_{i=1}^{n_1} \alpha_i s^{i-1}\right)^\top \tilde{A}(s) = 0$ for all s. Since the leading coefficient of $A(s)$ is the identity matrix, $A(s)$ and $\tilde{A}(s)$ have full rank for almost all s. Thus $\sum_{i=1}^{n_1} \alpha_i s^{i-1} = 0$ for almost all s and therefore $\alpha_1 = \cdots = \alpha_{n_1} = 0$. □

Lemma 3.2. *Let A and be an $d_1 \times d_2$ matrix with* $\ker A^\top = \{0\}$ *and B be a symmetric $d_2 \times d_2$ matrix. Then there exists $\rho > 0$ with $\lambda_{\min}(ABA^\top) \ge \rho\lambda_{\min}(B)$.*

Proof. By Theorem 3.2.1, Lancaster (1969, p. 109), any symmetric matrix C satisfies $\lambda_{\min}(C) = \min_{\|x\|=1}(x^\top Cx)$. Since $AA^\top$ is nonsingular, for any $0 \ne x \in \mathbb{R}^{d_1}$,

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

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

□

Having stated these basic results, we need the following representation of the system (3.1). Letting ι denote the unit shift operator (i.e., $\iota X_t = X_{t+1}$), the system (3.1) may be rewritten as

$$A(\iota)y_{t-n_1} = B(\iota)u_{t-n_2+1} + C(\iota)w_{t-n_3} + D(\iota)\epsilon_{t-n_4}, \qquad (3.52)$$

where we have set $u_{t-1} = y^e_{t-1,t+n_2}$ for the forecasts. The next Proposition 3.4 is a version of Theorem 2 in Lai & Wei (1986b, p. 235) for an additional source of inputs.

Proposition 3.4. *Suppose that the matrix polynomials $A(s)$, $B(s)$, $C(s)$, and $D(s)$ given (3.46) are left coprime and that (3.52) holds for all times t. Let*

$$\begin{aligned} X_{t-1}^\top &= (y_{t-1}^\top, \ldots, y_{t-n_1}^\top, u_{t-1}^\top, \ldots, u_{t-n_2-1}^\top, w_{t-1}^\top, \ldots, w_{t-n_3}^\top, \epsilon_{t-1}^\top, \ldots, \epsilon_{t-n_4}^\top), \\ z_{t-1}^\top &= (u_{t-1+dn_1}^\top, \ldots, u_{t-n_2+1}^\top, w_{t-1+dn_1}^\top, \ldots, w_{t-n_3}^\top, \epsilon_{t-1+dn_1}^\top, \ldots, \epsilon_{t-n_4}^\top). \end{aligned} \tag{3.53}$$

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

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

Proof. Let $\widetilde{A}(s)$ 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 $\widetilde{A}(\iota) A(\iota) X_{t-1} = a(\iota) X_{t-1}$, (3.52) implies

$$\begin{aligned} a(\iota) y_{t-j} &= \widetilde{A}(\iota)\left[B(\iota)\iota^{n_1-j} u_{t-n_2+1} + D(\iota)\iota^{n_1-j} w_{t-n_3} + C(\iota)\iota^{n_1-j}\epsilon_{t-n_4}\right], \\ a(\iota) u_{t-j} &= a(\iota)\iota^{n_2+1-j} u_{t-n_2+1}, \\ a(\iota) w_{t-j} &= a(\iota)\iota^{n_4-j} w_{t-n_3}, \\ a(\iota)\epsilon_{t-j} &= a(\iota)\iota^{n_4-j}\epsilon_{t-n_4}. \end{aligned} \tag{3.54}$$

Now set $V_t = (V_t^{(1)\top}, \ldots, V^{(4)\top})^\top = a(\iota) X_{t-1}$ with

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

and

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

Then (3.54) has a block-matrix representation as

$$\begin{pmatrix} V_t^{(1)} \\ V_t^{(2)} \\ V_t^{(3)} \\ V_t^{(4)} \end{pmatrix} = \mathcal{D} \begin{pmatrix} U_t \\ W_t \\ E_t \end{pmatrix} \quad \text{where} \quad \mathcal{D} = \begin{pmatrix} \mathcal{D}_1 & \mathcal{D}_2 & \mathcal{D}_3 \\ \mathcal{D}_4 & 0 & 0 \\ 0 & \mathcal{D}_5 & 0 \\ 0 & 0 & \mathcal{D}_6 \end{pmatrix} \tag{3.55}$$

The block matrix $\mathcal{D}$ in (3.55) 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 $d(n_1 + 1 + n_2 + n_4 + n_3)$, then there exist vectors $\alpha_1, \ldots, \alpha_{n_1}$, $\beta_1, \ldots, \beta_{n_2+1}$, $\gamma_1, \ldots, \gamma_{n_3}$, and $\delta_1, \ldots, \delta_{n_4}$ in $\mathbb{R}^d$ such that

$$\sum_{i=1}^{n_1} \alpha_i^\top \widetilde{A}(s) B(s) s^{i-1} = -a(s) \sum_{j=1}^{n_2+1} \beta_j^\top I_d s^{j-1},$$
$$\sum_{i=1}^{n_1} \alpha_i^\top \widetilde{A}(s) C(s) s^{i-1} = -a(s) \sum_{j=1}^{n_3} \gamma_j^\top I_d s^{j-1},$$
$$\sum_{i=1}^{n_1} \alpha_i^\top \widetilde{A}(s) D(s) s^{i-1} = -a(s) \sum_{j=1}^{n_4} \delta_j^\top I_d s^{j-1},$$

which is impossible by Lemma 3.1. Hence the matrix in (3.55) has full rank. Noting that $z_{t-1} = (U_t^\top, W_t^\top, E_t^\top)^\top$, we can apply Lemma 3.2 to see that there exists $\rho > 0$ such that

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

To complete the proof, it suffices to show that

$$\lambda_{\min} \left(\sum_{t=1}^{T-dn_1} V_t V_t^\top \right) \leq (dn_1 + 1) \left(\sum_{j=0}^{dn_1} a_j^2 \right) \lambda_{\min} \left(\sum_{t=1}^{T} X_t X_t^\top \right), \tag{3.56}$$

Let ζ be a unit vector. Then

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

implies

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

proving (3.56). □

Proposition 3.5. *Let the following hypotheses be satisfied.*

(i) The matrix polynomials $A(s)$, $B(s)$, $C(s)$, and $D(s)$ defined in (3.46) are left coprime and (3.52) holds for all t.
(ii) Assumptions 3.1 and 3.2 are fulfilled.

(iii) The inputs $u_t = y^e_{t,t+1+n_2}$ *satisfy the following conditions. Let* $\mathcal{G}_t := \sigma(\mathcal{F}_{t-1}, y_t, w_t)$ *be defined as before. 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.:*

$$\sup_{t\geq 1} \mathbb{E}[\|\tilde{u}_t\|^\alpha|\mathcal{G}_t] < \infty \quad \textit{for some} \quad \alpha > 2, \tag{3.57}$$

$$\frac{\lambda_{\min}\left(\sum_{t=1}^T \tilde{u}_t\tilde{u}_t^\top\right)}{\log T + \|\tilde{u}_T\|^2} \to \infty \quad \textit{as } T \to \infty, \tag{3.58}$$

$$\sum_{t=1}^T \|\mathbb{E}[u_t|\mathcal{G}_t]\|^2 = O\left(\lambda_{\min}\left(\sum_{t=1}^T \tilde{u}_t\tilde{u}_t^\top\right)\right). \tag{3.59}$$

For arbitrary $n > 0$*, set*

$$Z_t := (\epsilon_{t+n}^\top; u_{t+n-1}^\top, w_{t+n-1}^\top, \epsilon_{t+n-1}^\top; \ldots; u_t^\top, w_t^\top, \epsilon_t^\top)^\top. \tag{3.60}$$

Then

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

The proof of Proposition 3.5 will be prefaced by Theorem 3.4 and Corollary 3.4 which we state without proof.

Theorem 3.4. *(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* σ*-algebras. Suppose that the following properties hold with probability 1.*

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

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

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

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

(iii) $\tau_t \in \mathbb{R}^{d_1}$ *is* $\mathcal{H}_{t-1}$ *measurable with*

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

Then

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

Corollary 3.4. *Under the hypotheses of Theorem 3.4, if, in addition,*

$$\liminf_{T\to\infty} \frac{\lambda_{\min}\left(\sum_{t=1}^{T} \tau_t \tau_t^\top\right)}{\lambda_{\min}\left(\sum_{t=1}^{T} \eta_t \eta_t^\top\right)} > 0 \quad \mathbb{P} - a.s., \tag{3.61}$$

then

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

Proof of Proposition 3.5. The proof consists of an induction argument on n and of a repeated application of Theorem 3.4. To apply Theorem 3.4, 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 3.1 and 3.2, all requirements of Corollary 3.4 are fulfilled. In particular, since $\lambda_{\min}\left(\sum_{t=1}^{T} \tilde{u}_t \tilde{u}_t^\top\right) = O(T)$,

$$\liminf_{T\to\infty} \frac{\lambda_{\min}\left(\sum_{t=1}^{T} \zeta_t^{(1)} \zeta_t^{(1)\top}\right)}{\lambda_{\min}\left(\sum_{t=1}^{T} \tilde{u}_t \tilde{u}_t^\top\right)} > 0 \quad \mathbb{P} - \text{a.s.} \tag{3.62}$$

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 (3.57)-(3.59), $\zeta_t^{(2)}$ again meets all requirements of Corollary 3.4. Hence, we conclude from $\lambda_{\min}\left(\sum_{t=1}^{T} \tilde{u}_t \tilde{u}_t^\top\right) = O(T)$ that

$$\liminf_{T\to\infty} \frac{\lambda_{\min}\left(\sum_{t=1}^{T} \zeta_t^{(2)} \zeta_t^{(2)\top}\right)}{\lambda_{\min}\left(\sum_{t=1}^{T} \tilde{u}_t \tilde{u}_t^\top\right)} > 0 \quad \mathbb{P} - \text{a.s.} \tag{3.63}$$

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 3.1 (i) and (ii) that $\zeta_t^{(3)}$ satisfies all hypotheses of Corollary 3.4. From an argument analogous to Step 2, we obtain

$$\liminf_{T\to\infty} \frac{\lambda_{\min}\left(\sum_{t=1}^T \zeta_t^{(3)}\zeta_t^{(3)\top}\right)}{\lambda_{\min}\left(\sum_{t=1}^T \tilde{u}_t\tilde{u}_t^\top\right)} > 0 \quad \mathbb{P}-\text{a.s.} \tag{3.64}$$

This proves the case $n = 1$. By (3.58), we have

$$\lambda_{\min}\left(\sum_{t=1}^{T} \tilde{u}_t\tilde{u}_t^\top\right) \sim \lambda_{\min}\left(\sum_{t=1}^{T-m} \tilde{u}_t\tilde{u}_t^\top\right)$$

for any fixed m and a repeated application of the above three steps establishes the desired conclusion for arbitrary n. □

The key ingredient of the proof of Theorem 3.2 is the following Theorem 3.5 which is an extension of Lai & Wei (1986b, Thm. 5 (ii), p. 242) to a system with additional exogenous variables as inputs.

Theorem 3.5. *Let the hypotheses of Proposition 3.5 be satisfied. Then the following properties hold true* $\mathbb{P}$*-a.s.:*

(*a*) $\sum_{t=1}^T \|u_t\|^2 = O(T)$, (*b*) $\sum_{t=1}^T \|\tilde{u}_t\|^2 = O(T)$,

(*c*) $\liminf_{T\to\infty} \frac{\lambda_{\min}\left(\sum_{t=1}^T X_tX_t^\top\right)}{\lambda_{\min}\left(\sum_{t=1}^T \tilde{u}_t\tilde{u}_t^\top\right)} > 0.$

Proof. Property (*b*) follows from (3.57) 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 (*a*) follows from (3.59).

The last property (*c*) follows from Proposition 3.4 by first showing that the regressors (3.36) can be reduced to regressors (3.53) consisting only of inputs and then applying Proposition 3.5, noting that $\lambda_{\min}\left(\sum_{t=1}^T z_tz_t^\top\right) \geq \lambda_{\min}\left(\sum_{t=1}^T Z_tZ_t^\top\right)$ for any sub-vector z_T of Z_t, where the latter was defined in (3.60). Note in this connection that interchanging any vector entries of a regressor z_t will not change the eigenvalues of the matrix $z_tz_t^\top$, since it amounts to interchanging columns and rows simultaneously. □

Theorem 3.6 is an extension of Lai & Wei (1986b, Thm. 6 (ii)) and provides conditions under which strong consistency of the AML algorithm obtains.

Theorem 3.6. *Let the following hypotheses be satisfied.*

(i) Assumptions (i)-(iii) of Theorem 3.2 are satisfied.
(ii) Assumption (iii) of Proposition 3.5 is satisfied.

Then the AML-algorithm (3.41) generates strongly consistent estimates, i.e.,

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

Proof. By Theorem 3.5, $\sum_{t=1}^{T} \|u_t\|^2 = O(T)$. In view of Assumption (i), all prerequisites of Proposition 3.1 are then satisfied, implying that there exists $\alpha > 0$ such that $\|X_t\|^2 = O(t^\alpha)$ $\mathbb{P}$-a.s. Hence

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

It follows from (3.58) and Theorem 3.5 that

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

and hence (3.49) is satisfied. The assertion then follows from Proposition 3.3. □

We are now ready to establish Theorem 3.2.

Proof of Theorem 3.2. By Assumption (iii) and (iv) we have $\|w_t\| = o(t^{\frac{1}{2}})$, $\|\epsilon_t\| = o(t^{\frac{1}{2}})$, and $\|y^e_{t,t+j}\| = o(t^{\frac{1}{2}})$ $\mathbb{P}$-a.s. Hence by Assumption (i), the system (3.45) is stable in the sense of Proposition 3.1.

In order to apply Theorem 3.6 it remains to show that Conditions (3.57)-(3.59) of Proposition 3.5 are fulfilled. As before, set $u_t = y^e_{t,t+1+n_2}$. Since by (3.44) $\tilde{u}_t = u_t - \mathbb{E}[u_t|\mathcal{G}_t] = v_t$, Condition (3.57) is satisfied. By an analogous argument to that of Remark 3.7 (ii), Assumption (iv) on the dither v_t implies that $\mathbb{P}$-a.s.

(*a*) $\|v_T\| = o(T^{\frac{1}{2}})$, (*b*) $\sum_{t=1}^{T} \|v_t\|^2 = O(T)$,

(*c*) $\liminf_{T\to\infty} \frac{1}{T} \lambda_{\min} \left(\sum_{t=1}^{T} v_t v_t^\top \right) > 0$.

Properties (a) and (b) imply Condition (3.58). Since $\|\mathbb{E}[u_t|\mathcal{G}_t]\| = \|u_t - v_t\| \leq c_t$, it follows from (3.44) that $\sum_{t=1}^{T} \|\mathbb{E}[u_t|\mathcal{G}_t\|^2 = O(T)$. Using Property (c), this yields Condition (3.59). Thus Assumption (ii) of Theorem 3.6 is satisfied and we can apply Theorem 3.6 to conclude that $\hat{\theta}_t \to \theta$ $\mathbb{P}$-a.s. □

Proof of Theorem 3.3. Since all eigenvalues **A** must lie strictly within the unit circle, Assumptions (iii) and (iv) of Theorem 3.2 imply that the system (3.45) is stable in the sense of Cororollary 3.1, i.e., $\|y_t\| = o(t^{\frac{1}{2}})$ $\mathbb{P}$-a.s.

In view of Theorem 3.2 it remains to show that all orbits of (3.45) eventually become orbits with rational expectations. It follows from the recursive definition of the prediction errors (3.40) that $\|\hat{\epsilon}_t\| = o(t^{\frac{1}{2}})$ and

$$\hat{\epsilon}_t - \epsilon_t = - \sum_{i=1}^{n_4} D^{(i)} (\hat{\epsilon}_{t-i} - \epsilon_t) + (\theta - \hat{\theta}_t) \hat{X}_{t-1}, \quad t \in \mathbb{N}. \tag{3.65}$$

By Assumption 3.1 (ii), all eigenvalues of the matrix

$$\mathbf{D} = \begin{pmatrix} -D^{(1)} & \cdots & \cdots & -D^{(n_4)} \\ I_d & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I_d & 0 \end{pmatrix}$$

associated with (3.65) lie within the unit circle. Since the additive term in (3.65) vanishes by Assumption (iii), it follows that $\mathbb{P}$-a.s.

$$\lim_{t\to\infty} (\hat{\epsilon}_t - \epsilon_t) = 0.$$

As a consequence, there exists $t_{\mathrm{tol}} > 0$ such that

$$\|\hat{\psi}(\hat{x}_t, Y^e_{t-1}; \hat{\theta}_t)\| = o(t^{\frac{1}{2}}) \quad \mathbb{P}\text{-a.s. for all} \quad t \geq t_{\mathrm{tol}}.$$

Hence, if c_{tol} is chosen large enough, censoring c_t will not be applied anymore after time t_{tol}. An an application of Proposition 3.2 to the system under unbiased no-updating (3.20) and its approximation as given by (3.45) implies that all orbits of (3.45) eventually become orbits with rational expectations.

□

4

Economic Models Subject to Stationary Noise

In this chapter the concepts of Chapter 2 are generalized to the case in which the exogenous perturbations are driven by a stationary ergodic process. We establish necessary conditions for the existence and uniqueness of unbiased forecasting rules. For this purpose it is convenient to imbed our setup into the theory of *random dynamical systems*. In joining the dynamical systems theory with the theory of stochastic processes, the theory of random dynamical systems developed by Arnold (1998) and others provides a natural formal framework to analyze the behavior of economic systems subject to stationary ergodic perturbations.

To apply the theory of random dynamical systems, we extend the sequence of random variables $\{\xi_t\}_{t\in\mathbb{N}}$ with values in $\Xi \subset \mathbb{R}^{d_\xi}$ defining the random environment of the economy to a sequence $\{\xi_t\}_{t\in\mathbb{Z}}$ with so-called *two-sided time* $\mathbb{Z} := \{0, \pm 1, \pm 2, \dots\}$. We will represent the stochastic perturbations by means of a metric dynamical system which essentially amounts to using the equivalent *canonical realization process* of a stochastic process. To fix the notation, let $\vartheta : \Omega \to \Omega$ be a measurable invertible map on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which is measure preserving with respect to $\mathbb{P}$ and whose inverse ϑ^{-1} is again measurable. Assume that $\mathbb{P}$ is ergodic with respect to ϑ and let ϑ^t denote the t-th iterate of the map ϑ. The collection $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$ is called an *ergodic metric dynamical system*, cf. Arnold (1998, Appendix A.2, p. 542). It is well known that any stationary ergodic process $\{\xi_t\}_{t\in\mathbb{Z}}$ can be represented by an ergodic metric dynamical system. This means that there exists a measurable map $\xi : \Omega \to \Xi$ such that for each fixed $\omega \in \Omega$, a sample path of the noise process is given by

$$\xi_t(\omega) = \xi(\vartheta^t \omega), \quad t \in \mathbb{Z}. \tag{4.1}$$

The following assumption is a refinement of Assumption 2.1 of Chapter 2.

Assumption 4.1. *Let* $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$ *be an ergodic metric dynamical system on a complete probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. $\{\xi_t\}_{t\in\mathbb{Z}}$ *is a stationary ergodic* *stochastic process on* $(\Omega, \mathcal{F}, \mathbb{P})$ *with values in a subset* $\Xi \subset \mathbb{R}^{d_\xi}$ *which admits a representation of the form (4.1).*

In this setup, the set Ω may best be thought of the space of all sample paths of the noise process, so that $\omega \in \Omega$ denotes a particular sample path. Let us illustrate this point with a simple and standard example.

Example 4.1. *Consider the case in which* $\{\xi_t\}_{t\in\mathbb{Z}}$ *is a sequence of* Ξ*-valued iid random variables whose probability distribution is* μ*. Then* $(\Omega, \mathcal{F}, \mathbb{P}) = (\Xi^{\mathbb{Z}}, \mathcal{B}(\Xi)^{\mathbb{Z}}, \mu^{\mathbb{Z}})$*, where* Ω *is the space of all sample paths of the process,* $B(\Xi)^{\mathbb{Z}}$ *is the Borel* σ*-algebra of all cylinder sets, and* $\mu^{\mathbb{Z}}$ *is the product measure. Each element* $\omega = \{\omega(s)\}_{s\in\mathbb{Z}} \in \Omega$ *is a doubly infinite sequence describing a sample path of the process. The map*

$$\vartheta : \Omega \to \Omega, \quad \omega \mapsto \vartheta\omega$$

is defined by $(\vartheta\omega)(s) := \omega(s+1)$*,* $s \in \mathbb{Z}$ *and is called the left shift such that* $(\vartheta^t\omega)(s) = \omega(s+t)$ *for all* $t, s \in \mathbb{Z}$*. If*

$$\xi : \Omega \to \Xi, \quad \xi(\omega) := \omega(0)$$

denotes the evaluation map, a representation of the form (4.1) is obtained. More details are found in Appendix A.2 of Arnold (1998).

As before, let $\mathbb{X} \subset \mathbb{R}^{d_x}$ be the space of endogenous variables and $x_t \in \mathbb{X}$ denote the vector of endogenous variables describing the state of the economy at date t. Each $x_t \in \mathbb{X}$ is subdivided into $x_t = (\overline{x}_t, y_t) \in \overline{\mathbb{X}} \times \mathbb{Y} = \mathbb{X}$, where $y_t \in \mathbb{Y} \subset \mathbb{R}^{d_y}$ contains the endogenous variables for which expectations are formed, $\overline{x}_t \in \overline{\mathbb{X}} \subset \mathbb{R}^{\bar{d}_x}$ is the vector of the remaining variables. As before, an *economic law* is a map

$$G : \Xi \times \mathbb{X} \times \mathbb{Y}^m \longrightarrow \mathbb{X} \tag{4.2}$$

with the interpretation that the state of the economy at time $t+1$ is given by

$$x_{t+1} = G(\xi(\vartheta^t\omega), x_t, y^e_{t,t+1}, \ldots, y^e_{t,t+m}),$$

where x_t is the current observable state of the economy and

$$y^e_t := (y^e_{t,t+1}, \ldots, y^e_{t,t+m}) \in \mathbb{Y}^m$$

is the vector of forecasts for $(y_{t+1}, \ldots, y_{t+m})$ formed at time t. Again, x_t itself may be a vector of lagged endogenous variables including past realizations ξ_s, $s \le t$ of the exogenous noise process and past forecasts.

Remark 4.1. *Formally, each map*

$$G(\xi(\cdot), x, y^e) : \Omega \to \mathbb{X}, \quad (x, y^e) \in \mathbb{X} \times \mathbb{Y}^m$$

is a random variable which is allowed to depend on the whole path $\omega \in \Omega$. For many economic applications, two cases in which ξ_t is assumed to be observable at date t are of particular interest. This can be formalized as follows. Setting $\xi(\omega) = \omega(0)$, $\omega \in \Omega$ as in Example 4.1 with iid noise, one has $\xi_t(\omega) = \xi(\vartheta^t \omega)$ in the first case such that x_{t+1} becomes a deterministic function of the observable variables $\xi_t(\omega)$, x_t, and y_t^e at date t. Setting $\xi(\omega) = \omega(1)$, $\omega \in \Omega$ in the second case, one has $\xi_{t+1}(\omega) = \xi(\vartheta^t \omega)$ and $x_{t+1}(\cdot) = G(\xi_{t+1}(\cdot), x_t, y_t^e)$ becomes a random quantity in the sense that $\xi_{t+1}(\omega)$ is not observable prior to period $t+1$.

4.1 Forecasting Rules for Stationary Noise

As in the Markovian case, we assume that a forecasting agency which is boundedly rational in the sense of Sargent (1993) is responsible for issuing all relevant forecasts. At date t, the time-series information of the forecasting agency includes the actual history of the economy $\{x_s, y_{s-1}^e\}_{s \leq t}$, where

$$y_{s-1}^e = (y_{s-1,s}^e, \dots, y_{s-1,s-1+m}^e) \in \mathbb{Y}^m, \quad s \in \mathbb{Z} \tag{4.3}$$

as before. It will turn out that the vector of previous forecasts y_{t-1}^e is of particular importance. The forecasts $y_{t,t+1}^e, \dots, y_{t,t+m}^e$ are determined according to a *forecasting rule* depending on available information. In view of the present setup with stationary ergodic noise, a *forecasting rule* $\psi = (\psi^{(1)}, \dots, \psi^{(m)})$ is formally defined as a function

$$\psi : \Omega \times \mathbb{X} \times \mathbb{Y}^m \longrightarrow \mathbb{Y}^m, \quad (\omega, x, y^e) \longmapsto \psi(\omega, x, y^e), \tag{4.4}$$

such that

$$y_{t,t+i}^e = \psi^{(i)}(\vartheta^t \omega, x_t, y_{t-1}^e), \quad i = 1, \dots, m$$

is the forecast for the future realization y_{t+i}, $i = 1, \dots, m$, based on information available in period t.

To reflect the fact that the forecasting rule depends on the history of the economy, let $\mathcal{F}^0$ be a suitably defined sub-σ-algebra of $\mathcal{F}$ which describes the *past* of the economy. This σ-algebra will be specified in detail in Section 4.1.2 below. Let each forecasting rule

$$\psi(\cdot, x, y^e), \quad x \in \mathbb{X}, \ y^e \in \mathbb{Y}^m$$

be $\mathcal{F}^0/\mathcal{B}(\mathbb{Y}^m)$ measurable with $\mathcal{B}(\mathbb{Y}^m)$ denoting the σ-algebra of all Borel sets of $\mathbb{Y}^m$. Loosely speaking, this measurability assumption imposed on (4.4) together with the definition of ϑ implies that each $y_{t,t+i}^e$ will depend only on entries up to a certain time index of the path $\omega = (\dots, \omega_{-1}, \omega_0, \omega_1, \dots) \in \Omega$.

Inserting the forecasting rule (4.4) into the economic law (4.2) yields a discrete-time random dynamical system given by the *time-one map*

$$G_\psi : \Omega \times \mathbb{X} \times \mathbb{Y}^m \longrightarrow \mathbb{X} \times \mathbb{Y}^m, \tag{4.5}$$

defined by

$$G_\psi(\omega, x, y^e) = \begin{pmatrix} G(\xi(\omega), x, y^e) \\ \psi\big(\vartheta\omega, G(\xi(\omega), x, y^e), y^e\big) \end{pmatrix}.$$

Let x_{t-1} describe the state of the economy at time $t-1$ and y^e_{t-1} the vector of forecasts based on information available at time $t-1$. Then an application of the map (4.5) yields

$$\begin{pmatrix} x_t \\ y^e_t \end{pmatrix} = G_\psi(\vartheta^{t-1}\omega, x_{t-1}, y^e_{t-1})$$
$$= \begin{pmatrix} G(\xi(\vartheta^{t-1}\omega), x_{t-1}, y^e_{t-1}) \\ \psi(\vartheta^t\omega, G(\xi(\vartheta^{t-1}\omega), x_{t-1}, y^e_{t-1}), y^e_{t-1}) \end{pmatrix},$$

such that x_t describes the state of the economy at time t and y^e_t the vector of forecasts based on information available at that time.

The map G_ψ governs the evolution of the economy as follows. If the system started in some state (x_0, y^e_0), then the iteration of the map (4.5) under the perturbation ω induces a measurable map, again denoted by G_ψ, of the form $G_\psi : \mathbb{N} \times \Omega \times \mathbb{X} \times \mathbb{Y}^m \to \mathbb{X}$, which is recursively defined by

$$G_\psi(t, \omega, x_0, y^e_0) = \begin{cases} G_\psi\big(\vartheta^{t-1}\omega, G_\psi(t-1, \omega, x_0, y^e_0)\big) & \text{if } t > 0 \\ (x_0, y^e_0) & \text{if } t = 0 \end{cases}, \tag{4.6}$$

such that $(x_t, y^e_t) = G_\psi(t, \omega, x_0, y^e_0)$ is the state of the system at time t. By abuse[1] of notation, we write $G_\psi(1, \omega, x_0, y^e_0) \equiv G_\psi(\omega, x_0, y^e_0)$.

The map (4.6) generated by the time-one map G_ψ describes the *random dynamical system* over a metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$ in the sense of Arnold (1998). For arbitrary initial conditions $(x_0, y^e_0) \in \mathbb{X} \times \mathbb{Y}^m$ and any perturbation $\omega \in \Omega$, the sequence of points $\gamma(x_0, y^e_0) := \{(x_t, y^e_t)\}_{t\in\mathbb{N}}$ with $(x_t, y^e_t) = G_\psi(t, \omega, x_0, y^e_0)$, $t \in \mathbb{N}$ is called an *orbit* of the random dynamical system G_ψ starting at (x_0, y^e_0). In the sequel, (4.5) or, equivalently, (4.6) will be referred to as an *economic random dynamical system (ERDS).*

It is easily seen from (4.6) that the map G_ψ has two crucial properties. The first one is the fact that

$$G_\psi(0, \omega, x, y^e) = (x, y^e) \quad \text{for all } \ (x, y^e) \in \mathbb{X} \times \mathbb{Y}^m, \ \omega \in \Omega \tag{4.7}$$

is the identity map on $\mathbb{X} \times \mathbb{Y}^m$. The second property is

$$G_\psi(t_2 + t_1, \omega, x, y^e) = G_\psi\big(t_2, \vartheta^{t_1}\omega, G_\psi(t_1, \omega, x, y^e)\big) \tag{4.8}$$

for all $\omega \in \Omega$, $(x, y^e) \in \mathbb{X} \times \mathbb{Y}^m$, and arbitrary $t_1, t_2 \in \mathbb{N}$. To simplify the exposition and to define a filtration for the stochastic process (4.6) in Section

[1]This is done to avoid an inflation of symbols. No confusion should arise.

4.1.2, we will assume that for each $\omega \in \Omega$, the map $G_\psi(\omega,\cdot)$ defined in (4.5) is invertible. Writing

$$G_\psi^{-1}(\omega,\cdot) : \mathbb{X} \times \mathbb{Y}^m \to \mathbb{X} \times \mathbb{Y}^m$$

for the inverse of $G_\psi(\omega,\cdot)$, the map (4.6) can then be extended to a measurable map $G_\psi : \mathbb{Z} \times \Omega \times \mathbb{X} \times \mathbb{Y}^m \to \mathbb{X} \times \mathbb{Y}^m$, by recursively setting for each negative time $t < 0$,

$$G_\psi(t,\omega,x_0,y_0^e) = G_\psi^{-1}\big(\vartheta^t\omega, G_\psi(t+1,\omega,x_0,y_0^e)\big).$$

Then property (4.8) holds for all $t_1, t_2 \in \mathbb{Z}$ and G_ψ is said to satisfy the so-called *cocycle property.* It is shown in Arnold (1998, Thm. 1.1.6, p. 7) that the invertibility of the time-one map (4.5) follows from requiring the cocycle property.

We are now ready to characterize those forecasting rules which are correct in the sense that they provide best least-squares predictions conditional on available information and thus rational expectations.

4.1.1 Consistent Forecasting Rules

In this section we will develop the notion of a consistent forecasting rule for the case with stationary ergodic noise. Having defined a forecasting rule of the functional form (4.4), assume that the information available in a particular period $t \in \mathbb{Z}$ is given by the sub-σ-algebra

$$\mathcal{F}_{t,r} := \sigma\big(G_\psi(s,\cdot,x,y^e)\big|\ t-r-1 < s \le t,\ x \in \mathbb{X},\ y^e \in \mathbb{Y}^m\big) \tag{4.9}$$

of $\mathcal{F}$ which is generated by the past $r+1$ generalizations of the process (4.6). Writing $\mathcal{F}_t \equiv \mathcal{F}_{t,\infty}$ for $r = \infty$, the sequence of σ-algebras $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$ defines a filtration to which the stochastic process (4.6) is adapted.

We will stipulate the length of the memory to a fixed number $r \in \mathbb{N}$ which may be arbitrary large and which describes the amount of available information. This assumption may be justified both economically and mathematically. From a mathematical viewpoint, the system may be Markovian as before such that only a finite past influences the future states of the system, as discussed in Chapter 2.[2] From an economic point of view, there are many reasons why agents use only a finite memory when taking decisions, one being that their capacity to process information may be limited.

In Lemma 4.1 on p. 80 of Section 4.5 we show that $\mathcal{F}_{t,r} = \vartheta^{-t}\mathcal{F}_{0,r}$ for each $t \in \mathbb{Z}$ and arbitrary $r \in \mathbb{N}$. It is therefore natural to assume from now on that each map $\psi(\cdot,x,y^e)$, $x \in \mathbb{X}$, $y^e \in \mathbb{Y}^m$ is $\mathcal{F}_{0,r}/\mathcal{B}(\mathbb{Y}^m)$ measurable. Then each $\psi^{(i)}(\vartheta^t\cdot,x,y^e)$ is $\mathcal{F}_{t,r}/\mathcal{B}(\mathbb{Y}^m)$ measurable. As before, let

[2] Recall that if $\{\xi_t\}_{t\in\mathbb{Z}}$ is iid, then (4.6) is a Markov process, see also Arnold (1998, Sec. 2.1.3, p. 53).

$$y^e_{t,t+i} = \psi^{(i)}(\vartheta^t\omega, x_t, y^e_{t-1}), \quad i = 1, \dots, m \tag{4.10}$$

denote period-t's forecasts for the future realizations y_{t+i}, $i = 1, \dots, m$, respectively. The factorization lemma (Bauer 1992, p. 71) then guarantees the existence of a forecasting rule $\Psi = (\Psi^{(1)}, \dots, \Psi^{(m)})$ of the form (2.13) as defined in Chapter 2, given by a map $\Psi : \mathbb{X} \times \Sigma \to \mathbb{Y}^m$, such that for each $\omega \in \Omega$ and each $i = 1, \dots, m$,

$$y^e_{t,t+i} = \psi^{(i)}(\vartheta^t\omega, x_t, y^e_{t-1}) = \Psi^{(i)}\left(x_t, y^e_{t-1}, x_{t-1}, \dots, y^e_{t-r}, x_{t-r}\right) \tag{4.11}$$

$\mathbb{P}$-a.s., where

$$(x_s, y^e_s) = G_\psi(s, \omega, x_0, y^e_0), \quad s \leq t - 1.$$

In view of the estimation techniques which will be introduced in Chapter 5, this observation shows that a forecasting rule of the functional form (4.10) can always be written as a function depending on the history of observable quantities. The more formally defined forecasting rule (4.10) itself is a stationary object, whereas the form (4.11) can be thought as actually being applied by a forecasting agency within the model. Such a forecasting agency, naturally, would not be assumed to know the whole path $\omega \in \Omega$.

Recall that in each period t there are $m - 1$ forecasts

$$y^e_{t-1,t+1}, \dots, y^e_{t-1,t-1+m}$$

for y_{t+1} which were set prior to time t, where $y^e_{t-i,t+1}$ is based on information available in period $t - i$ and hence $\mathcal{F}_{t-i}$ measurable. As before, let $\mathbb{E}[\cdot|\mathcal{F}_{t-i}]$ denote the conditional expectations operator with respect to $\mathcal{F}_{t-i}$. All forecast errors $y_{t+1} - y^e_{t-i,t+1}$, $i = 0, \dots, m - 1$ vanish on average conditional on the respective σ-algebras $\mathcal{F}_{t-i}$ if and only if

$$\mathbb{E}\left[(y_{t+1} - y^e_{t-i,t+1})\middle|\mathcal{F}_{t-i}\right] = 0 \quad \mathbb{P}-\text{a.s.} \tag{4.12}$$

for all $i = 0, \dots, m - 1$. The law of iterated expectations (2.19) implies that the forecast $y^e_{t,t+i}$ for y_{t+i}, which is chosen in period t, has to satisfy the *consistency condition*

$$\mathbb{E}[y^e_{t,t+i}|\mathcal{F}_{t-1}] = y^e_{t-1,t+i} \quad \mathbb{P}-\text{a.s.} \tag{4.13}$$

Since the consistency condition (4.13) has to hold for all $i = 1, \dots, m - 1$ in all periods t, the law of iterated expectations implies

$$\mathbb{E}[y^e_{t,t+1}|\mathcal{F}_{t-i}] = y^e_{t-i,t+1} \quad \mathbb{P}-\text{a.s.} \tag{4.14}$$

for all $i = 1, \dots, m - 1$ and all $t \in \mathbb{N}$. Hence, if the consistency condition (4.13) holds for all forecasts in all periods prior to time t, then (4.12) is satisfied, provided that the most recent forecast $y^e_{t,t+1}$ has a vanishing conditional forecast error, that is, if

$$\mathbb{E}\Big[(y_{t+1} - y^e_{t,t+1})\Big|\mathcal{F}_t\Big] = 0 \quad \mathbb{P} - \text{a.s.}$$

In terms of forecasting rules $\psi = (\psi^{(1)}, \ldots, \psi^{(m)})$ of the form (4.4), the consistency conditions (4.13) can be expressed as follows. Let $\mathbb{E}_j[\cdot] \equiv \mathbb{E}[\cdot|\mathcal{F}_{j,r}]$ denote the conditional expectations operator with respect to the σ-algebra $\mathcal{F}_{j,r}$, where the length of the memory r may be arbitrarily long. In terms of a forecasting rule ψ, the consistency condition (4.13) is now rephrased as

$$\begin{aligned}\mathbb{E}_{t-1}[y^e_{t,t+i}](\omega) &= \mathbb{E}_{t-1}\left[\psi^{(i)}(\vartheta^t\cdot, G_\psi(t-1,\cdot,x_0,y^e_0)\right](\omega)\\ &= y^e_{t-1,t+i} \quad \mathbb{P} - \text{a.s.}\end{aligned} \tag{4.15}$$

for each $i = 1, \ldots, m-1$, $t \in \mathbb{N}$, where $(x_{t-1}, y^e_{t-1}) = G_\psi(t-1, \omega, x_0, y^e_0)$. Again, the requirement (4.15) states that for each $i = 1, \ldots, m-1$, the expected value of the new forecast $y^e_{t,t+i}$ conditional on $\mathcal{F}_{t,r}$ has to coincide with the respective old forecast $y^e_{t-1,t+i}$. Since $\mathcal{F}_{t,r} \subset \mathcal{F}_t$, condition (4.15) implies (4.13) and thus is weaker than the original consistency condition (4.13).

We seek a consistency notion which is time invariant. Recall to this end that by Lemma 4.1 of Section 4.5 $\mathcal{F}_{t,r} = \vartheta^{-t}\mathcal{F}_{0,r}$. Proposition 4.1 on page 81 in Section 4.5 then implies that

$$\begin{aligned}\mathbb{E}_{t-1}[y^e_{t,t+i}](\omega) &= \mathbb{E}_{t-1}\left[\psi^{(i)}(\vartheta^t\cdot, G(\xi(\vartheta^{t-1}\cdot), x_{t-1}, y^e_{t-1}), y^e_{t-1})\right](\omega)\\ &= \mathbb{E}_0\left[\psi^{(i)}(\vartheta\cdot, G(\xi(\cdot), x_{t-1}, y^e_{t-1}), y^e_{t-1})\right](\vartheta^{t-1}\omega),\end{aligned} \tag{4.16}$$

where $(x_{t-1}, y^e_{t-1}) = G_\psi(t-1, \omega, x_0, y^e_0)$. Combining (4.15) with (4.16), the desired notion of a consistent forecasting rule now takes the following form.

Definition 4.1. *A forecasting rule $\psi = (\psi^{(1)}, \ldots, \psi^{(m)})$ of type (4.4) is called* (locally) consistent *with respect to the sub-σ-algebra $\mathcal{F}_{0,r}$ of $\mathcal{F}$ if there exists a random set*[3] *$\mathcal{U} = \{\mathcal{U}(\omega)\}_{\omega\in\Omega}$, $\mathcal{U}(\omega) \subset \mathbb{X} \times \mathbb{Y}^m$, such that for each $i = 1, \ldots, m-1$,*

$$\mathbb{E}\left[\psi^{(i)}(\vartheta\cdot, G(\xi(\cdot), x, z), z)\middle|\mathcal{F}_{0,r}\right](\omega) = z^{(i+1)} \quad \mathbb{P} - a.s.$$

for all $(x, z) \in \mathcal{U}(\omega)$, where $z = (z^{(1)}, \ldots, z^{(m)})$. ψ is called globally consistent if $\mathcal{U}(\omega) = \mathbb{X} \times \mathbb{Y}^m$ $\mathbb{P}$-a.s.

A consistent forecasting rule as defined in Definition 4.1 is designed to satisfy (4.13) by construction on a possibly restricted random subset of the state space $\mathbb{X}\times\mathbb{Y}^m$. The definition contains a minimum requirement for a forecasting rule of the form (4.4) in order to generate best least-squares predictions. As a consequence of the law of iterated expectations, a forecasting rule that is not consistent cannot provide best least-squares predictions. Notice that the consistency requirement includes the case with infinite memory $\mathcal{F}_{0,\infty}$.

[3] A random set is simply a set-valued random variable $\mathcal{U} : \Omega \to \mathbf{P}(\mathbb{X} \times \mathbb{Y}^m)$, where $\mathbf{P}(\mathbb{X} \times \mathbb{Y}^m)$ denotes the space of all subsets of $\mathbb{X} \times \mathbb{Y}^m$, see Arnold (1998, Sec. 1.6, p. 32) for more details.

Example 4.2. *As already seen in Example 2.1, the simplest example of a consistent forecasting rule is a forecasting rule that never updates previous forecasts. A* no-updating *forecasting rule* $\psi = (\psi^{(1)}, \dots, \psi^{(m)})$ *is formally defined by setting for each* $i = 1, \dots, m-1$,

$$\psi^{(i)}(\omega, x, z) := z^{(i+1)} \quad \textit{for all } (x,z) \in \mathbb{X} \times \mathbb{Y}^m,\ \omega \in \Omega \tag{4.17}$$

and letting $\psi^{(m)}$ *be an arbitrary function depending on the history of the process. If* $x_t \in \mathbb{X}$ *is the current state of the economy at date* t *and* $y^e_{t-1} = (y^e_{t-1,t}, \dots, y^e_{t-1,t-1+m})$ *the vector of forecasts formed in the previous period* $t-1$, *then the current forecast* $y^e_t = (y^e_{t,t+1}, \dots, y^e_{t,t+m})$ *is*

$$\begin{cases} y^e_{t,t+i} &= y^e_{t-1,t+i}, \qquad i = 1, \dots, m-1, \\ y^e_{t,t+m} &= \psi^{(m)}(\vartheta^t\omega, x_t, y^e_{t-1}). \end{cases}$$

The forecasting rule (4.17) thus never updates forecasts formed at previous dates.

4.1.2 Unbiased Forecasting Rules

To see which forecasting rules generate best least-squares predictions, recall that

$$y_{t+1} = g\big(\xi(\vartheta^t\omega), x_t, y^e_t\big).$$

Invoking Proposition 4.1 (see Section 4.5, p. 81 below), the expected value of the future state y_{t+1} reads

$$\begin{aligned} \mathbb{E}_t[y_{t+1}](\omega) &= \mathbb{E}_t\big[g(\xi(\vartheta^t\cdot), x_t, y^e_t)\big](\omega) \\ &= \mathbb{E}_0\big[g(\xi(\cdot), x_t, y^e_t)\big](\vartheta^t\omega) \quad \mathbb{P}-\text{a.s.}, \end{aligned} \tag{4.18}$$

where $(x_t, y^e_t) = G_\psi(t, \omega, x_0, y^e_0)$. Given a vector of forecasts $y^e_t \in \mathbb{Y}^m$, the forecast error for y_{t+1} reads

$$y_{t+1} - y^e_{t,t+1} = g(\xi(\vartheta^t\omega), x_t, y^e_t) - y^e_{t,t+1},$$

while the forecast error conditional on information available at date t is

$$\mathbb{E}_t\big[(y_{t+1} - y^e_{t,t+1})\big](\omega) = \mathbb{E}_0\big[g(\xi(\cdot), x_t, y^e_t)\big](\vartheta^t\omega) - y^e_{t,t+1}. \tag{4.19}$$

Using (4.18), this gives rise to the definition of a *(conditional) mean error function* associated with the economic law G.

Definition 4.2. *Let* $G = (\overline{G}, g)$ *be an economic law. The* (conditional) mean error function *of the economic law* G *with respect to a given sub-*σ*-algebra* $\mathcal{F}_{0,r}$ *is a function* $e^{\mathbb{E}}_G : \Omega \times \mathbb{X} \times \mathbb{Y}^m \to \mathbb{R}^{d_y}$, *defined by*

$$e^{\mathbb{E}}_G(\omega, x, z) = \mathbb{E}\Big[g\big(\xi(\cdot), x, z\big)\Big|\mathcal{F}_{0,r}\Big](\omega) - z^{(1)}, \tag{4.20}$$

where $z = (z^{(1)}, \dots, z^{(m)})$.

Given an arbitrary state of the economy $x_t \in \mathbb{X}$ at some date t, the mean error function describes all possible mean forecast errors conditional on all information available at date t, regardless of which forecasting or learning rule has been applied. Given arbitrary forecasts $y_t^e = (y_{t,t+1}^e, \dots, y_{t,t+m}^e)$, the conditional forecast error (4.19) vanishes if and only if

$$\mathbb{E}_t[y_{t+1} - y_{t,t+1}^e](\omega) = e_G^{\mathbb{E}}(\vartheta^t\omega, x_t, y_t^e) = 0 \quad \mathbb{P}-\text{a.s.} \tag{4.21}$$

Geometrically, condition (4.21) is characterized by the zero-contour set $\mathcal{W}_G = \{\mathcal{W}_G(\omega)\}_{\omega\in\Omega}$ of the error function (4.20), given by the family of subsets

$$\mathcal{W}_G(\omega) := \Big\{(x, y^e) \in \mathbb{X}\times\mathbb{Y}^m \Big|\ e_G^{\mathbb{E}}(\omega, x, y^e) = 0\Big\}, \quad \omega \in \Omega.$$

The zero-contour set $\mathcal{W}_G$, also referred to as *constraint variety* of the economic law G, describes all states of the economy (ω, x, y^e) for which the conditional forecast error (4.19) vanishes. Let $\|\cdot\|$ be some norm on $\mathbb{R}^{d_y}$ and $\varepsilon \geq 0$ be a non-negative number. A random neighborhood $\mathcal{W}_G^\varepsilon := \big\{\mathcal{W}_G^\varepsilon(\omega)\big\}_{\omega\in\Omega}$ of the zero-contour set of the error function $e_G^{\mathbb{E}}$ is defined by setting

$$\mathcal{W}_G^\varepsilon(\omega) := \Big\{(x, y^e) \in \mathbb{X}\times\mathbb{Y}^m \Big|\ \big\|e_G^{\mathbb{E}}(\omega, x, y^e)\big\| \leq \varepsilon\Big\}, \quad \omega \in \Omega.$$

For $\varepsilon = 0$, we have $\mathcal{W}_G \equiv \mathcal{W}_G^0$. The constraint variety imposes a geometric constraint for an unbiased forecasting rule which will be defined next.

Definition 4.3. *Let G be an economic law. A forecasting rule $\psi = (\psi^{(1)}, \dots, \psi^{(m)})$ of type (4.4) is called* locally ε-unbiased *for G with respect to $\mathcal{F}_{0,r}$ if there exists a random set $\mathcal{U} = \{\mathcal{U}(\omega)\}_{\omega\in\Omega}$, $\mathcal{U}(\omega) \subset \mathbb{X}\times\mathbb{Y}^m$ such that*

$$G_\psi(\omega, x, y^e) = \begin{pmatrix} G(\xi(\omega), x, y^e) \\ \psi\big(\vartheta\omega, G(\xi(\omega), x, y^e), y^e\big) \end{pmatrix} \in \mathcal{W}_G^\varepsilon(\vartheta\omega)$$

for all $(x, y^e) \in \mathcal{U}(\omega)$, $\omega \in \Omega$. For $\varepsilon = 0$, locally ε-unbiased forecasting rules are called locally unbiased.

Let $\psi = (\psi^{(1)}, \dots, \psi^{(m)})$ be locally ε-unbiased in the sense of Definition 4.3 and $(x_{t-1}, y_{t-1}^e) \in \mathcal{U}(\vartheta^{t-1}\omega)$ for some time $t-1$. Then

$$\begin{pmatrix} x_t \\ y_t^e \end{pmatrix} = G_\psi(\vartheta^{t-1}\omega, x_{t-1}, y_{t-1}^e) \in \mathcal{W}_G^\varepsilon(\vartheta^t\omega),$$

implying that the conditional forecast error for

$$y_{t,t+1}^e = \psi^{(1)}(\vartheta^t\omega, x_t, y_{t-1}^e)$$

is less than ε, i.e.,

$$\big\|\mathbb{E}_t[y_{t+1} - y_{t,t+1}^e](\omega)\big\| = \big\|e_G^{\mathbb{E}}(\vartheta^t\omega, x_t, y_t^e)\big\| \leq \varepsilon \quad \mathbb{P}-\text{a.s.} \tag{4.22}$$

In other words, the forecasting rule ψ gives an approximate solution to (4.21) in the sense that $y^e_{t,t+1}$, up to an error of size ε, is the best least-squares prediction for y_{t+1} given the history $\{x_s, y^e_{s-1}\}_{s=t-r+1}^{t}$. If, in addition, a locally unbiased forecasting rule is consistent, then (4.22) implies

$$\left\|\mathbb{E}_{t-i}[y_{t+1} - y^e_{t-i,t+1}](\omega)\right\| = \left\|\mathbb{E}_{t-i}\left[\mathbb{E}_t[y_{t+1} - y^e_{t,t+1}]\right](\omega)\right\| \leq \varepsilon \quad \mathbb{P}-\text{a.s.}$$

for all $i = 1, \ldots, m-1$, so that the forecast errors of *all* forecasts $y^e_{t-i,t+1}$, $i = 0, \ldots, m-1$ is less that ε. This justifies the terminology locally ε-unbiased forecasting rule.

Let $\gamma(x_0, y^e_0) = \{x_t, y^e_t\}_{t\in\mathbb{N}}$ with $(x_t, y^e_t) = G_\psi(t, \omega, x_0, y^e_0)$ denote an orbit of the time-one map G_ψ corresponding to the economic law G and the forecasting rule ψ which starts at $(x_0, y^e_0) \in \mathbb{X} \times \mathbb{Y}^m$. A main problem with locally ε-unbiased forecasting rules is that the forecasting rule may well lose its property of being unbiased along such an orbit. This problem can be ruled out by imposing the following additional requirement. Let ψ be a forecasting rule that is locally ε-unbiased on a random set $\mathcal{U}$ in the sense of Definition 4.3 and assume that $\mathcal{U}(\omega) \subset \mathcal{W}^\varepsilon_G(\omega)$ for all $\omega \in \Omega$. Then the requirement that $\mathcal{U}$ is forward invariant under G_ψ, i.e.,

$$G_\psi(\omega, x, y^e) \in \mathcal{U}(\vartheta\omega) \quad \text{for all} \quad (x, y^e) \in \mathcal{U}(\omega),\ \omega \in \Omega \tag{4.23}$$

is sufficient for the existence of unbiased orbits. In this case, any orbit $\gamma(x_0, y^e_0)$ starting at $(x_0, y^e_0) \in \mathcal{U}(\omega)$ satisfies $(x_t, y^e_t) \in \mathcal{U}(\vartheta^t\omega) \subset \mathcal{W}^\varepsilon_G(\vartheta^t\omega)$ for all times t and for all perturbations ω and for this reason will be an ε-unbiased orbit. In other words, an ε-unbiased orbit is an orbit along which all forecasts are ε-unbiased. For $\varepsilon = 0$, ε-unbiased orbits are referred to as unbiased orbits. These correspond precisely to rational expectations equilibria (REE) in the classical sense. A forecasting rule that generates ε-unbiased orbits will be referred to as unbiased forecasting rule and is defined as follows.

Definition 4.4. *Let $G = (\overline{G}, g)$ be an economic law. A forecasting rule $\psi_\star = (\psi^{(1)}_\star, \ldots, \psi^{(m)}_\star)$ of type (4.4) is called* ε-unbiased *for G with respect to $\mathcal{F}_{0,r}$ if there exists a random set $\mathcal{U} = \{\mathcal{U}(\omega)\}_{\omega\in\Omega}$ with $\mathcal{U}(\omega) \in \mathbb{X} \times \mathbb{Y}^m$, such that the following conditions hold:*

(i) Consistency. For each $i = 1, \ldots, m-1$,

$$\mathbb{E}\left[\psi^{(i)}_\star(\vartheta\cdot, G(\xi(\cdot), x, z), z)\middle|\mathcal{F}_{0,r}\right](\omega) = z^{(i+1)} \quad \mathbb{P}-a.s.$$

for all $(x, z) \in \mathcal{U}(\omega)$, where $z = (z^{(1)}, \ldots, z^{(m)})$.

(ii) Local Unbiasedness. There exists a non-negative number $\varepsilon \geq 0$ such that

$$G_{\psi_\star}(\omega, x, z) = \begin{pmatrix} G(\xi(\omega), x, z) \\ \psi_\star\big(\vartheta\omega, G(\xi(\omega), x, z), z\big) \end{pmatrix} \in \mathcal{W}^\varepsilon_G(\vartheta\omega)$$

for all $(x, z) \in \mathcal{U}(\omega)$, $\omega \in \Omega$.

(iii) Invariance. $\mathcal{U}(\omega) \subset \mathcal{W}_G^{\varepsilon}(\omega)$ *for all* $\omega \in \Omega$ *and*

$$G_{\psi_\star}(\omega, x, z) \in \mathcal{U}(\vartheta\omega) \quad \text{for all } (x, z) \in \mathcal{U}(\omega),\ \omega \in \Omega.$$

For $\varepsilon = 0$, ε*-unbiased forecasting rules are called* unbiased.

Definition 4.4 implies that a forecasting rule $\psi_\star$ is ε-unbiased if and only if *all* orbits $\gamma(x_0, y_0^e)$ starting in $(x_0, y_0^e) \in \mathcal{U}(\omega)$ of $G_{\psi_\star}$ are ε-unbiased orbits. While Conditions (i) and (ii) state that the forecasts are locally unbiased, it is the invariance condition (iii) that guarantees that the forecasts of $\psi_\star$ remain unbiased through the course of time. A forecasting rule that generates ε-unbiased orbits for arbitrary initial values $x_0 \in \mathbb{X}$ is defined as follows.

Definition 4.5. *Under the hypotheses of Definition 4.4, a forecasting rule* $\psi_\star$ *of type (4.4) is called* globally ε-unbiased *for* G *with respect to* $\mathcal{F}_{0,r}$ *if, in addition, the following conditions holds true:*

(iv) Global existence. For each initial state $x_0 \in \mathbb{X}$ *and each perturbation* $\omega \in \Omega$*, there exists a vector of forecasts* $y_0^e \in \mathbb{Y}^m$ *such that* $(x_0, y_0^e) \in \mathcal{U}(\omega)$.

The conditions in Definitions 4.4 and 4.5 are clearly quite demanding. In view of economic applications and the learning issue treated in these notes, it will often be sufficient to have good forecasts in the long run. We will therefore relax the conditions of Definition 4.4 when we discuss the long-run behavior of a random dynamical system in Section 4.4 and conclude this section with two remarks.

Remark 4.2. *A particular case arises when the* g*-part of the economic law (4.2) is* $\mathcal{F}_{0,r}$ *measurable. Then the mean error function takes the form*

$$e_G^{\mathbb{E}}(\omega, x, z) = g\big(\xi(\omega), x, z\big) - z^{(1)}$$

and coincides essentially with the deterministic error function in Böhm & Wenzelburger (1999, 2002). An unbiased no-updating forecasting rule would generate perfect foresight in the classical sense. We will make use of this observation in Chapter 6.

Remark 4.3. *For a finite memory* $r < \infty$*, the factorization lemma and Proposition 4.1 below yield the existence of an error function*

$$\mathcal{E}_G : \mathbb{Y}^m \times \mathbb{X} \times \Sigma \to \mathbb{R}^{md_y}$$

as defined in Chapter 2, such that for each $t \in \mathbb{N}$ *and each* $\omega \in \Omega$

$$e_G^{\mathbb{E}}(\vartheta^t\omega, x_t, y_t^e) = \mathcal{E}_G(y_t^e, x_t, \dots, y_{t-r}^e, x_{t-r}) \quad \mathbb{P} - a.s.$$

along any orbit $\{x_t, y_t^e\}_{t\in\mathbb{N}}$*. This implies that an unbiased forecasting rule* $\psi_\star$ *in the sense of Definition 4.4 induces an unbiased forecasting rule*

$$\Psi_\star : \mathbb{X} \times \Sigma \to \mathbb{Y}^m$$

of the form (4.11), such that

$$\mathcal{E}_G(\Psi_\star(x_t, y^e_{t-1}, \dots, y^e_{t-r}, x_{t-r}), x_t, y^e_{t-1}, x_{t-1}, \dots, y^e_{t-r}, x_{t-r}) = 0$$

on a suitably defined subset of $\mathbb{X} \times \Sigma$.

4.2 Existence of Unbiased Forecasting Rules

At first sight, the problem of existence and uniqueness of unbiased no-updating forecasting rules for systems with an expectational lead looks much like the deterministic case treated in Böhm & Wenzelburger (2004). This is due to the fact that the mean error function defined in Definition 4.2 is a static and hence time-invariant object the properties of which decide whether or not unbiased forecasting rules exist. As in the deterministic case, the existence and uniqueness problem is encoded in the geometry of the zero-contour set of the error function. The stochastification of the deterministic case, however, requires some subtle adjustments which will be developed in this section.

The existence of unbiased forecasting rules is based on the following geometric observation. Let $\mathcal{V} = \{\mathcal{V}(\omega)\}_{\omega\in\Omega}$ with $\mathcal{V}(\omega) \subset \mathbb{X} \times \mathbb{Y}^{m-1}$, $\omega \in \Omega$ be a random set and assume that for each $\omega \in \Omega$, there exists a function $H(\omega, \cdot) : \mathcal{V}(\omega) \to \mathbb{Y}$ such that

$$e^{\mathbb{E}}_G\big(\omega, x, \eta, H(\omega, x, \eta)\big) = 0 \quad \text{for all} \quad (x, \eta) \in \mathcal{V}(\omega) \quad \mathbb{P}-\text{a.s.} \tag{4.24}$$

Assume, in addition, that each $H(\cdot, x, \eta)$ is $\mathcal{F}_{0,r}/\mathcal{B}(\mathbb{Y})$ measurable. The idea now is to construct an unbiased no-updating forecasting rule

$$\psi_\star = (\psi^{(1)}_\star, \dots, \psi^{(m)}_\star) : \Omega \times \mathbb{X} \times \mathbb{Y}^m \to \mathbb{Y}^m, \tag{4.25}$$

using the function H. Since in this case all $\psi^{(i)}_\star$, $i = 1, \dots, m-1$, are projections, we are left to define the forecasting rule $\psi^{(m)}_\star$.

To this end, let $x_t \in \mathbb{X}$ be the current state of the economy at date t and $y^e_{t-1} = (y^e_{t-1,t}, \dots, y^e_{t-1,t-1+m})$ the vector of forecasts formed in the previous period $t-1$ such that

$$\big(x_t, (y^e_{t-1,t+1}, \dots, y^e_{t-1,t-1+m})\big) \in \mathcal{V}(\vartheta^t\omega).$$

The forecast $y^e_t = (y^e_{t,t+1}, \dots, y^e_{t,t+m})$ formed in period t can now be obtained by setting

$$\begin{aligned} y^e_{t,t+i} &= \psi^{(i)}_\star(\vartheta^t\omega, x_t, y^e_{t-1}) := y^e_{t-1,t+i}, \qquad i = 1, \dots, m-1, \\ y^e_{t,t+m} &= \psi^{(m)}_\star(\vartheta^t\omega, x_t, y^e_{t-1}) := H\big(\vartheta^t\omega, x_t, \mathrm{pr}_{-1} y^e_{t-1}\big), \end{aligned}$$

where

$$\mathrm{pr}_{-1} : \mathbb{Y}^m \to \mathbb{Y}^{m-1}, \quad (z^{(1)}, \dots, z^{(m)}) \mapsto (z^{(2)}, \dots, z^{(m)}).$$

This defines $\psi_\star(\omega, \cdot)$ on $\mathcal{V}(\omega) \times \mathbb{Y}$ for each perturbation $\omega \in \Omega$. Outside of $\mathcal{V}$, $\psi_\star^{(m)}$ may be an arbitrary function.

Since $\psi_\star$ never updates forecasts formed at previous dates, it is a consistent rule in the sense of Definition 4.1. Moreover, $\psi_\star$ is locally unbiased in the sense of Definition 4.3, because the function H was assumed to satisfy (4.24). For this reason the conditional forecast error vanishes, that is,

$$\mathbb{E}_t[y_{t+1} - y^e_{t,t+1}](\omega) = e^{\mathbb{E}}_G\big(\vartheta^t\omega, x_t, \psi_\star(\vartheta^t\omega, x_t, y^e_{t-1})\big) = 0 \quad \mathbb{P}-\text{a.s.}$$

By construction of (4.25) all forecasts for y_{t+1} fulfill

$$\mathbb{E}_t[y_{t+1} - y^e_{t-i,t+1}] = 0 \quad \mathbb{P}-\text{a.s.}, \quad i = 0, \dots, m-1, \tag{4.26}$$

as long as the *no-updating rule* (4.25) has been applied for the past m periods. This shows that ψ_* is locally unbiased in the sense of Definition 4.3, provided that (4.24) holds.

The question of existence of an unbiased no-updating rule can now be addressed as follows. Consider the function $\mathbf{E}_G$ induced by $e^{\mathbb{E}}_G$, which is defined by

$$\mathbf{E}_G : \begin{cases} \Omega \times \mathbb{X} \times \mathbb{Y}^m \longrightarrow & \Omega \times \mathbb{X} \times \mathbb{Y}^{m-1} \times \mathbb{R}^{d_y} \\ \quad (\omega, x, z) \longmapsto & \big(\omega, x, \mathrm{pr}_{-m} z,\ e^{\mathbb{E}}_G(\omega, x, z)\big) \end{cases}, \tag{4.27}$$

where

$$\mathrm{pr}_{-m} : \mathbb{Y}^m \to \mathbb{Y}^{m-1}, \quad (z^{(1)}, \dots, z^{(m)}) \mapsto (z^{(1)}, \dots, z^{(m-1)}).$$

If the *induced error function* (4.27) is globally invertible, then (4.27) has an inverse function $\mathbf{H}$ which must be of the functional form

$$\mathbf{H} : \begin{cases} \Omega \times \mathbb{X} \times \mathbb{Y}^{m-1} \times \mathbb{R}^{d_y} & \to \Omega \times \mathbb{X} \times \mathbb{Y}^{m-1} \times \mathbb{Y} \\ \quad (\omega, x, \eta, \zeta) & \mapsto \big(\omega, x, \eta, H_0(\omega, x, \eta, \zeta)\big) \end{cases} \tag{4.28}$$

with some function $H_0 : \Omega \times \mathbb{X} \times \mathbb{Y}^{m-1} \times \mathbb{R}^{d_y} \to \mathbb{Y}$. If the inverse function $\mathbf{H}$ exists, then

$$e^{\mathbb{E}}_G\big(\omega, x, \eta, H_0(\omega, x, \eta, 0)\big) = 0 \quad \text{for all} \quad (\omega, x, \eta) \in \Omega \times \mathbb{X} \times \mathbb{Y}^{m-1}.$$

Defining

$$H(\omega, x, \eta) := H_0(\omega, x, \eta, 0), \quad (\omega, x, \eta) \in \Omega \times \mathbb{X} \times \mathbb{Y}^{m-1},$$

this implies

$$e^{\mathbb{E}}_G\big(\omega, x, \eta, H(\omega, x, \eta)\big) = 0, \quad (\omega, x, \eta) \in \Omega \times \mathbb{X} \times \mathbb{Y}^{m-1}.$$

If $\mathcal{V}(\omega) = \mathbb{X} \times \mathbb{Y}^{m-1}$, $\omega \in \Omega$, then the forecasting rule $\psi_\star$ as defined in (4.25) is unbiased in the sense of Definition 4.4 globally on $\mathbb{X} \times \mathbb{Y}^m$.

The question whether unbiased no-updating rules of the form (4.25) exist is therefore reduced to the question whether the induced error function (4.27) is invertible. Invoking the Inverse Function Theorem or, equivalently, the Implicit Function Theorem, it is now possible to provide sufficient conditions for the existence of a unique unbiased no-updating forecasting rule for an economic law G.

Theorem 4.1. *Let $G = (\overline{G}, g)$ be an economic law. Suppose that the following conditions hold:*

(i) There exists a $\mathcal{F}_{0,r}/\mathcal{B}(\mathbb{X})$ measurable random variable $x_0 : \Omega \to \mathbb{X}$ and a $\mathcal{F}_{0,r}/\mathcal{B}(\mathbb{Y}^m)$ measurable random variable $z_0 : \Omega \to \mathbb{Y}^m$, such that

$$e_G^{\mathbb{E}}\big(\omega, x_0(\omega), z_0(\omega)\big) = 0 \quad \mathbb{P} - a.s.$$

(ii) For each $\omega \in \Omega$, the function

$$h(\omega, x, z) := \mathbb{E}[g(\xi(\cdot), x, z)|\mathcal{F}_{0,r}](\omega)$$

is continuously differentiable with respect to $x \in \mathbb{X}$ and $z = (z^{(1)}, \ldots, z^{(m)}) \in \mathbb{Y}^m$; the derivative $\mathrm{D}_{z^{(m)}} h\big(\omega, x_0(\omega), z_0(\omega)\big)$ is invertible for all $\omega \in \Omega$, where $\mathrm{D}_{z^{(m)}}$ denotes the derivative with respect to $z^{(m)}$.

Then there exists a family of open sets $\mathcal{V} = \{\mathcal{V}(\omega)\}_{\omega \in \Omega}$ with $\mathcal{V}(\omega) \subset \mathbb{X} \times \mathbb{Y}^{m-1}$, $\omega \in \Omega$ and functions $H(\omega, \cdot) : \mathcal{V}(\omega) \to \mathbb{Y}$, $\omega \in \Omega$ satisfying

$$H\big(\omega, x_0(\omega), \mathrm{pr}_{-1} z_0(\omega)\big) = z_0^{(m)}(\omega), \quad \omega \in \Omega,$$

such that

$$e_G^{\mathbb{E}}(\omega, x, \eta, H\big(\omega, x, \eta)\big) = 0 \quad \textit{for all} \ \ (x, \eta) \in \mathcal{V}(\omega) \quad \mathbb{P} - a.s.$$

Moreover, the no-updating rule $\psi_\star$ given in (4.25) is locally unbiased in the sense of Definition 4.3 and each $\psi_\star(\cdot, x, y^e)$, $(x, y^e) \in \mathbb{X} \times \mathbb{Y}^m$ is $\mathcal{F}_{0,r}/\mathbb{B}(\mathbb{Y}^m)$ measurable.

Proof. The existence of the map H and thus of $\psi_\star$ follows from a ω-wise application of the Implicit Function Theorem, e.g., see Lang (1968, p. 355). The measurability property of $\psi_\star$ follows from the fact that each

$$e_G^{\mathbb{E}}(\cdot, x, z), \quad (x, z) \in \mathbb{X} \times \mathbb{Y}^m$$

is $\mathcal{F}_{0,r}/\mathbb{B}(\mathbb{R}^{d_y})$ measurable. □

To formulate a global existence theorm we will invoke what in the mathematical literature is known as a Global Inverse Function Theorem. Recall to this end the notion of a simply connected set. Loosely speaking, a simply connected set contains no holes. More formally, a set $\mathbb{X} \subset \mathbb{R}^d$ is *simply connected* if every closed path $p : [0,1] \to \mathbb{X}$, $p(0) = p(1)$ is homotopic to a constant path, that is if there exists a continuous function $P : [0,1] \times [0,1] \to \mathbb{X}$ with $P(t,0) = p(t)$ and $P(t,1) \equiv x_0 \in \mathbb{X}$ for all $t \in [0,1]$, cf. Lipschutz (1965, p. 186). Simple examples for simply connected sets are $\mathbb{X} = \mathbb{R}^{d_x}$ and $\mathbb{Y} = \mathbb{R}^{d_y}$.

Theorem 4.2. *Under the hypotheses of Theorem 4.1, assume that the following conditions hold:*

(i) For each $\omega \in \Omega$, $h(\omega,\cdot) \in C^1(\mathbb{X} \times \mathbb{Y}^m, \mathbb{Y})$ *and* $\mathrm{D}_{z^{(m)}} h(\omega,x,z)$ *is invertible for all* $(\omega,x,z) \in \Omega \times \mathbb{X} \times \mathbb{Y}^m$.

(ii) For each $\omega \in \Omega$, *there exist positive constants* $\alpha(\omega)$ *and* $\beta(\omega)$ *such that*

$$\|\mathrm{D}_{z^{(m)}} h(\omega,x,z)^{-1} \widehat{\mathrm{D}} e_G^{\mathbb{E}}(\omega,x,z)\| + \|\mathrm{D}_{z^{(m)}} h(\omega,x,z)^{-1}\| \le \alpha(\omega)\|(x,z)\| + \beta(\omega)$$

$\mathbb{P}$*-a.s. for all* $(x,z) \in \mathbb{X} \times \mathbb{Y}^m$, *where* $\widehat{\mathrm{D}}$ *denotes the derivative with respect to the variables* $(x, z^{(1)}, \ldots, z^{(m-1)})$.

(iii) $\mathbb{X}$ *and* $\mathbb{Y}$ *are simply connected.*

Then all conclusions of Theorem 4.1 hold and a unique no-updating rule

$$\psi_\star : \Omega \times \mathbb{X} \times \mathbb{Y}^m \to \mathbb{Y}^m$$

as defined in (4.25) exists, which is unbiased in the sense of Definition 4.5, globally on $\mathbb{X} \times \mathbb{Y}^m$.

Proof. We only have to assure that (4.28) exists. The derivative of $\mathbf{E}_G$ with respect to $(x,z) \in \mathbb{X} \times \mathbb{Y}^m$ takes the (block matrix) form

$$\mathrm{D}\mathbf{E}_G(\omega,x,z) = \begin{pmatrix} I & 0 \\ \widehat{\mathrm{D}} e_G^{\mathbb{E}}(\omega,x,z) & \mathrm{D}_{z^{(m)}} h(\omega,x,z) \end{pmatrix}.$$

Indeed, $\mathrm{D}\mathbf{E}_G(\omega,x,z)$ is invertible if and only if $D_{z^{(m)}} h(\omega,x,z)$ is invertible. The inverse is

$$\mathrm{D}\mathbf{E}_G(\omega,x,z)^{-1} = \begin{pmatrix} I & 0 \\ -\mathrm{D}_{z^{(m)}} h(\omega,x,z)^{-1} \widehat{\mathrm{D}} e_G^{\mathbb{E}}(\omega,x,z) & \mathrm{D}_{z^{(m)}} h(\omega,x,z)^{-1} \end{pmatrix}.$$

Using the respective matrix norms,

$$\|\mathrm{D}\mathbf{E}_G(\omega,x,z)^{-1}\| \le 1 + \|\mathrm{D}_{z^{(m)}} h(\omega,x,z)^{-1} \widehat{\mathrm{D}} e_G^{\mathbb{E}}(\omega,x,z)\| + \|\mathrm{D}_{z^{(m)}} h(\omega,x,z)^{-1}\|.$$

Conditions (i) and (ii) are precisely the two hypotheses needed for a Global Inverse Function Theorem (Deimling 1980, Thm. 15.4, p. 153) to be applicable, whereas Condition (iii) is precisely that additional requirement needed for the case in which $\mathbb{X}$ and $\mathbb{Y}$ are not vector spaces. This guarantees the existence of a unique globally defined inverse function $\mathbf{H} \equiv \mathbf{E}_G^{-1}$. □

It is straightforward to generalize Theorem 4.2 to the case of ε-unbiased forecasting rules. Theorem 4.2 reduces the problem of existence and uniqueness of an unbiased forecasting rule to the problem of global invertibility of the induced error function. Condition (i) in Theorem 4.2 provides a local invertibility criterion for the induced error function (4.27), whereas (ii) is an upper bound sufficient for the existence of a unique global inverse and (iii) is a topological criterion. Notice that for the existence of an unbiased forecasting rule it suffices to have invertibility of (4.27) on an open neighborhood of $\mathbb{X} \times \mathbb{Y}^{m-1} \times \{0\}$ in the image of $\mathbf{E}_G$. By virtue of the Inverse Function Theorem (see also Gale & Nikaido 1965), locally as well as globally unbiased forecasting rules are uniquely determined by the properties of $\mathcal{W}_G$ and thus by the fundamentals of the economy. This stipulates the functional form of an unbiased forecasting rule.

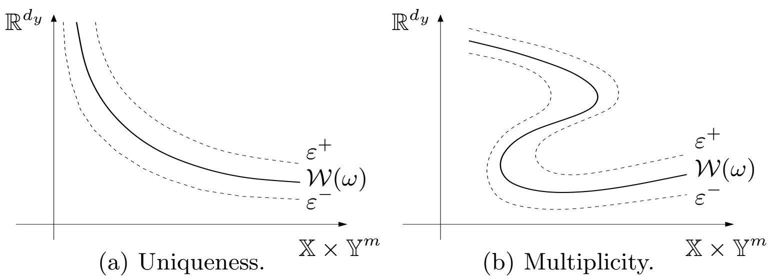

Fig. 4.1. Constraint varieties.

If the induced error function (4.27) is not globally invertible, it may happen that there exist multiple local inverses and thus several locally unbiased no-updating forecasting rules. Uniqueness and existence of unbiased forecasting rules are therefore encoded in the geometry of the zero-contour set $\mathcal{W}_G$, as graphically stylized in Figure 4.1. Topological methods to analyze the structure of $\mathcal{W}_G$ along with a global implicit function theorem can, for instance, be found in Rheinboldt (1969). However, even if all conditions of a theorem like Theorem 4.2 hold, for economic laws $G = (\overline{G}, g)$ with non-linear g–parts, it will, in general, be impossible to construct an unbiased forecasting rule $\psi_\star$ analytically.

Recall the remarkable feature of economic systems with expectational leads ($m > 1$). An unbiased no-updating forecasting rule (4.25), if it exists, generates rational expectations in the sense that *all* forecast errors (4.26) conditional on information available at date t vanish, including those for forecasts which have been made prior to that date, see also Theorem 2.1 of Chapter 2.

The no-updating rule (4.25) is designed in such a way that $\mathbb{E}_t[y_{t+1}] = y^e_{t,t+1}$ holds essentially for all choices of $y^e_{t,t+i}$, $i = 1, \dots, m-1$. The idea is to choose $y^e_{t,t+m}$ in such a way that the forecast error for $y^e_{t,t+1}$ is minimal. This implies a considerable freedom in selecting the first $m-1$ forecasts because any martingale difference sequence could, in principle, be added to these forecasts without violating the unbiasedness. Apart from the geometry of $\mathcal{W}_G$, this fact constitutes a primary cause for the emergence of multiple unbiased forecasting rules and thus of multiple rational expectations equilibria in the presence of expectational leads.

4.3 MSV Predictors

In linear models the notion of minimal-state-variable solutions introduced by McCallum (1983, 1998, 1999) is a popular concept to describe situations in which agents have rational expectations along particular trajectories of the system. In Chapter 3, Section 3.2.2, we showed in a linear context that MSV predictors may be seen as generators for minimal-state-variable solutions. We will now specialize the notion of a MSV predictor to non-linear economic laws of the form (4.2).

By virtue of Theorem 4.1, the existence of a set $\mathcal{U}$ on which unbiased forecasts are possible depends solely on the economic law G. This holds in particular for any subset $\mathcal{V} \subset \mathcal{U}$ which is forward invariant under $G_{\psi_\star}$ because the restriction $\psi_\star|_{\mathcal{V}}$ of a locally unbiased forecasting rule is uniquely determined by $\mathcal{W}_G$. Hence, the system $G_{\psi_\star}$ when restricted to the invariant set $\mathcal{V}$ is already determined by the economic fundamentals and, in principle, may exhibit any type of dynamic behavior. As pointed out in the previous section, initial conditions have to lie in a forward-invariant set $\mathcal{V}$ of $G_{\psi_\star}$ in order to guarantee rational expectations along the whole orbit.

There is a considerable freedom in choosing the initial forecasts $y^e_{0,i}$ $i = 1, \dots, m$, for an initial state $x_0 \in \mathbb{X}$ of the economy. In particular, different initial forecasts could, in principle, lead to different long-run outcomes of the economy even under rational expectations. To eliminate this degree of freedom, we will derive forecasting rules from a random dynamical system as follows. Let ϕ be a random dynamical system on $\mathbb{X}$ over a metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$, given by

$$x_{t+1} = \phi(1, \vartheta^t\omega, x_t) = \phi(t+1, \omega, x_0), \quad x_0 \in \mathbb{X}. \tag{4.29}$$

One may assume here that (4.29) describes a *perceived law of motion* in which a forecasting agency believes. Let $\mathcal{V} := \{\mathcal{V}(\omega)\}_{\omega\in\Omega}$ be a random set which is forward invariant under ϕ, i.e.,

$$\phi(1, \omega, x) \in \mathcal{V}(\vartheta\omega), \quad \text{for all } \ x \in \mathcal{V}(\omega)$$

and

$$\mathcal{F}^\phi_t := \sigma\big(\phi(s, \cdot, x)\big|\ s \le t, x \in \mathbb{X}\big),\ t \in \mathbb{Z}$$

be sub-σ-algebras pertaining to (4.29). Define forecasting rules $\varphi^{(i)}_{\mathbb{X}} : \Omega \times \mathbb{X} \to \mathbb{X}$, $i = 1, \dots, m$, by setting

$$\varphi_{\mathbb{X}}^{(i)}(\omega, x) := \mathbb{E}[\phi(i, \cdot, x)|\mathcal{F}_0^{\phi}](\omega), \quad i = 1, \ldots, m. \tag{4.30}$$

As before, let $x_t \in \mathbb{X}$ denote the state of the economy at date t. Using (4.30), we let

$$x_{t,t+i}^e := \varphi_{\mathbb{X}}^{(i)}(\vartheta^t \omega, x_t), \quad i = 1, \ldots, m$$

be the period-t forecasts for x_{t+i}, $i = 1, \ldots, m$, respectively, which are derived from the perceived law (4.29). By Proposition 4.1 (see Section 4.5 below),

$$x_{t,t+i}^e = \mathbb{E}[\phi(i, \vartheta^t \cdot, x_t)|\mathcal{F}_t^{\phi}](\omega) = \mathbb{E}[x_{t+i}|\mathcal{F}_t^{\phi}](\omega) \quad \mathbb{P}-\text{a.s.} \tag{4.31}$$

for all $i = 1, \ldots, m$. If x_{t+i} is thought of being generated from the perceived law (4.29), then the forecasts $x_{t,t+i}^e$, $i = 1, \ldots, m$, become the best least-squares predictions, given the history of realizations $\{x_s\}_{s \le t}$. By (4.31) and the law of iterated expectations, each forecast derived from (4.30) satisfies

$$\mathbb{E}[x_{t,t+i}^e|\mathcal{F}_{t-1}^{\phi}](\omega) = \mathbb{E}[x_{t+i}|\mathcal{F}_{t-1}^{\phi}](\omega) = \varphi_{\mathbb{X}}^{(i+1)}(\vartheta^{t-1}\omega, x_{t-1}) \tag{4.32}$$

for all $i = 1, \ldots, m$. This implies that $\varphi_{\mathbb{X}} = (\varphi_{\mathbb{X}}^{(1)}, \ldots, \varphi_{\mathbb{X}}^{(m)})$ is a consistent forecasting rule in the sense of Definition 4.1.

To construct a forecasting rule for an economic law G as defined in (4.2), let $\pi_{\mathbb{Y}}^{\mathbb{X}} : \mathbb{X} \to \mathbb{Y}$, $(\overline{x}, y) \mapsto y$ denote the natural projection onto $\mathbb{Y}$. For each $i = 1, \ldots, m$, let

$$y_{t,t+i}^e = \varphi^{(i)}(\vartheta^t \omega, x_t) := \pi_{\mathbb{Y}}^{\mathbb{X}}\big(\varphi_{\mathbb{X}}^{(i)}(\vartheta^t \omega, x_t)\big) \tag{4.33}$$

be the forecast for y_{t+i}, $i = 1, \ldots, m$, respectively. Since the forecasting rule $\varphi := (\varphi^{(1)}, \ldots, \varphi^{(m)})$ defined in (4.33) uses a minimal amount of endogenous variables, it will be referred to as a *minimal-state-variable* predictor (MSV predictor) for an economic law G of the form (4.2). It follows from (4.32) that φ is a consistent forecasting rule in the sense of Definition 4.1. φ is locally unbiased in the sense of Definition 4.3 if

$$e_G^{\mathbb{E}}\big(\omega, x, \varphi(\omega, x)\big) = 0 \quad \text{for all} \;\; x \in \mathcal{V}(\omega) \quad \mathbb{P}-\text{a.s.} \tag{4.34}$$

For each $\omega \in \Omega$, let

$$\Gamma_{\varphi}(\omega) := \{(x, z) \in \mathcal{V}(\omega) \times \mathbb{Y}^m |\; z = \varphi(\omega, x)\}$$

denote the graph of the map $\varphi(\omega, \cdot)$ and denote by G_{φ} the time-one map pertaining to the economic law G and the MSV predictor φ. The predictor φ is then unbiased in the sense of Definition 4.4 if Γ_{φ} is forward invariant under G_{φ}. Notice again that (4.34) incorporates an invertibility condition of the error function as discussed in Section 4.2.

A particular case arises when

$$\phi(1, \omega, x) = G\big(\xi(\omega), x, \varphi(\omega, x)\big), \quad \text{for all} \;\; x \in \mathcal{V}(\omega) \quad \mathbb{P}-\text{a.s.} \tag{4.35}$$

Then the additional degree of freedom in the unbiasedness condition (4.34), which requires coincidence of objective and subjective first moments only, is eliminated and each orbit of the perceived law (4.29) starting in $\mathcal{V}(\omega)$ corresponds to an orbit of G_φ. These orbits correspond precisely to *minimal-state-variable solutions* in the sense of McCallum (1983). The three relationships (4.29), (4.33), and (4.35) may also be seen to define a *functional rational expectations equilibrium* in the sense of Spear (1988). This observation allows us to interpret minimal state variable solutions and functional rational expectations equilibria as being generated from a particular kind of unbiased forecasting rules, referred to as unbiased MSV predictors.

The two conditions (4.33) and (4.35) constitute a nonlinear equation on an appropriately defined function space for a time-one map (4.29) which determines functional rational expectations equilibria (minimal state variable solutions). In general, such an equation will be much harder to solve than finding the inverse of an error function on a finite-dimensional space. Recently, Böhm, Kikuchi & Vachadze (2006) used a numerical method to comupte a MSV predictor for a particular model. Unbiased MSV predictors will most likely not be uniquely determined, as the analysis of the linear case treated in Chapter 3 indicates. However, they retain their importance as a possibility of singling out unbiased forecasting rules under which the resulting rational-expectations dynamics are stable.

4.4 Random Fixed Points

For many economic applications it suffices to have good forecasts in the long run, that is, ε–unbiased forecasts on special (random) sets or along special orbits of the system. This is the case when the set $\{\mathcal{U}(\omega)\}_{\omega\in\Omega}$ appearing in Definition 4.4 contains a random attractor. Random attractors are the random analogue of attractors for deterministic dynamical systems (see Arnold 1998, p. 483). Each orbit starting from the corresponding domain of attraction will then eventually become an ε-unbiased orbit. Otherwise an orbit may lose this property. Typical candidates for these special orbits are generated by asymptotically stable *random fixed points* of an ERDS. The following definition of a random fixed point (Schmalfuß 1996, 1998)[4] includes a stability notion given as Definition 7.4.6 in Arnold (1998).

Definition 4.6. *A* random fixed point *of the ERDS G_ψ is a random variable $(x_\star, y^e_\star) : \Omega \to \mathbb{X} \times \mathbb{Y}^m$, $\omega \mapsto \big(x_\star(\omega), y^e_\star(\omega)\big)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ such that*

$$\big(x_\star(\vartheta\omega), y^e_\star(\vartheta\omega)\big) = G_\psi\big(1, \omega, x_\star(\omega), y^e_\star(\omega)\big) \quad \textit{for all} \ \ \omega \in \Omega',$$

[4]Random fixed points can also be defined via a special class of invariant measures, also referred to as *random Dirac measures*, see Arnold (1998, p. 25).

where $\Omega' \subset \Omega$ is a ϑ-invariant subset of full measure $\mathbb{P}(\Omega') = 1$.[5]

A random fixed point $(x_\star, y_\star^e)$ is called asymptotically stable *with respect to a norm $\|\cdot\|$ on $\mathbb{X} \times \mathbb{Y}^m$ if there exists a random neighborhood $\mathcal{U}(\omega) \subset \mathbb{X} \times \mathbb{Y}^m$, $\omega \in \Omega$ such that*

$$\lim_{t\to\infty} \|G_\psi\big(t, \omega, x_0(\omega), y_0^e(\omega)\big) - \big(x_\star(\vartheta^t\omega), y_\star^e(\vartheta^t\omega)\big)\| = 0$$

for all $\big(x_0(\omega), y_0^e(\omega)\big) \in \mathcal{U}(\omega)$ $\mathbb{P}$ – a.s.

The first part in Definition 4.6 implies

$$\big(x_\star(\vartheta^{t+1}\omega), y_\star^e(\vartheta^{t+1}\omega)\big) = G_\psi\big(1, \vartheta^t\omega, x_\star(\vartheta^t\omega), y_\star^e(\vartheta^t\omega)\big)$$

for all times t. Hence, a random fixed point generates orbits

$$\big\{\big(x_\star(\vartheta^t\omega), y_\star^e(\vartheta^t\omega)\big)\big\}_{t\in\mathbb{N}}, \quad \omega \in \Omega$$

which solve the random difference equation

$$(x_{t+1}, y_{t+1}^e) = G_\psi(1, \vartheta^t\omega, x_t, y_t^e), \quad (x_0, y_0^e) \in \mathbb{X} \times \mathbb{Y}^m$$

induced by (4.5). If G_ψ is independent of the perturbation ω, then Definition 4.6 coincides with the notion of a deterministic fixed point. A random fixed point $(x_\star, y_\star^e)$ is asymptotically stable if for almost all perturbations $\omega \in \Omega$ all orbits starting at sufficiently close points $\big(x_0(\omega), y_0^e(\omega)\big) \in \mathcal{U}(\omega)$ eventually converge to orbits of the random fixed point. By stationarity and ergodicity of ϑ, the process $\{\big(x_\star(\vartheta^t\cdot), y_\star^e(\vartheta^t\cdot)\big)\}_{t\in\mathbb{N}}$ is stationary and ergodic. Let $\mathbb{P}_\star$ denote the probability distribution of $(x_\star, y_\star^e)$. If, in addition, $\mathbb{E}\left[\|(x_\star, y_\star^e)\|\right] < \infty$, then the ergodicity and the stability property of $(x_\star, y_\star^e)$ imply that

$$\begin{aligned}
&\lim_{T\to\infty} \tfrac{1}{T+1} \sum_{t=0}^{T} \chi_B\big(G_\psi(t, \omega, x_0(\omega), y_0^e(\omega))\big) \\
&= \lim_{T\to\infty} \tfrac{1}{T+1} \sum_{t=0}^{T} \chi_B\big(x_\star(\vartheta^t\omega), y_\star^e(\vartheta^t\omega)\big) \\
&= \mathbb{P}_\star(B)
\end{aligned}$$

for all $\big(x_0(\omega), y_0^e(\omega)\big) \in \mathcal{U}(\omega)$ $\mathbb{P}$-a.s. In other words, the empirical law of any 'nearby' orbit $\{G_\psi\big(t, \omega, x_0(\omega), y_0^e(\omega)\big)\}_{t\in\mathbb{N}}$ is for $\mathbb{P}$-almost all $\omega \in \Omega$ well defined and converges to the true probability distribution of $(x_\star, y_\star^e)$. We are now ready to establish the following notion.

[5] In the context of random dynamical systems, the term almost surely (a.s.) is used in a non-standard sense: a property holds a.s. if there exists a ϑ-invariant set $\Omega' \subset \Omega$ ($\vartheta\Omega' = \Omega'$) such that the property holds for all $\omega \in \Omega'$.

Definition 4.7. *Given an economic law G, a forecasting rule ψ is called* asymptotically ε-unbiased *if there exists an asymptotically stable random fixed point $(x_\star, y_\star^e)$ of G_ψ such that ψ is locally ε-unbiased along $(x_\star, y_\star^e)$, i.e.,*

$$\big(x_\star(\omega), y_\star^e(\omega)\big) \in \mathcal{W}_G^\varepsilon(\omega) \quad \mathbb{P} - a.s.$$

For $\varepsilon = 0$, ψ is called asymptotically unbiased.

Let ψ be an asymptotically unbiased forecasting rule which is locally unbiased along a random fixed point $(x_\star, y_\star^e)$ of G_ψ. Since $\mathbb{X} = \overline{\mathbb{X}} \times \mathbb{Y}$, we can split $x_\star$ into $x_\star = (\overline{x}_\star, y_\star) : \Omega \to \overline{\mathbb{X}} \times \mathbb{Y}$. The forecast errors along $(x_\star, y_\star^e)$ are

$$\zeta_{t+1}^\star(\omega) = y_\star(\vartheta^{t+1}\omega) - \psi\big(\vartheta^t\omega, x_\star(\vartheta^t\omega), y_\star^e(\vartheta^{t-1}\omega)\big), \quad \omega \in \Omega,\ t \in \mathbb{N}$$

and satisfy $\mathbb{E}_t[\zeta_{t+1}^\star] = 0$ for all times $t \in \mathbb{N}$. That is, all mean forecast errors conditional on the maximum amount of information available prior to actual realizations vanish. This implies in particular that all forecast errors are zero in the mean, $\mathbb{E}_0[\zeta_t^\star] = 0$ for all $t \in \mathbb{N}$. It follows from the stability property and the ergodicity of $(x_\star, y_\star^e)$ that the sample mean errors and all sample autocorrelation coefficients of all sufficiently close orbits $\{x_t, y_t^e\}_{t\in\mathbb{N}}$ as well as those of the corresponding forecast errors $\{\zeta_t\}_{t\in\mathbb{N}}$ converge almost surely to the respective true means and the true autocorrelation coefficients associated with the random fixed point $(x_\star, y_\star^e)$.

Remark 4.4. *The observation just made illustrates that an asymptotically unbiased forecasting rule induces a (stochastic)* consistent expectations equilibrium *(CEE) in the sense of Hommes & Sorger (1998) and Hommes, Sorger & Wagner (2004), provided that agents in the model have a* perceived law of motion *whose hypothesized autocorrelation structure is confirmed by the autocorrelation structure of the realizations. Notice, however, that an unbiased forecasting rule is more precise than a forecasting rule which induces a CEE because the former provides the best least-squares prediction given the history of the process. Moreover, the autocorrelation structure of the forecast errors associated with an unbiased forecasting rule may well be non-trivial, cf. Böhm & Wenzelburger (2002) for a concrete example. Hence, learning schemes designed to find a CEE such as* consistent adaptive learning schemes *in the sense of Schönhofer (1999, 2001) might only be meaningful in situations with additive iid noise.*

We conclude this section by reviewing a simplified version of an existence theorem for random fixed points due to Schmalfuß (1996, 1998). Let $\mathcal{H}(\omega) \subset \mathbb{R}^d, \omega \in \Omega$ be a random closed set, i.e., each $\mathcal{H}(\omega)$ is closed a.s. and $\{\omega \in \Omega \mid \mathcal{H}(\omega) \cap U \neq \emptyset\}$ is measurable for all open sets $U \subset \mathbb{R}^d$ (Arnold 1998, Prop. 1.6.2, p. 32). Consider random variables $h : \Omega \longrightarrow \mathbb{R}^d$ which are tempered, meaning that

$$\lim_{t\to\infty} e^{-\delta t}\|h(\vartheta^t\omega)\| = 0 \quad \text{for all } \delta > 0.$$

Let

$$\mathbb{H} := \left\{h : \Omega \longrightarrow \mathbb{R}^{d+md_y} \;\middle|\; h \text{ is tempered and } \; h(\omega) \in \mathcal{H}(\omega),\; \omega \in \Omega\right\}$$

denote the set of all tempered random variables with values $h(\omega)$ in $\mathcal{H}(\omega)$.

Theorem 4.3. *Let $G_\psi : \mathbb{Z} \times \Omega \times \mathbb{X} \times Y^m \to \mathbb{X} \times \mathbb{Y}^m$ be a random dynamical system over a metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$ as defined in (4.6). Suppose there exists a random set $\mathcal{H}(\omega) \subset \mathbb{X} \times \mathbb{Y}^m$, $\omega \in \Omega$, such that $\mathcal{H}$ is nonempty and the following conditions holds:*

(i) The map $\omega \longmapsto G_\psi\big(1, \vartheta^{-1}\omega, h(\vartheta^{-1}\omega)\big)$ is contained in $\mathcal{H}$ for all $h \in \mathbb{H}$.

(ii) There exists a random variable c with $\mathbb{E}[c] < 0$, such that

$$\sup_{(x,y^e)\neq(\tilde{x},\tilde{y}^e)\in\mathcal{H}(\omega)} \log\left(\frac{\|G_\psi(1,\omega,x,y^e) - G_\psi(1,\omega,\tilde{x},\tilde{y}^e)\|}{\|(x,y^e) - (\tilde{x},\tilde{y}^e)\|}\right) \le c(\omega),\; \omega \in \Omega.$$

(iii) If, for some $h \in \mathbb{H}$, $\{G_\psi\big(t, \vartheta^{-t}\omega, h(\vartheta^{-t}\omega)\big)\}_{t\in\mathbb{N}}$ is a Cauchy sequence for each $\omega \in \Omega$, then its limit is contained in $\mathbb{H}$.

Then there exists a unique random fixed point $h^\star \in \mathbb{H}$ of G_ψ with

$$\lim_{t\to\infty} \|G_\psi\big(t,\omega,h(\omega)\big) - h^\star(\vartheta^t\omega)\| = 0 \quad \textit{for all} \;\; h \in \mathbb{H} \quad \mathbb{P} - a.s.$$

If, in addition, the time-one map $(x, y^e) \mapsto G_\psi(1, \omega, x, y^e)$ is continuously differentiable a.s., then Condition (ii) can be replaced by

(ii') There exists random variable c with $\mathbb{E}[c] < 0$, such that

$$\sup_{(x,y^e)\in\mathcal{H}(\omega)} \log\|\mathrm{D}_{(x,y^e)}G_\psi(1,\omega,x,y^e)\| \le c(\omega),\; \omega \in \Omega$$

with $\mathrm{D}_{(x,y^e)}$ denoting the derivative with respect to (x, y^e).

The concept of an asymptotically stable random fixed point in Definition 4.6 is a natural extension of the notion of an asymptotically stable fixed point in the deterministic case. Combined with the property of unbiasedness, these criteria seem to be natural stability requirements for an ERDS. On a purely distributional level, a simplified version of Theorem 4.3 is given as Theorem 12.7.2 on p. 430 of Lasota & Mackey (1994).

Theorem 4.3 improves the purely statistical viewpoint of considering invariant distributions and Markov equilibria, as done in most economic applications, e.g., see Duffie, Geanakoplos, Mas-Colell & McLennan (1994), Bhattacharya & Majumdar (2001), Spear (1988), or Spear & Srivastava (1986). In fact, the required conditions can be established for a number of economic models, yielding a characterization of the long-run behavior of orbits in models with noise. See Chapters 6 and 7 or Böhm & Wenzelburger (2002) for economic examples. A discussion of the relationship between random fixed points and Markov equilibria is found in Schenk-Hoppé & Schmalfuß (1998).

Concluding Remarks

Extending and confirming results of Chapter 2 to the case of stationary noise, it was shown that the class of unbiased forecasting rules which generate rational expectations are themselves time-invariant functions. The most remarkable phenomenon in stationary economic systems with expectational leads is the notion of an unbiased no-updating forecasting rule which yields the most precise forecasts in the sense that forecast errors vanish conditional on information which was not available at the stage in which they were issued.

The existence of unbiased forecasting rules was reduced to the existence of a global inverse of the mean error function associated with the system. This error function depends exclusively on the fundamentals of the economy and is independent of any expectations formation procedure. The information necessary to construct an unbiased forecasting rule requires detailed knowledge of the whole economic system and amounts to the ability of computing the global inverse of the mean error function. The static nature of this error function, however, opens up the possibility to estimate and approximate unbiased forecasting rules from historical data, whenever they exist.

4.5 Mathematical Appendix

A random dynamical system in the sense of Arnold (1998) consists of two basic ingredients, a model of the exogenous noise and a model of the system which is perturbed by noise. In what follows we restrict ourselves to the case of discrete time. The noise will be modeled by an (ergodic) metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$ in the sense of ergodic theory as defined in Assumption 4.1. Recall that any stationary ergodic process can be represented by an ergodic dynamical system.

Let $\mathbb{X} \subset \mathbb{R}^d$ be given and $\mathcal{B}(\mathbb{X})$ be the σ-algebra of all Borel sets. A *random dynamical system* (RDS) on $\mathbb{X}$ over a metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$ is a mapping

$$\phi : \mathbb{Z} \times \Omega \times \mathbb{X} \to \mathbb{X}, \quad (t, \omega, x) \mapsto \phi(t, \omega, x)$$

with the following properties:

(i) Each $\phi(t, \cdot)$, $t \in \mathbb{Z}$ is $\mathcal{F} \otimes \mathcal{B}(\mathbb{X})/\mathcal{B}(\mathbb{X})$ measurable.

(ii) The mappings $\phi(t, \omega, \cdot) : \mathbb{X} \to \mathbb{X}$ form a *cocycle* over $\{\vartheta^t\}_{t\in\mathbb{Z}}$, that is, they satisfy

$$\phi(0, \omega, x) = x \quad \text{for all} \;\; x \in \mathbb{X}, \; \omega \in \Omega$$

and

$$\phi(t_1 + t_2, \omega, x) = \phi\big(t_1, \vartheta^{t_2}\omega, \phi(t_2, \omega, x)\big)$$

for all $x \in \mathbb{X}$, $\omega \in \Omega$, and $t_1, t_2 \in \mathbb{Z}$.

If, in addition, each map $\phi(1,\omega,\cdot):\mathbb{X}\to\mathbb{X}$, $\omega\in\Omega$, is continuous or even of class C^k, then ϕ is said to be continuous or of class C^k, respectively. If the system started at $x_0\in\mathbb{X}$, then $x_t=\phi(t,\omega,x_0)$ describes the state of the system at time t. For an arbitrary initial condition $x_0\in\mathbb{X}$ and any perturbation $\omega\in\Omega$, the sequence of points $\gamma(x_0):=\{x_t\}_{t\in\mathbb{N}}$ with $x_t=\phi(t,\omega,x_0)$, $t\in\mathbb{N}$, is called an orbit of the random dynamical system ϕ.

Having defined a random dynamical system, one may now define the past and the future associated with such systems. The following definition is taken from Arnold (1998, Def. 1.7.1, p. 37).

Definition 4.8. *A sub-σ-algebra $\mathcal{F}^-$ of $\mathcal{F}$ is called* past *of the RDS ϕ if*

(i) each map $\phi(t,\cdot,x)$, $x\in\mathbb{X}$, is $\mathcal{F}^-/\mathcal{B}(\mathbb{X})$-measurable for all $t\le 0$, and
(ii) $\vartheta^{-t}\mathcal{F}^-\subset\mathcal{F}^-$ for all $t\le 0$.

A sub-σ-algebra $\mathcal{F}^+$ of $\mathcal{F}$ is called future *of the RDS ϕ if*

(i) each map $\phi(t,\cdot,x)$, $x\in\mathbb{X}$, is $\mathcal{F}^+/\mathcal{B}(\mathbb{X})$-measurable for all $t\ge 0$, and
(ii) $\vartheta^{-t}\mathcal{F}^+\subset\mathcal{F}^+$ for all $t\ge 0$.

The smallest possible choice for the past of ϕ is

$$\mathcal{F}^-:=\sigma(\phi(s,\cdot,x)|\ s\le 0,\ x\in\mathbb{X}),$$

and the smallest possible choice for the future of ϕ is

$$\mathcal{F}^+:=\sigma(\phi(s,\cdot,x)|\ s\ge 0,\ x\in\mathbb{X}).$$

see Arnold (1998, Sec. 1.7, p. 37). Setting for each $t\in\mathbb{Z}$, $\mathcal{F}_t^-:=\vartheta^{-t}\mathcal{F}^-$ and $\mathcal{F}_t^+:=\vartheta^{-t}\mathcal{F}^+$, it is clear that $\mathcal{F}_t^-$ is increasing and $\mathcal{F}_t^+$ is decreasing in $t\in\mathbb{Z}$, respectively. Consider now the sub-σ-algebras

$$\mathcal{F}_t:=\sigma(\phi(s,\cdot,x)|\ s\le t,\ x\in\mathbb{X})$$

and

$$\mathcal{F}'_t:=\sigma(\phi(s,\cdot,x)|\ s\ge t,\ x\in\mathbb{X})$$

of $\mathcal{F}$ and for each $r\in\mathbb{N}$, set

$$\mathcal{F}_{t,r}=\mathcal{F}_t\cap\mathcal{F}'_{t-r}.$$

The sub-σ-algebras $\mathcal{F}_t$ and $\mathcal{F}_{t,r}$ have been used in the main text of this chapter, as these seem to be intuitively easier to comprehend.

Lemma 4.1. *Let ϕ be a RDS over a metric dynamical system $(\Omega,\mathcal{F},\mathbb{P},\{\vartheta^t\}_{t\in\mathbb{Z}})$. Then, with the notation introduced above, the following holds for each $t\in\mathbb{Z}$:*

$$(i)\ \mathcal{F}_t=\mathcal{F}_t^-,\quad (ii)\ \mathcal{F}'_t=\mathcal{F}_t^+,\quad (iii)\ \mathcal{F}_{t+s,r}=\vartheta^{-s}\mathcal{F}_{t,r},\ r\in\mathbb{N},\ s\in\mathbb{Z}.$$

Proof. (i) *Step 1.* We show that $\mathcal{F}_t^- \subset \mathcal{F}_t$. Let $s \le 0$, $y \in \mathbb{X}$, and $B \in \mathcal{B}(\mathbb{X})$ be arbitrary but fixed and consider a typical set

$$F_y^- := \{\omega \in \Omega |\ \phi(s, \vartheta^t \omega, y) \in B\} \in \mathcal{F}_t^- .$$

To show that $F_y^- \in \mathcal{F}_t$, let

$$F_x := \{\omega \in \Omega |\ \phi(t, \omega, x) = y \ \text{ and } \ \phi(t + s, \omega, x) \in B\}, \quad x \in \mathbb{X}.$$

Since $s \le 0$, $F_x \in \mathcal{F}_t$. The cocycle property of ϕ reads

$$\phi(t + s, \omega, x) = \phi\big(s, \vartheta^t \omega, \phi(t, \omega, x)\big), \quad x \in \mathbb{X}.$$

Hence $\bigcup_{x\in\mathbb{X}} F_x \subset F_y^-$. On the other hand, let $\omega \in F_y^-$ be arbitrary. Since $\phi(t, \omega, \cdot)$ is invertible, there exists a unique $x_\omega \in \mathbb{X}$ such that $\phi(t, \omega, x_\omega) = y$. Hence $\phi(t + s, \omega, x_\omega) = \phi(s, \vartheta^t \omega, y) \in B$, implying that $\omega \in F_{x_\omega}$. This shows that $F_y^- \subset \bigcup_{x\in\mathbb{X}} F_x$ and hence $F_y^- = \bigcup_{x\in\mathbb{X}} F_x$. Since $s \le 0$, $y \in \mathbb{X}$, and $B \in \mathcal{B}(\mathbb{X})$ were arbitrary, this proves $\mathcal{F}_t^- \subset \mathcal{F}_t$.

Step 2. The proof is complete if one can show that $\mathcal{F}_t \subset \mathcal{F}_t^-$. To do so, let $s \le t$, $x \in \mathbb{X}$, and $B \in \mathcal{B}(\mathbb{X})$ be arbitrary but fixed. Consider a typical set

$$F_x := \{\omega \in \Omega |\ \phi(s, \omega, x) \in B\} \in \mathcal{F}_t.$$

The cocycle property of ϕ implies

$$\phi(s - t, \vartheta^t \omega, y) = \phi\big(s, \omega, \phi(-t, \vartheta^t \omega, y)\big), \quad y \in \mathbb{X}.$$

Consider now sets of the form ($s \le t$ fixed as above)

$$F_y^- := \{\omega \in \Omega |\ \phi(-t, \vartheta^t \omega, y) = x \ \text{ and } \ \phi(s - t, \vartheta^t \omega, y) \in B\} \in \mathcal{F}_t^-, \quad y \in \mathbb{X}.$$

By a reasoning analogous to Step 1, $F_x = \bigcup_{y\in\mathbb{X}} F_y^-$ which yields $\mathcal{F}_t \subset \mathcal{F}_t^-$.

(ii) The proof of Property (ii) is analogous to that of Property (i).

(iii) Using (i) and (ii), one has

$$\vartheta^{-s}\mathcal{F}_{t,r} = \vartheta^{-s}\mathcal{F}_t^- \cap \vartheta^{-s}\mathcal{F}_t^+ = \mathcal{F}_{t+s}^- \cap \mathcal{F}_{t+s}^+ = \mathcal{F}_{t+s,r}.$$

□

We use Lemma 4.1 to prove the following Proposition.

Proposition 4.1. *Let ϕ be a RDS over a metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$ as defined above and $h : \mathbb{X} \to \mathbb{R}^d$ be an integrable map. Let $\mathcal{F}_t = \vartheta^{-t}\mathcal{F}^0$, $t \in \mathbb{Z}$ for some sub-σ-algebra $\mathcal{F}^0 \subset \mathcal{F}$. Then, with the notation introduced above, for each $x_0 \in \mathbb{X}$,*

$$\begin{aligned}\mathbb{E}\big[h\big(\phi(t + 1, \cdot, x_0)\big)\,\big|\mathcal{F}_t\big]\ (\omega) &= \mathbb{E}\big[h\big(\phi(1, \vartheta^t \cdot, x_t)\big)\,\big|\mathcal{F}_t\big]\ (\omega)\\ &= \mathbb{E}_0\big[h\big(\phi(1, \cdot, x_t)\big)\,\big|\mathcal{F}^0\big]\ (\vartheta^t\omega)\end{aligned}$$

$\mathbb{P}$-a.s., where $x_t = \phi(t, \omega, x_0)$.

Proof. The assertion follows from Lemma 4.2 and Lemma 4.3 below. □

Lemma 4.2. *Let* $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$ *be a metric dynamical system,* $\mathcal{F}_t = \vartheta^{-t}\mathcal{F}^0$, $t \in \mathbb{Z}$ *for some sub-σ-algebra* $\mathcal{F}^0 \subset \mathcal{F}$*, and* $u : \Omega \to \mathbb{X}$ *be a* $\mathcal{F}/\mathcal{B}(\mathbb{X})$*-measurable random variable. Then*

$$\mathbb{E}[u \circ \vartheta^t | \mathcal{F}_t] = \mathbb{E}[u|\mathcal{F}^0] \circ \vartheta^t \quad \mathbb{P} - a.s.$$

Proof. By the definition of a metric dynamical system, ϑ is measure preserving with respect to $\mathbb{P}$. This implies $\vartheta^t\mathbb{P} = \mathbb{P}$ with $\vartheta^t\mathbb{P}$ denoting the image measure under ϑ^t. The change-of-variable formula then gives

$$\int_F u \, \mathrm{d}\mathbb{P} = \int_F u \, \mathrm{d}(\vartheta^t\mathbb{P}) = \int_{\vartheta^{-t}(F)} u \circ \vartheta^t \, \mathrm{d}\mathbb{P}, \quad F \in \mathcal{F}^0. \tag{4.36}$$

On the other hand, the definition of the conditional expectations and the change-of-variable formula imply

$$\int_F u \mathrm{d}\mathbb{P} = \int_F \mathbb{E}[u|\mathcal{F}^0] \, \mathrm{d}\mathbb{P} = \int_{\vartheta^{-t}(F)} \mathbb{E}[u|\mathcal{F}^0] \circ \vartheta^t \, \mathrm{d}\mathbb{P}, \quad F \in \mathcal{F}^0 \quad \mathbb{P}-\text{a.s.} \tag{4.37}$$

Noting that $\mathcal{F}_t = \vartheta^{-t}\mathcal{F}^0$ and applying the definition of the conditional expectations again, (4.36) and (4.37) yield

$$\int_{\vartheta^{-t}(F)} \mathbb{E}[u \circ \vartheta^t | \mathcal{F}_t] \, \mathrm{d}\mathbb{P} = \int_{\vartheta^{-t}(F)} \mathbb{E}[u|\mathcal{F}^0] \circ \vartheta^t \, \mathrm{d}\mathbb{P}, \quad F \in \mathcal{F}^0 \quad \mathbb{P} - \text{a.s.}$$

This completes the proof. □

Lemma 4.3. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *a probability space,* $\Phi : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}^{d_3}$ *be an integrable map, and* $x : \Omega \to \mathbb{R}^{d_1}$ *and* $y : \Omega \to \mathbb{R}^{d_2}$ *be integrable random variables. Then*

$$\mathbb{E}[\Phi(x, y)|y](\omega) = \mathbb{E}\big[\Phi\left(x, y(\omega)\right)|y\big](\omega), \quad \mathbb{P} - a.s.$$

Here, both conditional expectations are taken with respect to the sub-σ-algebra $\sigma(y)$ *and the second one is the conditional expectation of the random variable* $\omega \mapsto \Phi\big(x(\omega), \eta\big)$ *with arbitrary but fixed* $\eta \in \mathbb{R}^{d_2}$*.*

Proof. The factorization lemma (Bauer 1992, p. 71) ensures the existence of a $\sigma(y)$-measurable map h such that

$$\mathbb{E}[\Phi(x, y)|y](\omega) = h\big(y(\omega)\big), \quad \mathbb{P} - a.s.$$

and Proposition 15.9 in Bauer (1991, p. 128) shows that

$$h(\eta) = \frac{1}{\mathbb{P}(y = \eta)} \int_{y^{-1}(\eta)} \Phi(x, y) \mathrm{d}\mathbb{P} = \frac{1}{\mathbb{P}(y = \eta)} \int_{y^{-1}(\eta)} \Phi(x, \eta) \mathrm{d}\mathbb{P} \quad \mathbb{P} - \text{a.s.},$$

provided that $\mathbb{P}(y = \eta) > 0$. Similarly, for each fixed $\zeta \in \mathbb{R}^{d_2}$, there exists a $\sigma(y)$-measurable map $H(\cdot, \zeta)$ such that

$$\mathbb{E}[\varPhi(x, \zeta)|y](\omega) = H\big(y(\omega), \zeta\big), \quad \mathbb{P} - a.s.$$

and

$$H(\eta, \zeta) = \frac{1}{\mathbb{P}(y = \eta)} \int_{y^{-1}(\eta)} \varPhi(x, \zeta) \mathrm{d}\mathbb{P}, \quad \mathbb{P} - a.s.$$

Hence, $h\big(y(\omega)\big) = H\big(y(\omega), y(\omega)\big)$ which completes the proof. □

5

Nonparametric Adaptive Learning

In this chapter we introduce a learning scheme for unbiased forecasting rules of the non-linear system (2.8) as introduced in Chapter 2, using nonparametric methods developed in Chen & White (1996, 1998, 2002). Based on results of Yin & Zhu (1990), Yin (1992), and many other authors, these contributions generalize an approach in stochastic approximation theory initiated by Kushner & Clark (1978) to infinite-dimensional spaces.We will specialize these results to economic laws with expectational leads.

The main difference between the learning scheme proposed in this chapter and the approach of the traditional learning literature is that we distinguish between the estimation of an unbiased forecasting rule associated with an economic law and the actual application of a forecasting rule within the system. As in the linear case of Chapter 3, forecasts will be treated systematically as exogenous inputs in the control theoretical sense. One advantage of such an approach is that, contrary to Chen & White (1998), the unknown functional relationship which has be to estimated is uniquely determined. This considerably reduces the mathematical equipment needed to establish convergence results. It allows to select between multiple unbiased forecasting rules on economic grounds and, in principle, it enables a forecasting agency to stabilize the evolving system lat all stages of the learning procedure.

The general idea of the learning scheme is to construct an approximation of an unbiased forecasting rule from historical data. A forecasting rule will be updated at each point in time, where we will assume that this rule uses a finite amount of historical data only. Although a forecasting rule of the general form (2.13) could be taken, our existence and uniqueness results (Theorems 4.1 and 4.2) of Chapter 4 suggest to take no-updating rules. Moreover, as Theorem 2.1 of Chapter 2 showed, unbiased no-updating rules provide the most precise forecasts.

Assuming existence and uniqueness, we will show how to approximate an unbiased no-updating forecasting rule from an estimate of the 'inverse' of the economic law. If it is unknown whether or not such a forecasting rule exists, then the economic law G should be directly estimated. As all multiplicities are embodied in the economic law, an approximated unbiased forecasting rule could then be computed from an estimated economic law. The estimation techniques are completely analogous. We start with an outline of the general methodology.

5.1 The Method of Stochastic Approximation

Recall that, given the information vector of past observed states

$$Z_t = (y^e_{t-1}, x_{t-1}, \ldots, y^e_{t-r}, x_{t-r}) \in \Sigma$$

in an arbitrary period t, the vector of forecasts $y^e_t = (y^e_{t,t+1}, \ldots, y^e_{t,t+m})$ formed in period t by means of a no-updating forecasting rule $\Psi = (\Psi^{(1)}, \ldots, \Psi^{(m)})$ is

$$\begin{cases} y^e_{t,t+i} &= \Psi^{(i)}(x_t, Z_t) \;:=\; y^e_{t-1,t+i}, \quad i = 1, \ldots, m-1, \\ y^e_{t,t+m} &= \Psi^{(m)}(x_t, Z_t). \end{cases} \tag{5.1}$$

If $(x_{t-1}, Z_{t-1}) \in \mathbb{X} \times \Sigma \subset \mathbb{R}^{d_{\mathbb{H}}}$ denotes the state of the system in period $t-1$, then in view of Assumption 2.1

$$\begin{cases} x_t &= G\big(\xi_t(\omega), x_{t-1}, y^e_{t-1}\big) \\ Z_t &= \big(y^e_{t-1}, x_{t-1}, \mathrm{pr}^{\Sigma}_{-r} Z_{t-1}\big) \\ y^e_t &= \Psi(x_t, Z_t) \end{cases} \tag{5.2}$$

describes the state of the system in period t under the perturbation $\omega \in \Omega$, where as before

$$\mathrm{pr}^{\Sigma}_{-r} : \Sigma \to \prod_{i=1}^{r-1} (\mathbb{Y}^m \times \mathbb{X}), \quad (Z^{(1)}, \ldots, Z^{(r)}) \mapsto (Z^{(1)}, \ldots, Z^{(r-1)}) \tag{5.3}$$

denotes the projection into the first $r-1$ components.

Remark 5.1. *The information contained in Z_t is redundant when using no-updating forecasting rules of the form (5.1). In this case*

$$y^e_{t,t-j+m} = y^e_{t-j,t-j+m}, \quad j = 1, \ldots, m-1,$$

so that the first $m-1$ components of $y^e_t = (y^e_{t,t+1}, \ldots, y^e_{t,t+m})$ are redundant if $r > m$.

In order to formulate a learning scheme according to which the forecasting rule Ψ is updated, we need to specify the set of *feasible* forecasting rules from which an agent or a forecasting agency can choose. Let $\mathbb{X} \times \Sigma$ be an open and convex subset of $\mathbb{R}^{d_{\mathbb{H}}}$ and $\|\cdot\|$ denote the Euclidean norm. Restricting attention to no-updating forecasting rules, it suffices to consider the function $\Psi^{(m)}$ for the m-th forecast of a forecasting rule Ψ. Let $C_B^p(\mathbb{X}\times\Sigma;\mathbb{R}^{d_y})$ denote the space of all bounded p-times continuously differentiable functions $\psi:\mathbb{X}\times\Sigma\to\mathbb{R}^{d_y}$ with bounded derivatives and $C_B(\mathbb{X}\times\Sigma;\mathbb{R}^{d_y})$ be the space of all functions which are continuous and bounded.[1] It is well known when $C_B^p(\mathbb{X}\times\Sigma;\mathbb{R}^{d_y})$ is a dense subspace of a real separable Hilbert space $\mathbb{H}$. This could be either the space of all square integrable functions $L_2(\mathbb{X}\times\Sigma;\mathbb{R}^{d_y})$ or the Sobolev space $\mathbb{W}^{p,2}(\mathbb{X}\times\Sigma;\mathbb{R}^{d_y})$.[2]

We will choose the Sobolev spaces $\mathbb{H}:=\mathbb{W}^{p,2}(\mathbb{X}\times\Sigma;\mathbb{R}^{d_y})$ and $\mathbb{H}':=\mathbb{W}^{p,2}(\mathbb{X}\times\Sigma;\mathbb{R}^{md_y})$ with $p>\frac{d_{\mathbb{H}}}{2}$ as the underlying Hilbert spaces of our learning scheme. By the Sobolev Imbedding Theorem (see Theorem 5.4 below) which will be applied in the sequel, $\mathbb{H}$ can be imbedded into $C_B(\mathbb{X}\times\Sigma;\mathbb{R}^{d_y})$ in the sense that if $\psi\in\mathbb{H}$, then there exists $\psi_0\in C_B(\mathbb{X}\times\Sigma;\mathbb{R}^{d_y})$ such that $\psi=\psi_0$ almost everywhere. The Sobolev Imbedding Theorem requires that $\mathbb{X}\times\Sigma$ has a geometric property, called the cone property. Open balls and open rectangular domains are examples for sets having the cone property, so that from an economic view point this assumption imposes no serious restriction. A brief discussion of all necessary details is found in Section 5.5.1 below.

The Hilbert space $\mathbb{H}$ is endowed with an inner product and inner-product induced norm, denoted by $\langle\cdot,\cdot\rangle_{\mathbb{H}}$ and $\|\cdot\|_{\mathbb{H}}$, respectively. The σ-algebra generated by the subsets of $\mathbb{H}$ which are open in the norm-induced topology of $\mathbb{H}$ is denoted by $\mathcal{B}(\mathbb{H})$. Similarly, we denote by $\mathbb{H}'$ the Hilbert space of all functions $\Psi:\mathbb{X}\times\Sigma\to\mathbb{R}^{md_y}$ endowed with the norm $\|\cdot\|_{\mathbb{H}'}$. Except for the dimension of the target space $\mathbb{R}^{md_y}$, we assume throughout that $\mathbb{H}$ and $\mathbb{H}'$ are Hilbert spaces of the same sort. Fix $p>\frac{d_{\mathbb{H}}}{2}$ as above and set

$$\mathbb{S}:=C_B^p(\mathbb{X}\times\Sigma;\mathbb{Y}^m) \tag{5.4}$$

for the space of *feasible forecasting rules*. Note that any feasible forecasting rule Ψ must take values in $\mathbb{Y}^m$ and that $\mathbb{S}$ is a subspace of $\mathbb{H}'$. Let $\pi_{\mathbb{Y}}$ denote a projection of $\mathbb{R}^{d_y}$ onto $\mathbb{Y}$. Define an imbedding $\mathbf{I}$ of $\mathbb{H}$ into $\mathbb{H}'$ by

$$\mathbf{I}:\mathbb{H}\to\mathbb{H}',\quad \psi\mapsto\mathbf{I}(\psi), \tag{5.5}$$

such that $\mathbf{I}(\psi):\mathbb{X}\times\Sigma\to\mathbb{Y}^m$ is given by

$$\mathbf{I}(\psi)\big(x,Z\big):=\big(z^{(2)},\dots,z^{(m)},\pi_{\mathbb{Y}^m}\circ\psi(x,Z)\big),\quad (x,Z)\in\mathbb{X}\times\Sigma,$$

where $Z=(z^{(1)},\dots,z^{(m)},Z')\in\Sigma$ and $(z^{(1)},\dots,z^{(m)})\in\mathbb{Y}^m$. In particular, $\mathbf{I}$ transforms any $\psi\in C_B^p(\mathbb{X}\times\Sigma;\mathbb{R}^{d_y})$ into a feasible forecasting rule $\mathbf{I}(\psi)\in C_B^p(\mathbb{X}\times\Sigma;\mathbb{Y}^m)$.

[1] Throughout this chapter we change the notation with respect to Chapter 4 and denote the forecasting function for the m-th forecast by ψ for ease of notation.

[2] As an alternative, the space of all p-times continuously differentiable functions with compact support could be chosen. By Theorem 3.18, p. 54 of Adams (1975), this space is a dense subset of a Sobolev space as well if $\mathbb{X}\times\Sigma$ has an additional geometric property, called the segment property.

With these choices, the learning scheme which will be introduced below may require that the forecasting rule obtained from an estimation procedure can be censored. The main reason for such a manipulation is to generate certain stochastic properties of the estimates that guarantee convergence of the algorithm. To this end, we introduce for each $t \in \mathbb{N}$ a *censor map*

$$\Upsilon_t : \mathbb{H} \to \mathbb{H}', \quad \psi \mapsto \Upsilon_t(\psi) \tag{5.6}$$

which maps $C_B^p(\mathbb{X} \times \Sigma; \mathbb{R}^{d_y})$ onto $\mathbb{S}$. The simplest example is $\Upsilon_t \equiv \mathbf{I}$ for all $t \in \mathbb{N}$ when no censoring is applied.

We next define the way in which a forecasting rule will be updated. Throughout this chapter, denote by $\psi_\star = \Psi_\star^{(m)}$ the m-th component function of an unbiased no-updating $\Psi_\star$ introduced in Definition 2.3 of Section 2.5 in Chapter 2. Clearly, $\Psi_\star = \mathbf{I}(\psi_\star)$. Let $\{a_t\}_{t\in\mathbb{N}}$ be a sequence of positive real numbers, which will be specified later in this section, and for each $t \in \mathbb{N}$, let

$$M_t : \mathbb{X} \times \Sigma \times \mathbb{H} \longrightarrow \mathbb{H}$$

be a function that governs the updating of a forecasting rule at time $t \in \mathbb{N}$. Assume that each $M_t(x, Z, \cdot)$, $t \in \mathbb{N}$, $(x, Z) \in \mathbb{X} \times \Sigma$ maps differentiable functions onto differentiable functions. The resulting learning scheme then evolves over time according to the list of mappings

$$\begin{cases} x_t &= G\big(\xi_t(\omega), x_{t-1}, y_{t-1}^e\big) \\ Z_t &= \big(y_{t-1}^e, x_{t-1}, \mathrm{pr}_{-r}^{\Sigma} Z_{t-1}\big) \\ y_t^e &= \Psi_t(x_t, Z_t) \\ \Psi_t &= \Upsilon_t(\hat{\psi}_t) \\ \hat{\psi}_t &= \hat{\psi}_{t-1} + a_{t-1} M_{t-1}(x_t, Z_t, \hat{\psi}_{t-1}) \end{cases} \tag{5.7}$$

with $t \in \mathbb{N}$. The first three equations in the recursive system (5.7) are already given in (5.2) and play the role of the former time-one map (2.15) [or, equivalently, (4.5)]. $\hat{\psi}_t$ is a period-t approximation of the m-th component $\psi_\star$ of the desired unbiased no-updating forecasting rule $\Psi_\star = \mathbf{I}(\psi_\star)$. The censored map $\Psi_t = \Upsilon_t(\hat{\psi}_t)$ is the forecasting rule which is applied in period t. As mentioned above this may become necessary in order to fulfill certain stochastic requirements for the estimation technique and in order to stabilize the evolving system (5.7). The fifth equation in (5.7) specifies the way in which the estimate $\hat{\psi}_t$ is updated. This updating scheme is commonly referred to as a *Robbins-Monro procedure with feedback*.

The basic idea of the algorithm (5.7) is to find the unknown function $\psi_\star$ as follows. Suppose that $\psi_\star$ can be characterized as the zero of an unknown map $M_\star : \mathbb{H} \to \mathbb{H}$. Assume that the unknown map $M_\star$ is approximated by a sequence of functions M_t, $t \in \mathbb{N}$. Then $\psi_\star$ could recursively be estimated by taking some initial element $\hat{\psi}_0 \in \mathbb{H}$ and setting

$$\begin{aligned}\hat{\psi}_{t+1} &= \hat{\psi}_t + a_t M_t(x_{t+1}, Z_{t+1}, \hat{\psi}_t) \\ &= \hat{\psi}_t + a_t \big[U_t(x_{t+1}, Z_{t+1}, \hat{\psi}_t) + M_\star(\hat{\psi}_t) \big], \quad t = 0, 1, \ldots,\end{aligned}$$

where $U_t(x_{t+1}, Z_{t+1}, \hat{\psi}_t) := M_t(x_{t+1}, Z_{t+1}, \hat{\psi}_t) - M_\star(\hat{\psi}_t)$ is the approximation error of period t. We will show that $\|\hat{\psi}_t - \psi_\star\|_{\mathbb{H}}$ converges almost surely to zero as t tends to infinity, provided that the approximation errors $\big\{U_t(x_{t+1}, Z_{t+1}, \hat{\psi}_t)\big\}_{t\in\mathbb{N}}$ fulfill certain stochastic requirements.

In order to establish such convergence results, the magnitudes $\|\hat{\psi}_t\|_{\mathbb{H}}$ of the estimates $\hat{\psi}_t$ must be hindered from diverging. This will be achieved by modifying the *learning scheme* (5.7) as follows. The idea is to *'reset'* the estimate $\hat{\psi}_t$ to some fixed forecasting rule $\overline{\psi} \in \mathbb{H}$ whenever the sequence $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ attempts to escape. To this end, let χ_J denote the indicator function of a subset $J \subset \Omega$ and $\{b_t\}_{t\in\mathbb{N}}$ be a sequence of strictly increasing positive real numbers. Define a sequence of positive integer-valued random variables $\{\gamma(t)\}_{t\in\mathbb{N}}$ by

$$\gamma(0) = 0, \quad \gamma(t+1) = \gamma(t) + \chi_{J_t^c}, \tag{5.8}$$

where

$$J_t := \Big\{ \|\hat{\psi}_{t-1} + a_{t-1} M_{t-1}(x_t, Z_t, \hat{\psi}_{t-1})\|_{\mathbb{H}} \le b_{\gamma(t)} \Big\},$$

$$J_t^c := \Big\{ \|\hat{\psi}_{t-1} + a_{t-1} M_{t-1}(x_t, Z_t, \hat{\psi}_{t-1})\|_{\mathbb{H}} > b_{\gamma(t)} \Big\}.$$

Fix some $\bar{b}$ with $0 < \bar{b} < b_1$ and choose a fixed $\overline{\psi} \in \mathbb{H}$ with $\|\overline{\psi}\|_{\mathbb{H}} < \bar{b}$. For each $t \in \mathbb{N}$, set

$$\begin{cases} x_t &= G\big(\xi_t(\omega), x_{t-1}, y^e_{t-1}\big) \\ Z_t &= \big(y^e_{t-1}, x_{t-1}, \mathrm{pr}^{\Sigma}_{-r} Z_{t-1}\big) \\ y^e_t &= \Psi_t(x_t, Z_t) \\ \Psi_t &= \Upsilon_t(\hat{\psi}_t) \\ \hat{\psi}_t &= \Big[\hat{\psi}_{t-1} + a_{t-1} M_{t-1}(x_t, Z_t, \hat{\psi}_{t-1})\Big] \chi_{J_t} + \overline{\psi}\, \chi_{J_t^c} \end{cases} \tag{5.9}$$

Both systems (5.7) and (5.9) are variants of learning schemes developed in Chen & White (1998, 2002). The main difference is that we distinguish between the recursive approximation of a forecasting rule and its application to the system such that the forecasts which actually feed back into the process may be altered or, in other words, censored. The learning scheme (5.9) is also referred to as *randomly truncated Robbins-Monro procedure with feedback* cf. Chen & White (1998).

Remark 5.2. *In the case in which $\{\xi_t\}_{t\in\mathbb{N}}$ is iid, Remark 2.4 of Chapter 2 showed that $\psi_\star$ is a function of the form $\psi_\star : \mathbb{X} \times \mathbb{Y}^{m-1} \to \mathbb{Y}$. In this case, $\Sigma = \mathbb{Y}^m$, $r = 0$ and one may set*

$$\mathbb{S} := C_B^p(\mathbb{X} \times \mathbb{Y}^{m-1}; \mathbb{Y}^m)$$

for the space of feasible forecasting rules. The Hilbert spaces $\mathbb{H}$ *and* $\mathbb{H}'$ *are defined accordingly. The same remark applies to the case of stationary ergodic noise with* $r = 0$ *treated in Chapter 4.*

Remark 5.3. *The stochastic approximation schemes in (5.7) and (5.9) involve a generalization of a stochastic approximation method developed by Kushner & Clark (1978) to infinite-dimensional systems with feedback, see also Kushner & Yin (2003) and references therein. The original algorithm without an expectations feedback dates back to Robbins & Monro (1951). Kushner & Clark (1978) extended this method to finite-dimensional systems with feedback, Yin & Zhu (1990), Yin (1992), and others developed a generalization to systems with an infinite-dimensional state space. Chen & White (1996, 1998, 2002) extended and refined these results to feedback systems with an infinite-dimensional state space. For an account of the literature we refer to Chen and White's as well as to Yin and Zhu's work.*

Remark 5.4. *The system (5.9) is not necessarily a random dynamical system in the sense of Arnold (1998) because the fourth and fifth equation in (5.9) will most generally violate the cocycle property (see Chapter 4).*

5.2 A Basic Convergence Result

We are now in a position to formulate a convergence result which is based on Yin & Zhu (1990) and Chen & White (1996, 2002). The convergence result stated in Theorem 5.1 below is quite general, whereas the required assumptions are rather abstract. Indeed, the crucial assumption which guarantees convergence will, in general, be difficult if not impossible to verify. A list of assumptions which is easier to verify will be formulated in Section 5.4.

There are various approaches using different assumptions to prove almost sure convergence of the sequence of estimates $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ to $\psi_\star$. In essence, all try to establish the following two relations:

1. Boundedness of the sequence of estimates, that is,

$$\sup_{t\in\mathbb{N}} \|\hat{\psi}_t\|_{\mathbb{H}} < \infty \quad \mathbb{P}-\text{a.s.}$$

2. For each $T > 0$, convergence of the sum

$$\lim_{t\to\infty} \sup_{t<n\leq \mathrm{n}(t,T)} \Big\| \sum_{j=t}^{n-1} a_j \left[M_j(x_{j+1}, Z_{j+1}, \hat{\psi}_j) - M_\star(\hat{\psi}_j)\right] \Big\|_{\mathbb{H}} = 0 \quad \mathbb{P}-\text{a.s.},$$

where for each $t \in \mathbb{N}$,

$$\mathrm{n}(t,T) := \max\Big\{ i > t \Big| \sum_{j=t}^{i-1} a_j \leq T \Big\}.$$

Before stating our assumptions, let us introduce some notation. Given some feasible forecasting rule $\Psi \in \mathbb{S}$ and an economic law G of the form (2.8), it will turn out to be convenient to allow the time-one map as defined in (2.15) to depend on Ψ. Formally, we define

$$\mathbf{G} : \begin{cases} \Xi \times \mathbb{X} \times \Sigma \times \mathbb{S} & \longrightarrow \quad \mathbb{X} \times \Sigma \\ (\xi, x, Z, \Psi) & \longmapsto \quad \begin{pmatrix} G\big(\xi, x, \Psi(x, Z)\big) \\ \big(\Psi(x, Z), x, \mathrm{pr}^{\Sigma}_{-r} Z\big) \end{pmatrix} \end{cases}. \tag{5.10}$$

The map (5.10) will replace the role of the former time-one map (2.15) as follows. Let $\prod_{i=1}^{t+1} \mathbb{S}$ denote the $(t+1)$-fold product of $\mathbb{S}$ and

$$\Psi_0^t = (\Psi_0, \ldots, \Psi_t) \in \prod_{i=1}^{t+1} \mathbb{S}$$

be a list of forecasting rules, where Ψ_s is the forecasting rule applied in period s. Define maps

$$\mathbf{G}(t, \cdot) : \Omega \times \mathbb{X} \times \Sigma \times \prod_{i=1}^{t} \mathbb{S} \to \mathbb{X} \times \Sigma, \qquad t \in \mathbb{N} \tag{5.11}$$

recursively by setting

$$\begin{aligned} &\mathbf{G}\left(t, \omega, x_0, Z_0, \Psi_0^{t-1}\right) \\ &\qquad := \begin{cases} \mathbf{G}\left(\xi_t(\omega), \mathbf{G}\left(t-1, \omega, x_0, Z_0, \Psi_0^{t-2}\right), \Psi_{t-1}\right) & \text{if } t > 0, \\ (x_0, Z_0) & \text{if } t = 0. \end{cases} \end{aligned}$$

If the system started in some state $(x_0, Z_0) \in \mathbb{X} \times \Sigma$, then the t-th iteration of the map (5.10) under the perturbation $\omega \in \Omega$ induces maps (5.11) such that

$$(x_t, Z_t) = \mathbf{G}\left(t, \omega, x_0, Z_0, \Psi_0^{t-1}\right)$$

describes the state of the system at time t. If $\Psi_0^{t-1} = (\Psi, \ldots, \Psi)$ is a constant list of functions so that $\Psi_s = \Psi$ for $s = 0, \ldots, t-1$, we write

$$\mathbf{G}\left(t, \omega, x_0, Z_0, \Psi\right) = \mathbf{G}\left(t, \omega, x_0, Z_0, \Psi_0^{t-1}\right) \tag{5.12}$$

for notational convenience.

The following assumptions are taken from Yin & Zhu (1990) and Chen & White (2002) and adjusted to our setup. The first assumption is a slight extension of our previous Assumption 2.1. The second one concerns the dynamic stability of our system under learning. Recall that in the context of infinite dimensional spaces, a mapping is said to be bounded if it maps bounded sets into bounded sets, where we mean bounded with respect to the norm topology.

Assumption 5.1. *The exogenous noise is driven by a sequence of random variables* $\{\xi_t\}_{t\in\mathbb{N}}$ *on a complete probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *with values in a subset* $\Xi \subset \mathbb{R}^{d_\xi}$*, such that for each* $t \in \mathbb{N}$, $\xi_t : \Omega \to \Xi$.

Assumption 5.2. *(i)* $\mathbb{X}\times\Sigma \subset \mathbb{R}^{d_\mathbb{H}}$ *is open, convex and has the cone property.*
(ii) Each map $\Upsilon_t : \mathbb{H} \to \mathbb{S} \subset \mathbb{H}'$, $t \in \mathbb{N}$ *is uniformly continuous on bounded sets. That is, for each bounded subset* $\mathbb{A} \subset \mathbb{H}$ *and each* $\epsilon > 0$*, there exists some* $\delta = \delta(\epsilon, \mathbb{A})$ *such that*

$$\|\psi - \tilde{\psi}\|_\mathbb{H} < \delta \text{ implies } \|\Upsilon_t(\psi) - \Upsilon_t(\tilde{\psi})\|_{\mathbb{H}'} < \epsilon$$

for all ψ, $\tilde{\psi} \in \mathbb{A}$.
(iii) The economic law G *is a continuous map and there exists a compact subset* $\mathbb{D} \subset \mathbb{X} \times \Sigma$ *which is forward-invariant in the sense that*

$$\mathbf{G}\big(\cdot, \Upsilon_t(\cdot)\big) : \Xi \times \mathbb{D} \times \mathbb{H} \to \mathbb{D},$$

uniformly in $t \in \mathbb{N}$.

The boundedness condition in Assumption 5.2 imposes an important stability condition on the recursive system (5.9). This sets aside the problem that forecasting rules could well destabilize the system in the sense that the endogenous state variables $\{(x_t, Z_t)\}_{t\in\mathbb{N}}$ generated by the Robbins-Monro procedure (5.9) diverge. This possibility constitutes a serious control problem. It is ruled out by imposing the strong assumption that a list of *censoring maps* $\{\Upsilon_t\}_{t\in\mathbb{N}}$ is known which keeps the system stable no matter what the actual estimate for a forecasting rule is. It is beyond the scope of these notes to analyze the case in which the unknown system has to be stabilized without this a-priori knowledge. However, in principle, stability could be achieved through a reset mechanism similar to the one in (5.9). Notice in this context that the forecasting rule $\overline{\psi}$ would have to be known so that the system $\mathbf{G}$ is stable when $\overline{\psi}$ is applied in the sense that a bounded forward-invariant set exists.

Example 5.1. *As an example for a time-invariant censor map, consider forecasting rules* $\psi \in C^1_B(\mathbb{X} \times \Sigma; \mathbb{R}^{d_y})$ *with bounded first derivative. Let* $\mathrm{D}\psi(x, Z)$ *denote the derivative (i.e., the Jacobi matrix) of* Ψ *at* (x, Z) *and*

$$\big\|\mathrm{D}\psi\big\|_\infty := \sup\Big\{\big\|\mathrm{D}\psi(x, Z)\big\|\Big|\ (x, Z) \in \mathbb{X} \times \Sigma\Big\}$$

be the supremum norm of the derivative. Choose a positive constant $0 < c_\Upsilon < 1$ *and define*

$$\Upsilon' : \begin{cases} C^1_B(\mathbb{X} \times \Sigma; \mathbb{R}^{d_y}) & \longrightarrow C^1_B(\mathbb{X} \times \Sigma; \mathbb{R}^{d_y}) \\ \qquad \psi & \longmapsto \Upsilon'(\psi) \end{cases} \tag{5.13}$$

by setting

$$\Upsilon'(\psi) = \begin{cases} \psi & \text{if } \|\mathrm{D}\psi\|_\infty < c_\Upsilon \\ \frac{c_\Upsilon}{\|\mathrm{D}\psi\|_\infty}\,\psi & \text{if } \|\mathrm{D}\psi\|_\infty \geq c_\Upsilon \end{cases}.$$

The function (5.13) maps differentiable functions onto differentiable contractions with a prescribed Lipschitz constant c_Υ*, so that*

$$\|\Upsilon'(\psi)(x,Z) - \Upsilon'(\psi)(\tilde{x},\tilde{Z})\| \leq c_\Upsilon \|(x,Z) - (\tilde{x},\tilde{Z})\|$$

for all (x,Z), $(\tilde{x},\tilde{Z}) \in \mathbb{X} \times \Sigma$. *For each* $t \in \mathbb{N}$*, define*

$$\Upsilon_t \equiv \Upsilon : C_B^p(\mathbb{X} \times \Sigma; \mathbb{R}^{d_y}) \to \mathbb{S}, \quad \psi \mapsto \mathbf{I} \circ \Upsilon'(\psi), \tag{5.14}$$

where $\mathbf{I}$ *is defined in (5.5) and* $\mathbb{S}$ *is the space of all feasible forecasting rules. Based on the Sobolev Imbedding Theorem, Lemma 5.6 of Section 5.5.1 below shows that the censor map (5.14) can be continuously extended to a map on* $\mathbb{H}$ *which satisfies Assumption 5.2 (ii). Notice, however, that* $\Upsilon(\psi_\star) = \Psi_\star$ *only if* $\|\mathrm{D}\psi_\star\|_\infty \leq c_\Upsilon$.

Let us describe now the assumptions concerning the updating of estimates $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$. The first assumption stipulates the properties of the gains $\{a_t\}_{t\in\mathbb{N}}$ involved in the algorithm (5.9), also referred to as *step sizes*.

Assumption 5.3. $\{a_t\}_{t\in\mathbb{N}}$ *is a sequence of exogenously given non-increasing positive numbers such that the following conditions hold:*

$$(i)\ a_t \to 0 \quad \text{as } t \to \infty, \quad (ii)\ \textstyle\sum_{t=0}^{\infty} a_t = \infty,$$
$$(iii)\ a_t \leq a_{t+1}(1 + \overline{a}\, a_t) \quad \text{for some } 0 < \overline{a} \leq 1.$$

A popular choice satisfying Assumption 5.3 is $a_t = \frac{1}{t}$, $t \in \mathbb{N}$. The updating functions $\{M_t\}_{t\in\mathbb{N}}$ are specified next. Let $\{\mathbb{H}_k\}_{k\in\mathbb{N}}$ be a sequence of closed, bounded, and convex subsets of $\mathbb{H}$. Suppose that $\mathbb{H}_k \subset \mathbb{H}_{k+1}$ for all $k \in \mathbb{N}$ and that the closure of $\bigcup_{k\in\mathbb{N}} \mathbb{H}_k$ with respect to the norm topology is equal to the Hilbert space $\mathbb{H}$.

Assumption 5.4. *The updating functions* $\{M_t\}_{t\in\mathbb{N}}$ *take the following form. Let* $k(\cdot) : \mathbb{N} \to \mathbb{N}$, $t \mapsto k(t)$ *be a non-decreasing function with*

$$1 \leq k(t) \leq k(t+1) \leq k(t) + 1 \quad \text{and} \quad \lim_{t\to\infty} k(t) = \infty.$$

For each $t \in \mathbb{N}$*, let*

$$M_t : \mathbb{X} \times \Sigma \times \mathbb{H} \to \mathbb{H}_{k(t+1)}$$

be a $\mathcal{B}(\mathbb{X} \times \Sigma \times \mathbb{H})/\mathcal{B}(\mathbb{H})$ *measurable and bounded map.*

Example 5.2. *Let* $\pi_k : \mathbb{H} \to \mathbb{H}_k$ *denote a projection operator, i.e.* $\pi_k \circ \pi_k = \pi_k$ *and* $M : \mathbb{X} \times \Sigma \times \mathbb{H} \to \mathbb{H}$ *be a measurable and bounded map. Then* $M_t := \pi_{k(t)} \circ M$, $t \in \mathbb{N}$ *satisfies the conditions of Assumption 5.4. Such updating functions have been proposed in Chen & White (1998, 2000)*

Assumption 5.4 allows us to introduce nonparametric recursive estimators as kernel, orthogonal series, splines, wavelets, and neural-network estimators, etc., see White, Gallant, Hornik, Stinchcombe & Woolridge (1992) or Prakasa Rao (1983) for examples.

Example 5.3. *An example for a sequence* $\{\mathbb{H}_{k(t)}\}_{t\in\mathbb{N}}$ *of closed, bounded, and convex subsets of* $\mathbb{H}$ *can be defined as follows. Recall that* $d_{\mathbb{H}}$ *is the dimension of* $\mathbb{X}\times\Sigma$ *and let* $\phi\in C^p_B(\mathbb{X}\times\Sigma;\mathbb{R})$ *be a cumulative distribution function. For each* $l=1,\ldots,d_y$, $j\in\mathbb{N}$, *let each* $A_{lj}:\mathbb{R}^{d_{\mathbb{H}}}\to\mathbb{R}$ *be a linear map and* $\eta_{lj}\in\mathbb{R}^{d_{\mathbb{H}}}$ *be a vector. Choose an increasing sequence of positive real numbers* $\{\beta_t\}_{t\in\mathbb{N}}$ *such that* $\beta_t\to\infty$ *as* $t\to\infty$. *For each* $t\in\mathbb{N}$, *define*

$$\mathbb{H}_{k(t)} := \Big\{\psi=(\psi_1,\ldots,\psi_{d_y}):\mathbb{X}\times\Sigma\to\mathbb{R}^{d_y}\,\Big|\, \psi_l(\zeta)=\sum_{j=1}^{k(t)}\alpha_{lj}\;\phi(A_{lj}\zeta+\eta_{lj}),\ \textit{where}\ \alpha_{lj}\in\mathbb{R}\ \textit{and}\ \max_{1\le j\le k(t)}\big\{\|\eta_{lj}\|,\|A_{lj}\|\big\}\le\beta_t,\ l=1,\ldots,d_y\Big\}.$$

It is shown in White, Gallant, Hornik, Stinchcombe & Woolridge (1992) that the closure of $\bigcup_{t\in\mathbb{N}}\mathbb{H}_{k(t)}$ *is equal to* $\mathbb{H}$, *whereas* $\mathbb{H}_{k(t)}\subset\mathbb{H}_{k(t+1)}$ *for all* $t\in\mathbb{N}$.

Our next assumption concerns the unknown *'limiting map'* $M_\star:\mathbb{H}\to\mathbb{H}$ which characterizes the desired unbiased forecasting rule $\Psi_\star$.

Assumption 5.5. *There exists a measurable map* $M_\star:\mathbb{H}\to\mathbb{H}$ *such that the following holds:*

(i) $M_\star$ *has a unique zero* $\psi_\star\in\mathbb{H}$, *that is,* $M_\star(\psi_\star)=0$. $\Psi_\star=\mathbf{I}(\psi_\star)$ *is the unique unbiased no-updating forecasting rule associated with the economic law* G.

(ii) $M_\star$ *is uniformly continuous on bounded sets. That is, for each bounded subset* $\mathbb{A}\subset\mathbb{H}$ *and each* $\epsilon>0$, *there exists* $\delta=\delta(\epsilon,\mathbb{A})$ *such that*

$$\|\psi-\tilde{\psi}\|_{\mathbb{H}}<\delta\ \textit{implies}\ \|M_\star(\psi)-M_\star(\tilde{\psi})\|_{\mathbb{H}}<\epsilon$$

for all ψ, $\tilde{\psi}\in\mathbb{A}$.

Assumption 5.5 (ii) implies that $M_\star$ is continuous and maps bounded sets into bounded sets. Yin & Zhu (1990) give examples of $M_\star$ that satisfy Assumption 5.5 (ii). These include linear operators, Hölder and Lipschitz operators, continuous operators with 'polynomial growth' and some compact operators.

Remark 5.5. *The limiting function $M_\star$ can be related to our mean error function $\mathcal{E}_G$ as given in Definition 2.2 of Chapter 2. A typical candidate for $M_\star$ could be of the form*

$$M_\star(\psi) := [\psi_\star - \psi]\, f(\psi), \tag{5.15}$$

where $f(\psi)$ is a suitably defined $\mathbb{R}_+$-valued function on $\mathbb{X} \times \Sigma$ which depends on the forecasting rule ψ. Thus $M_\star$ defines weighted forecast errors in terms of forecasting rules.

The next assumption may be interpreted as a 'stability' assumption which is used quite often when treating stochastic approximation problems.

Assumption 5.6. *There exists a bounded and twice continuously Fréchet differentiable functional $\mathbf{V} : \mathbb{H} \to \mathbb{R}$, also referred to as* Liapunov function. *Let $\mathbf{V}'$ denote the gradient of the Fréchet derivative*[3] *of $\mathbf{V}$. Suppose that $\mathbf{V}$ satisfies the following properties:*

(i) $\mathbf{V}(\psi_\star) = 0$ *and* $\lim_{\|\psi\|_{\mathbb{H}} \to \infty} \mathbf{V}(\psi) = \infty$.

(ii) $\mathbf{V}(\psi) > 0$ *and* $\langle \mathbf{V}'(\psi), M_\star(\psi) \rangle_{\mathbb{H}} < 0$ *for all* $\psi \in \mathbb{H}$, $\psi \neq \psi_\star$.

(iii) $\mathbf{V}(\overline{\psi}) < d_{\mathbf{V}} := \inf\{\mathbf{V}(\psi) \mid \|\psi\|_{\mathbb{H}} > \bar{b}\}$.

Assumption 5.6 implies that $\mathbf{V}$ maps bounded subsets of $\mathbb{H}$ into bounded subsets of $\mathbb{R}$. When $M_\star$ is Fréchet differentiable at the unique zero $\psi_\star$, we can choose $\mathbf{V}$ to be of a local quadratic form

$$\mathbf{V}(\psi) = \langle (\psi - \psi_\star), \mathrm{D}M_\star(\psi_\star)(\psi - \psi_\star) \rangle_{\mathbb{H}} + o\big(\|\psi - \psi_\star\|_{\mathbb{H}}^2\big),$$

where $\mathrm{D}M_\star$ is the Fréchet derivative of $M_\star$. In general, however, differentiability of $M_\star$ will not be imposed.

Example 5.4. *A typical example for a Liapunov function is the functional $\mathbf{V} : \mathbb{H} \to \mathbb{R}$, defined by $\mathbf{V}(\psi) := \|\psi - \psi_\star\|_{\mathbb{H}}^2$. It is easy to show that in this case the gradient is $\mathbf{V}'(\psi) = 2[\psi - \psi_\star]$. If, in addition, the limiting function $M_\star$ is given by (5.15), then the first two conditions of Assumptions 5.6 are satisfied.*

We are now ready to state our first result which assures that the sequence of estimates $\{\hat{\psi}_t\}_{t \in \mathbb{N}}$ remains bounded with respect to the norm $\|\cdot\|_{\mathbb{H}}$ on $\mathbb{H}$.

[3] If $\mathrm{D}\mathbf{V}(\psi)$ denotes Fréchet derivative at $\psi \in \mathbb{H}$, then the gradient $\mathbf{V}'$ is defined as

$$\mathrm{D}\mathbf{V}(\psi)(\varphi) = \langle \mathbf{V}'(\psi), \varphi \rangle_{\mathbb{H}}, \quad \psi, \varphi \in \mathbb{H},$$

see Berger (1977).

Lemma 5.1. *Let Assumptions 5.1-5.6 be satisfied and* $(x_0, Z_0) \in \mathbb{D}$, $\hat{\psi}_0 \in \mathbb{H}_0$ *be given. If*

$$\limsup_{t\to\infty} \left\| a_t \sum_{i=1}^{t} \Big[M_i\big(\mathbf{G}\big(\xi_{i+1}(\omega), x_i, Z_i, \Upsilon_i(\hat{\psi}_i)\big), \hat{\psi}_i\big) - M_\star(\hat{\psi}_i) \Big] \right\|_{\mathbb{H}} < 1$$

$\mathbb{P}$-*a.s., then*

$$\lim_{t\to\infty} \gamma(t) = \gamma_{\max} < \infty \quad \mathbb{P}\text{-}a.s.$$

for the integer-valued random variable $\gamma(t)$ *defined in (5.8). Moreover, the learning scheme (5.9) satisfies*

$$\sup_{t\in\mathbb{N}} \|\hat{\psi}_t\|_{\mathbb{H}} \le b_{\gamma_{\max}} \ \mathbb{P} - a.s.$$

The proof of Lemma 5.1 is that of Proposition 4.1 on p. 124 in Yin & Zhu (1990). By Lemma 5.1 there exists a random integer $t_{\min}(\omega) \in \mathbb{N}$ such that for all $t \ge t_{\min}(\omega)$,

$$\hat{\psi}_t = \hat{\psi}_{t-1} + a_{t-1}\, M_{t-1}(x_t, Z_t, \hat{\psi}_{t-1}).$$

In other words, after some time $t_{\min}(\omega)$ the algorithm (5.9) produces no truncations. The critical assumption which will guarantee convergence is given next.

Assumption 5.7. *Let* $(x_0, Z_0) \in \mathbb{D}$ *and* $\hat{\psi}_0 \in \mathbb{H}_0$ *be arbitrary with* $\mathbb{D}$ *as given in Assumption 5.2 (iii). Assume that*

$$\limsup_{t\to\infty} \left\| a_t \sum_{i=1}^{t} \Big[M_i\big(\mathbf{G}\big(\xi_{i+1}(\omega), x_i, Z_i, \Upsilon_i(\hat{\psi}_i)\big), \hat{\psi}_i\big) - M_\star(\hat{\psi}_i) \Big] \right\|_{\mathbb{H}} = 0$$

$\mathbb{P}$-*a.s.*

It is shown in Theorem 3.3 of Yin & Zhu (1990) that Assumption 5.7 is necessary for almost-sure convergence of the estimates. Kronecker's Lemma (see Lukacs 1975, Lemma 4.3.3, p. 96) together with Assumption 5.3 implies that Assumption 5.7 is satisfied, provided that $\mathbb{P}$-a.s.

$$\limsup_{t\to\infty} \left\| \sum_{i=1}^{t} a_i \Big[M_i\big(\mathbf{G}\big(\xi_{i+1}(\omega), x_i, Z_i, \Upsilon_i(\hat{\psi}_i)\big), \hat{\psi}_i\big) - M_\star(\hat{\psi}_i) \Big] \right\|_{\mathbb{H}} < \infty$$

for arbitrary $(x_0, Z_0) \in \mathbb{D}$, $\hat{\psi}_0 \in \mathbb{H}_0$.

We are now ready to state our first convergence result. Based on Yin & Zhu (1990, Thm. 3.2, p. 123) this result is given as Theorem 2.2 in Chen & White (2002). We present a slightly extended version in which the expectations feedback is incorporated. In the literature, this situation is also referred to as the case with correlated noise.

Theorem 5.1. *Let Assumptions 5.1-5.7 be satisfied and let* $(x_0, Z_0) \in \mathbb{D}$ *and* $\hat{\psi}_0 \in \mathbb{H}_0$ *be arbitrary initial conditions. If* $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ *is a sequence of functions in* $\mathbb{H}$ *generated by the learning scheme (5.9), then*

$$\lim_{t\to\infty} \|\hat{\psi}_t - \psi_\star\|_{\mathbb{H}} = 0 \quad \mathbb{P} - a.s.$$

If, in addition, $\lim_{t\to\infty} \Upsilon_t(\psi_\star) = \Psi_\star$, *then*

$$\lim_{t\to\infty} \|\Psi_t - \Psi_\star\|_{\mathbb{H}'} = 0 \quad \mathbb{P} - a.s.$$

Proof. The first part of the proof is that of Theorem 3.1 of Yin & Zhu (1990), see also Theorem 2.2 in Chen & White (2002). The second part follows from the uniform continuity of Υ_t, $t \in \mathbb{N}$ presumed in Assumption 5.2 (ii). □

Theorem 5.1 shows that the learning scheme (5.9) generates estimates $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ which in the strong norm topology on $\mathbb{H}$ converge almost surely to the m-th component $\psi_\star$ of the unbiased forecasting rule $\Psi_\star = \mathbf{I}(\psi_\star)$ for all initial conditions $\hat{\psi}_0 \in \mathbb{H}_0$ and thus globally on $\mathbb{H}$. Contrary to Chen & White (1998), we distinguished between the convergence of the estimates $\hat{\psi}_t$ and the convergence of the respective forecasting rules $\Psi_t = \Upsilon_t(\hat{\psi}_t)$ which are actually used for forecasting. Notice that the convergence result of Theorem 5.1 does not require that censoring can be removed as time tends to infinity.

The difficulty with Theorem 5.1 is that in practice it will be hard if not impossible to verify the crucial Assumption 5.7 directly. For this reason conditions which are easier to verify are needed. We split these conditions into assumptions concerning the updating functions and into an assumption concerning the feedback effect of forecasts.

Assumption 5.8. *The updating functions* $\{M_t\}_{t\in\mathbb{N}}$ *given in Assumption 5.4 satisfy the following additional conditions.*

(i) The sequence $\{M_t\}_{t\in\mathbb{N}}$ *is bounded on bounded sets uniformly in* t*. That is, for any bounded subset* $\mathbb{B} \subset \mathbb{X} \times \Sigma \times \mathbb{H}$, *there exists a constant* $c_{\mathbb{B}}$ *such that for each* $t \in \mathbb{N}$,

$$\|M_t(x, Z, \psi)\|_{\mathbb{H}} \le c_{\mathbb{B}} \quad \text{for all} \quad (x, Z, \psi) \in \mathbb{B}.$$

(ii) There exists a sequence $\{m_t\}_{t\in\mathbb{N}}$ *of* $\mathcal{B}(\mathbb{H})/\mathcal{B}(\mathbb{R})$*-measurable functions* $m_t : \mathbb{H} \to \mathbb{R}_+$, $t \in \mathbb{N}$, *and a constant* $b_{\mathbb{H}} \in \mathbb{R}_+$, *such that:*
(a) for all $(x, Z), (\tilde{x}, \tilde{Z}) \in \mathbb{X} \times \Sigma$, *all* $\psi \in \mathbb{H}$, *and each* $t \in \mathbb{N}$,

$$\|M_t(x, Z, \psi) - M_t(\tilde{x}, \tilde{Z}, \psi)\|_{\mathbb{H}} \le m_t(\psi) \, \big\|(x, Z) - (\tilde{x}, \tilde{Z})\big\|,$$

(b) there exists $\overline{m} \in \mathbb{R}_+$, *so that for each* $t \in \mathbb{N}$,

$$\sup\big\{m_t(\psi) \big| \psi \in \mathbb{H} \text{ with } \|\psi\|_{\mathbb{H}} \le b_{\mathbb{H}}\big\} \le \overline{m}.$$

(iii) There exists a sequence $\{h_t\}_{t\in\mathbb{N}}$ *of* $\mathcal{B}(\mathbb{X}\times\Sigma)/\mathcal{B}(\mathbb{R})$*-measurable functions* $h_t : \mathbb{X}\times\Sigma \to \mathbb{R}_+$, $t\in\mathbb{N}$, *and a constant* $\overline{h}\in\mathbb{R}_+$, *such that:*

(a) for all $(x,Z)\in\mathbb{X}\times\Sigma$, *all* ψ, $\tilde{\psi}\in\mathbb{H}$, *and each* $t\in\mathbb{N}$,

$$\|M_t(x,Z,\psi)-M_t(x,Z,\tilde{\psi})\|_{\mathbb{H}} \le h_{t+1}(x,Z)\,\|\psi-\tilde{\psi}\|_{\mathbb{H}}.$$

(b) for each $(x_0,Z_0)\in\mathbb{D}$, *with* $\mathbb{D}$ *as given in Assumption 5.2 (iii), and each differentiable* $\psi\in\mathbb{H}$ *with* $\|\psi\|_{\mathbb{H}}\le b_{\mathbb{H}}$,

$$\sup_{t\in\mathbb{N}} \mathbb{E}\big[h_{t+1}\big(\mathbf{G}(\xi_{t+1}(\cdot),x_t,Z_t,\Upsilon_t(\psi))\big)\big] \le \overline{h}$$

and

$$\lim_{t\to\infty} a_t \sum_{j=1}^{t}\Big[h_{j+1}\big(\mathbf{G}(\xi_{j+1}(\omega),x_j,Z_j,\Upsilon_j(\psi))\big) - \mathbb{E}\big[h_{j+1}\big(\mathbf{G}(\xi_{j+1}(\cdot),x_j,Z_j,\Upsilon_j(\psi))\big)\big]\Big] = 0 \quad \mathbb{P}-a.s.$$

(iv) For each $(x_0,Z_0)\in\mathbb{D}$ *and each differentiable* $\psi\in\mathbb{H}$ *with* $\|\psi\|_{\mathbb{H}}\le b_{\mathbb{H}}$,

$$\lim_{t\to\infty}\Big\|a_t\sum_{j=1}^{t}\Big[M_j(\mathbf{G}(\xi_{j+1}(\omega),x_j,Z_j,\Upsilon_j(\psi)),\psi)-M_\star(\psi)\Big]\Big\|_{\mathbb{H}} = 0 \quad \mathbb{P}-a.s.$$

The assumption on the feedback effect of forecasts concerns essentially the effect of the sequence of censor maps $\{\Upsilon_t\}_{t\in\mathbb{N}}$ on the economic law G and is formalized as follows.

Assumption 5.9. *Let* $(x_0,Z_0)\in\mathbb{D}$ *be arbitrary. Assume that a sequence of censor maps* $\{\Upsilon_t\}_{t\in\mathbb{N}}$ *is given so that for each* $\epsilon>0$, *there exists* $\delta=\delta(\epsilon)$ *such that* $\mathbb{P}$*-a.s.*

$$\big\|\mathbf{G}(t+1,\omega,x_0,Z_0,\Psi_0^t)-\mathbf{G}(t+1,\omega,x_0,Z_0,\Psi_0)\big\| < \epsilon, \quad t\in\mathbb{N} \qquad (5.16)$$

for any sequence of differentiable forecasting rules $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ *in* $\mathbb{H}$ *with* $\Psi_t = \Upsilon_t(\hat{\psi}_t)$ *which satisfies*

$$\|\hat{\psi}_t-\hat{\psi}_0\|_{\mathbb{H}} < \delta \quad \textit{for all } t\in\mathbb{N}.$$

Assumptions 5.8 and 5.9 will now be used to establish the result of Lemma 5.1 for the case of an expectations feedback. The following proposition is a variant of Lemma B.1 in Chen & White (1998).

Proposition 5.1. *Let Assumptions 5.1-5.6 and Assumptions 5.8 and 5.9 be satisfied with an arbitrarily large* $b_{\mathbb{H}}$. *Then for arbitrary* $(x_0,Z_0)\in\mathbb{X}\times\Sigma$ *and* $\hat{\psi}_0\in\mathbb{H}_0$,

$$\lim_{t\to\infty} \gamma(t) = \gamma_{\max} < \infty \quad \mathbb{P}\text{-}a.s.$$

for the integer-valued random variable $\gamma(t)$ defined in (5.8). Moreover, the estimates of the learning scheme (5.9) satisfy

$$\sup_{t\in\mathbb{N}} \|\hat{\psi}_t\|_{\mathbb{H}} \leq b_{\gamma_{\max}} \ \mathbb{P} - a.s.$$

and there exists $t_{\min}(\omega)$, so that no truncations occur for all $t \geq t_{\min}(\omega)$.

Throughout the remainder of this chapter we tacitly assume that $b_{\gamma_{\max}} \leq b_{\mathbb{H}}$. The following Theorem 5.2 shows that Assumptions 5.8 and 5.9 imply Assumption 5.7, so that convergence of the learning scheme (5.9) obtains.

Theorem 5.2. *Let Assumptions 5.1-5.6 be satisfied. Let $(x_0, Z_0) \in \mathbb{D}$ and $\hat{\psi}_0 \in \mathbb{H}_0$ be arbitrary initial conditions and $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ be a sequence of functions in $\mathbb{H}$ generated by the learning scheme (5.9). If, in addition, Assumptions 5.8 and 5.9 hold, then Assumption 5.7 is satisfied and all conclusions of Theorem 5.1 hold.*

The proofs of Proposition 5.1 and Theorem 5.2 are given in Section 5.5.2 below. In order to provide further conditions which are easier to verify than the ones given so far, we need some probabilistic prerequisites. These will be discussed next.

5.3 Some Probabilistic Prerequisites

We briefly summarize some concepts for Banach-space-valued random variables which may be found in Egghe (1984, pp. 1-9) or Yosida (1978). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ denote a real separable Banach space with norm $\|\cdot\|_{\mathbb{B}}$. $\mathbb{B}$ is endowed with the Borel σ-algebra $\mathcal{B}(\mathbb{B})$ generated by the subsets of $\mathbb{B}$ which are open in the norm topology. A $\mathcal{F}/\mathcal{B}(\mathbb{B})$-measurable map $D : \Omega \to \mathbb{B}$ is called $\mathbb{B}$-*valued random variable.* The Bochner integral of a Banach-space-valued map is defined as follows.

Definition 5.1. *Let $D : \Omega \to \mathbb{B}$ be $\mathcal{F}/\mathcal{B}(\mathbb{B})$-measurable. D is* Bochner integrable *(with respect to $\mathbb{P}$) if there exists a sequence $\{D_n\}_{n\in\mathbb{N}}$ of simple functions, i.e. $D_n := \sum_{i=1}^n h_i^n \chi_{A_i^n}$, where $h_i^n \in \mathbb{B}$, $A_i^n \in \mathcal{F}$ with $A_i^n \cap A_j^n = \emptyset$ for $i \neq j$, such that $\mathbb{P}$-a.s.*

$$(i) \ \lim_{n\to\infty} \|D - D_n\|_{\mathbb{B}} = 0, \quad (ii) \ \lim_{n\to\infty} \int_\Omega \|D(\omega) - D_n(\omega)\|_{\mathbb{B}} \, \mathbb{P}(\mathrm{d}\omega) = 0.$$

The Bochner integral *of D is defined as*

$$\int_\Omega D(\omega)\mathbb{P}(d\omega) := \lim_{n\to\infty} \int_\Omega D_n(\omega) \, \mathbb{P}(d\omega).$$

It can be shown that a $\mathbb{B}$-valued function D is Bochner integrable if and only if $\|D\|_{\mathbb{B}}$ is Lebesgue integrable, see Yosida (1978, Thm. 1, p. 133). In this case

$$\mathbb{E}[D] := \int_{\Omega} D(\omega)\, \mathbb{P}(\mathrm{d}\omega)$$

is finite and will be referred to as *expectation* of D. We denote by

$$L_p(\Omega;\mathbb{B}) \equiv L_p(\Omega,\mathcal{F},\mathbb{P};\mathbb{B}), \quad 1 \le p < \infty$$

the space of (equivalence classes of) functions $D : \Omega \to \mathbb{B}$ which are $\mathcal{F}/\mathcal{B}(\mathbb{B})$-measurable such that the map $\|D(\cdot)\|_{\mathbb{B}}^p$ is Lebesgue integrable. That is, $\|D(\cdot)\|_{\mathbb{B}} \in L_p(\Omega,\mathcal{F},\mathbb{P})$. For $1 \le p < \infty$, $L_p(\Omega;\mathbb{B})$ becomes a Banach space when endowed with the norm

$$\|D\|_{L_p} := \Big(\int_{\Omega} \|D(\omega)\|_{\mathbb{B}}^p\, \mathbb{P}(d\omega)\Big)^{\frac{1}{p}}$$

and we write $D \in L_p(\Omega;\mathbb{B})$ if $\|D\|_{L_p} < \infty$. For any sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$, let

$$L_p(\Omega,\mathcal{A};\mathbb{B}) \equiv L_p(\Omega,\mathcal{A},\mathbb{P}|\mathcal{A};\mathbb{B}), \quad 1 \le p < \infty$$

denote the space of all $\mathcal{A}/\mathcal{B}(\mathbb{B})$-measurable functions D with $D \in L_p(\Omega;\mathbb{B})$. By Theorem I.2.2.1 in Egghe (1984, p. I.2) there exists a unique map $\mathbb{E}[\cdot|\mathcal{A}] : L_1(\Omega;\mathbb{B}) \to L_1(\Omega,\mathcal{A};\mathbb{B})$ such that

$$\int_A \mathbb{E}[D|\mathcal{A}](\omega)\, \mathbb{P}(\mathrm{d}\omega) = \int_A D(\omega)\, \mathbb{P}(\mathrm{d}\omega) \quad \text{for all} \;\; A \in \mathcal{A},\; D \in L_1(\Omega;\mathbb{B}).$$

$\mathbb{E}[D|\mathcal{A}]$ is called *conditional expectation* of D.

Let $\mathbb{H}$ be a real separable Hilbert space as, for example, defined in (5.4). We call the map $W : \Omega \to \mathbb{H}$ an $\mathbb{H}$-valued random variable if W is $\mathcal{F}/\mathcal{B}(\mathbb{H})$-measurable. All concepts for Banach-space-valued random variables carry over naturally to Hilbert-space-valued random variables. With the help of Bochner-integrable random variables we will now introduce the class of *mixingale processes*, a notion first introduced by McLeish (1975) and later extended to Hilbert-space valued processes by Chen & White (1996).

Definition 5.2. *Let $\{W_t\}_{t\in\mathbb{N}}$ be a sequence of $\mathbb{H}$-valued random variables on $(\Omega,\mathcal{F},\mathbb{P})$ and $\{\mathcal{A}_t\}_{t\in\mathbb{Z}}$ be a filtration of non-decreasing sub-σ-algebras of $\mathcal{F}$. Then $\{W_t,\mathcal{A}_t\}$ is called an $L_p(\Omega;\mathbb{H})$* mixingale (process) *if there exists a bounded sequence of non-negative real constants $\{d_t\}_{t\in\mathbb{N}}$ and a sequence of non-negative real constants $\{\delta_k\}_{k\in\mathbb{N}}$ with $\delta_k \to 0$ as $k \to \infty$, such that for each $t, k \ge 0$,*

$$\big\|\mathbb{E}[W_t|\mathcal{A}_{t-k}]\big\|_{L_p} \le d_t\delta_k \quad \textit{and} \quad \big\|W_t - \mathbb{E}[W_t|\mathcal{A}_{t+k}]\big\|_{L_p} \le d_t\delta_{k+1}.$$

If, in addition, $\sum_{k=0}^{\infty} \delta_k^{\varsigma} < \infty$ for some ς with $0 < a < (1/\varsigma)$, then $\{W_t,\mathcal{A}_t\}$ is called a mixingale of size $-a$.

If W_t is $\mathcal{A}_t$ measurable, then $\{W_t, \mathcal{A}_t\}$ is an adapted $L_p(\Omega;\mathbb{H})$ mixingale and the condition on $\|W_t - \mathbb{E}[W_t|\mathcal{A}_{t+k}]\|_{L_p}$ holds automatically. It suffices to have $\delta_k = o(k^{-1/\varsigma})$ for some $a < (1/\varsigma)$ or $\delta_k = O(k^{-\lambda})$ for some $\lambda > a$ to obtain a mixingale $\{W_t, \mathcal{A}_t\}$ of size $-a$. The sequence $\{d_t\}_{t\in\mathbb{N}}$ is a sequence of magnitude indices and $\{\delta_k\}_{k\in\mathbb{N}}$ measures the decay of $\|\mathbb{E}[W_t|\mathcal{A}_{t-k}]\|_{L_p}$ to zero as $k \to \infty$. This stipulates a form of asymptotic martingale difference property. As special cases, mixingale processes include independent processes, martingale difference sequences, and certain types of moving average processes, cf. Gallant & White (1988)

Exploiting certain trade-offs between the sequences $\{d_t\}_{t\in\mathbb{N}}$ and $\{\delta_k\}_{k\in\mathbb{N}}$, one obtains various strong and weak laws of large numbers which are found in Chen & White (1996). We need their Corollary 3.8 (ii) (see p. 291) which is formulated next.

Lemma 5.2. *Let $\{W_t, \mathcal{A}_t\}$ be a $L_p(\Omega;\mathbb{H})$ mixingale in the sense of Definition 5.2 with $p \geq 2$ and $\delta_k = O\big(k^{-1/2}(\log k)^{-2}\big)$. If $\sum_{t=0}^{\infty}(a_t d_t)^2 < \infty$, where $\{a_t\}_{t\in\mathbb{N}}$ is a decreasing sequence of positive real numbers with $a_t \to 0$ as $t \to \infty$, then*

$$\Big\|a_t \sum_{k=0}^{t} W_k\Big\|_{\mathbb{H}} \to 0 \quad \mathbb{P} - a.s. \text{ as } t \to \infty.$$

To introduce a certain class of mixingales, we need a notion of stochastic dependence. Let $\mathcal{A}$ and $\mathcal{C}$ be two sub-σ-algebras of $\mathcal{F}$. As two measures of dependence define

$$\alpha(\mathcal{A},\mathcal{C}) := \sup\Big\{|\mathbb{P}(A\cap C) - \mathbb{P}(A)\mathbb{P}(C)| \;\Big|\; A \in \mathcal{A},\ C \in \mathcal{C}\Big\},$$
$$\phi(\mathcal{A},\mathcal{C}) := \sup\Big\{|\mathbb{P}(C|A) - \mathbb{P}(C)| \;\Big|\; A \in \mathcal{A},\ \mathbb{P}(A) > 0,\ C \in \mathcal{C}\Big\}.$$

With these two measures of dependence, we introduce the notion of a *mixing sequence* of random variables.

Definition 5.3. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, $\mathbb{B}$ be a real separable Banach space, and $\{D_t\}_{t\in\mathbb{Z}}$ be a sequence of $\mathbb{B}$-valued random variables defined on $(\Omega,\mathcal{F},\mathbb{P})$. Denote $\mathcal{A}_a^b := \sigma(D_t|\ a \leq t \leq b)$ for all $a \leq b$, $a, b \in \mathbb{Z}$ and define*

$$\alpha_k := \sup_{t\in\mathbb{Z}}\{\alpha(\mathcal{A}_{-\infty}^t, \mathcal{A}_{t+k}^{\infty})\} \quad \text{and} \quad \phi_k := \sup_{t\in\mathbb{Z}}\{\phi(\mathcal{A}_{-\infty}^t, \mathcal{A}_{t+k}^{\infty})\}.$$

(i) If $\lim_{k\to\infty}\alpha_k = 0$, then $\{D_t\}_{t\in\mathbb{Z}}$ is called a strongly *or* α-mixing *sequence of $\mathbb{B}$-valued random variables.*
(ii) If $\lim_{k\to\infty}\phi_k = 0$, then $\{D_t\}_{t\in\mathbb{Z}}$ is called a uniformly *or* ϕ-mixing *sequence of $\mathbb{B}$-valued random variables.*

The mixing property of sequences defined in Definition 5.3 is a form of asymptotic independence. Mixing sequences of random variables constitute another class of mixingales. Finally, we introduce the notion of *near-epoch dependency* for Hilbert-space valued random variables. The following definition is taken from Chen & White (1996).

Definition 5.4. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space,* $\mathbb{B}$ *be a real separable Banach space, and* $\mathbb{H}$ *be a real separable Hilbert space. Let* $\{D_t\}_{t\in\mathbb{Z}}$ *be a sequence of* $\mathbb{B}$*-valued random variables and* $\{W_t\}_{t\in\mathbb{N}}$ *be a sequence of* $\mathbb{H}$*-valued random variables defined on* $(\Omega, \mathcal{F}, \mathbb{P})$*, where* $\mathcal{A}_{t-k}^{t+k} := \sigma(D_s | t-k \le s \le t+k)$ *as before. Then* $\{W_t\}_{t\in\mathbb{N}}$ *is called* $L_p(\Omega;\mathbb{H})$-near-epoch dependent (NED) *on* $\{D_t\}_{t\in\mathbb{Z}}$ *if the following conditions hold:*

(i) $\|W_t\|_{L_p} < \infty$ *for each* $t \in \mathbb{N}$.
(ii) There exists a sequence of non-negative numbers $\{\delta_k\}_{k\in\mathbb{N}}$ *with* $\delta_k \to 0$ *as* $k \to \infty$ *and a sequence of non-negative numbers* $\{d_t\}_{t\in\mathbb{N}}$ *such that*

$$\left\|W_t - \mathbb{E}[W_t | \mathcal{A}_{t-k}^{t+k}]\right\|_{L_p} \le d_t \delta_k \quad \textit{for all} \;\; t, k \ge 0.$$

If, in addition, $\sum_{k=0}^{\infty} \delta_k^{\varsigma} < \infty$ *for some* ς *with* $0 < a < (1/\varsigma)$*, then* $\{W_t\}_{t\in\mathbb{N}}$ *is called* $L_p(\Omega;\mathbb{H})$*-NED on* $\{D_t\}_{t\in\mathbb{Z}}$ *of size* $-a$.

It again suffices to have $\delta_k = o(k^{-1/\varsigma})$ for some $a < (1/\varsigma)$ or $\delta_k = O(k^{-\lambda})$ for some $\lambda > a$ in order for last condition in Definition 5.4 to be satisfied. The sequence $\{d_t\}_{t\in\mathbb{N}}$ is a sequence of magnitude indices and $\{\delta_k\}_{k\in\mathbb{N}}$ measures the intertemporal dependence of $\{W_t\}_{t\in\mathbb{N}}$ on $\{D_t\}_{t\in\mathbb{Z}}$. Near-epoch dependent processes form another class of mixingales. In particular, we have the following lemma which is given as Lemma 4.3 in Chen & White (1996, p. 293).

Lemma 5.3. *Let* $\{D_t\}_{t\in\mathbb{Z}}$ *be a sequence of* $\mathbb{B}$*-valued random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$ *which satisfies one of the following two conditions:*

(i) Either $\{D_t\}_{t\in\mathbb{Z}}$ *is strongly mixing with* α_k *of size* $-b\varrho p/(\varrho - p)$ *with* $\varrho > p$*, or*
(ii) $\{D_t\}_{t\in\mathbb{Z}}$ *is uniformly mixing with* ϕ_k *of size* $-b\varrho/(\varrho - 1)$*, where* $\varrho > 1$.

Suppose $\{W_t\}_{t\in\mathbb{N}}$ *is a sequence of* $\mathbb{H}$*-valued random variables in* $L_\varrho(\Omega;\mathbb{H})$ *with zero means which is* $L_p(\Omega;\mathbb{H})$*-NED on* $\{D_t\}_{t\in\mathbb{Z}}$ *of size* $-a$*, where* $\varrho \ge p \ge 1$*. Then* $\{W_t, \mathcal{A}_t\}$*, where* $\mathcal{A}_t := \sigma(D_k | -\infty < k \le t)$*, is an* $L_p(\Omega;\mathbb{H})$ *mixingale of size* $-\min\{a, b\}$.

5.4 Sufficient Conditions for Contractions

In this section we formulate a set of more tractable conditions which assure almost-sure convergence of our learning scheme (5.9). These are a specialization of those given in Section 5.2 and are easier to verify and interpret. They are designed to assure convergence of the estimations generated by (5.9) in the case in which the economic law is a contraction. The key idea is that the forecasts as inputs of the economic law are censored in such a way that the stochastic process generated by (5.9) becomes near-epoch dependent. Let $\mathbb{S}$ be the space of feasible forecasting rules introduced in Section 5.2. For simplicity of the exposition, we assume throughout this section that censoring is carried out by the time invariant censor map (5.14) as introduced in Example 5.1, so that $\Upsilon_t \equiv \Upsilon$.

The following conditions are a modification of conditions given in Chen & White (1998, Sec. 3) and Chen & White (2002, Sec. 2). These are listed with some redundancy to enhance readability. All proofs of this Section are given in Section 5.5.3 below.

Assumption 5.10. *The exogenous noise is driven by a sequence of random variables $\{\xi_t\}_{t\in\mathbb{N}}$ on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a subset $\Xi \subset \mathbb{R}^{d_\xi}$, such that for each $t \in \mathbb{N}$, $\xi_t : \Omega \to \Xi$. Moreover:*

(i) $\{D_t\}_{t\in\mathbb{Z}}$ is a sequence of $\mathbb{B}$-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ which is either strongly mixing with α_k of size -1 or uniformly mixing with ϕ_k of size $-\frac{1}{2}$, and
(ii) the process $\{\xi_t\}_{t\in\mathbb{N}}$ is $L_2(\Omega; \mathbb{R}^{d_\xi})$-NED on $\{D_t\}_{t\in\mathbb{Z}}$ of size $-\frac{1}{2}$ and for each $t \in \mathbb{N}$, $\xi_t \in L_\varrho(\Omega; \mathbb{R}^{d_\xi})$ for some $\varrho \geq 2$.

Assumption 5.11. *The economic law G as given in (2.8) satisfies the following conditions.*

(i) $\mathbb{X} \times \Sigma \subset \mathbb{R}^{d_\mathbb{H}}$ is open, convex, and has the cone property.
(ii) For each $(x, y^e) \in \mathbb{X} \times \mathbb{Y}^m$, the map $G(\cdot, x, y^e) : \Xi \to \mathbb{X}$ satisfies the Lipschitz condition

$$\|G(\xi, x, y^e) - G(\tilde{\xi}, x, y^e)\| \leq c_1 \|\xi - \tilde{\xi}\|, \quad \xi, \tilde{\xi} \in \Xi,$$

uniformly in (x, y^e) with Lipschitz constant $c_1 \geq 0$.
(iii) For each $\xi \in \Xi$, the map $G(\xi, \cdot) : \mathbb{X} \times \mathbb{Y}^m \to \mathbb{X}$ is Lipschitz continuous uniformly in $\xi \in \Xi$, that is, there exists constants $0 \leq c_2 < 1$ and $c_3 \geq 0$, independent of ξ, such that for each $(x, y^e), (\tilde{x}, \tilde{y}^e) \in \mathbb{X} \times \mathbb{Y}^m$,

$$\|G(\xi, x, y^e) - G(\xi, \tilde{x}, \tilde{y}^e)\| \leq c_2 \|x - \tilde{x}\| + c_3 \|y^e - \tilde{y}^e\|.$$

The important part of Assumption 5.11 is that the economic law is assumed to be contractive with respect to the endogenous variable x. The following three assumptions are a mere repetition of Assumptions 5.3, 5.5, and 5.6.

Assumption 5.12. *There exists a measurable map $M_\star : \mathbb{H} \to \mathbb{H}$ such that the following holds:*

(i) $M_\star$ has a unique zero $\psi_\star \in \mathbb{H}$, that is, $M_\star(\psi_\star) = 0$. $\Psi_\star = \mathbf{I}(\psi_\star)$ is the unique unbiased no-updating forecasting rule associated with the economic law G.

(ii) $M_\star$ *is uniformly continuous on bounded sets. That is, for each bounded subset* $\mathbb{A} \subset \mathbb{H}$ *and each* $\epsilon > 0$*, there exists* $\delta = \delta(\epsilon, \mathbb{A})$ *such that*

$$\|\psi - \tilde{\psi}\|_{\mathbb{H}} < \delta \text{ implies } \|M_\star(\psi) - M_\star(\tilde{\psi})\|_{\mathbb{H}} < \epsilon$$

for all ψ, $\tilde{\psi} \in \mathbb{A}$.

Assumption 5.13. *There exists a bounded and twice continuously Fréchet differentiable functional* $\mathbf{V} : \mathbb{H} \to \mathbb{R}$*, also referred to as* Liapunov function*. Let* $\mathbf{V}'$ *denote the gradient of the Fréchet derivative of* $\mathbf{V}$*. Suppose that* $\mathbf{V}$ *satisfies the following properties:*

(i) $\mathbf{V}(\psi_\star) = 0$ *and* $\lim_{\|\psi\|_{\mathbb{H}} \to \infty} \mathbf{V}(\psi) = \infty$.
(ii) $\mathbf{V}(\psi) > 0$ *and* $\langle \mathbf{V}'(\psi), M_\star(\psi)\rangle_{\mathbb{H}} < 0$ *for all* $\psi \in \mathbb{H}$, $\psi \neq \psi_\star$.
(iii) $\mathbf{V}(\overline{\psi}) < d_{\mathbf{V}} := \inf\{\mathbf{V}(\psi) \mid \|\psi\|_{\mathbb{H}} > \bar{b}\}$.

Assumption 5.14. $\{a_t\}_{t\in\mathbb{N}}$ *is a sequence of exogenously given non-increasing positive numbers such that the following conditions hold:*

$$(i)\ a_t \to 0 \quad \text{as } t \to \infty, \quad (ii)\ \textstyle\sum_{t=0}^{\infty} a_t = \infty,$$
$$(iii)\ a_t \le a_{t+1}(1 + \overline{a}\, a_t) \quad \text{for some } \ 0 < \overline{a} \le 1.$$

The assumptions on the updating functions are modified as follows.

Assumption 5.15. *The updating functions* $\{M_t\}_{t\in\mathbb{N}}$ *given in Assumption 5.4 satisfy the following conditions.*

(i) The sequence $\{M_t\}_{t\in\mathbb{N}}$ *is bounded on bounded sets uniformly in* t*. That is, for any bounded subset* $\mathbb{B} \subset \mathbb{X} \times \Sigma \times \mathbb{H}$*, there exists a constant* $c_{\mathbb{B}}$ *such that for each* $t \in \mathbb{N}$,

$$\|M_t(x, Z, \psi)\|_{\mathbb{H}} \le c_{\mathbb{B}} \quad \text{for all } \ (x, Z, \psi) \in \mathbb{B}.$$

(ii) For any bounded subset $\mathbb{D} \subset \mathbb{X} \times \Sigma \times \mathbb{H}$*, there exists a sequence of positive real numbers* $\{c_{4,t}\}_{t\in\mathbb{N}}$ *such that for each* $t \in \mathbb{N}$ *and arbitrary* (x, Z, ψ), $(\tilde{x}, \tilde{Z}, \tilde{\psi}) \in \mathbb{D}$,

$$\big\|M_t(x, Z, \psi) - M_t(\tilde{x}, \tilde{Z}, \tilde{\psi})\|_{\mathbb{H}} \le c_{4,t}\big(\big\|(x, Z) - (\tilde{x}, \tilde{Z})\big\| + \|\psi - \tilde{\psi}\|_{\mathbb{H}}\big).$$

The sequence $\{c_{4,t}\}_{t\in\mathbb{N}}$ *satisfies*

$$\sum_{t=0}^{\infty} (c_{4,t} a_t)^2 < \infty$$

with $\{a_t\}_{t\in\mathbb{N}}$ *as given in Assumption 5.14.*

(iii) For each $(x_0, Z_0) \in \mathbb{D}$*, any* $\omega \in \Omega$*, and any differentiable* $\psi \in \mathbb{H}$ *with* $\|\psi\|_{\mathbb{H}} \leq b_{\mathbb{H}}$*, where* $b_{\mathbb{H}} \geq b_{\gamma_{\max}}$ *with* $b_{\gamma_{\max}}$ *as given by Proposition 5.1,*

$$\lim_{t\to\infty} \left\|\mathbb{E}\big[M_t\big(\mathbf{G}\big(t+1,\cdot,x_0,Z_0,\Upsilon(\psi)\big),\psi\big)\big] - M_\star(\psi)\right\|_{\mathbb{H}} = 0.$$

It is straightforward to verify that Conditions (i) and (ii) of Assumption 5.15 imply Conditions (i)-(iii) of Assumption 5.8. The main task is to show that Assumptions 5.10-5.15 imply Assumption 5.8 (iv) and Assumption 5.9. We begin with the latter Assumption which is easier to establish. The following Lemma 5.4 demonstrates that the long-run deviating effect on realizations for two different sequences of forecasting rules can be controlled under the hypotheses of Assumption 5.11.

Lemma 5.4. *Let the hypotheses of Assumption 5.11 be satisfied and* $\{\Psi_i\}_{i\in\mathbb{N}}$ *and* $\{\tilde{\Psi}_i\}_{i\in\mathbb{N}}$ *be two sequences of forecasting rules in* $\mathbb{S}$*. For for each* $t \in \mathbb{N}$*, arbitrary initial conditions* (x_0, Z_0)*,* $(\tilde{x}_0, \tilde{Z}_0) \in \mathbb{X} \times \Sigma$*, and each* $\omega \in \Omega$*, set*

$$(x_t, Z_t) = \mathbf{G}(t,\omega,x_0,Z_0,\Psi_0^{t-1}) \quad \textit{and} \quad (\tilde{x}_t, \tilde{Z}_t) = \mathbf{G}(t,\omega,\tilde{x}_0,\tilde{Z}_0,\tilde{\Psi}_0^{t-1}).$$

Then

$$\|x_t - \tilde{x}_t\| \leq c_2^t\,\|x_0 - \tilde{x}_0\| + c_3 \sum_{i=0}^{t-1} c_2^i\,\|\Psi_{t-i} - \tilde{\Psi}_{t-i}\|_\infty.$$

Lemma 5.4 implies that Assumption 5.2 (iii) can be satisfied by ensuring that all forecasts remain bounded. It follows that a bounded subset $\mathbb{D} \subset \mathbb{X}\times\Sigma$ which is forward invariant under $\mathbf{G}\big(\cdot,\Psi\big)$ for all forecasting rules $\Psi \in \mathbb{S}$ which are uniformly bounded exists. Using the Sobolev Imbedding Theorem stated in Section 5.5.1 below, Lemma 5.4 is now applied to establish Assumption 5.9.

Proposition 5.2. *If Assumption 5.11 is satisfied, then Assumption 5.9 is satisfied as well.*

As regards Assumption 5.8 (iv), we start with the observation that this assumption is satisfied if the endogenous variables together with the forecasts form a near-epoch dependent process.

Proposition 5.3. *Let Assumptions 5.10–5.15 be satisfied. If, in addition, for each differentiable* $\psi \in \mathbb{H}$ *with* $\|\psi\|_{\mathbb{H}} \leq b_{\mathbb{H}}$*, the sequence of functions*

$$\big\{\mathbf{G}\big(t,\cdot,x_0,Z_0,\Upsilon(\psi)\big)\big\}_{t\in\mathbb{N}}$$

is $L_2(\Omega;\mathbb{R}^{d_{\mathbb{H}}})$*-NED on* $\{D_t\}_{t\in\mathbb{Z}}$ *of size* $-\frac{1}{2}$ *in the sense of Definition 5.4, then Assumption 5.8 (iv) is satisfied.*

With this result, the key idea to ensure that Assumption 5.8 (iv) holds is to censor the forecasts in such a way that the process of endogenous variables and forecasts $\{(x_t, y_t^e)\}_{t\in\mathbb{N}}$ inherits the near-epoch dependency of the exogenous noise process imposed by Assumption 5.10. One way of doing so is to apply Proposition 4.4 in Kuan & White (1994) stating that contractions inherit the near-epoch dependency of the exogenous perturbations. This result will be exploited in the proof of the following proposition.

Proposition 5.4. *Let the hypotheses of Assumptions 5.10 and 5.11 be satisfied and $(x_0, Z_0) \in \mathbb{X} \times \Sigma$ and $\omega \in \Omega$ be arbitrary initial conditions. Suppose $\Psi \in \mathbb{S}$ is a forecasting rule such that the sequence $\{y_t^e\}_{t\in\mathbb{N}}$ is $L_2(\Omega; \mathbb{R}^{d_y})$-NED on $\{D_t\}_{t\in\mathbb{Z}}$ of size $-\frac{1}{2}$ in the sense of Definition 5.4, where*

$$\begin{cases} x_t &= G\big(\xi_t(\omega), x_{t-1}, \Psi(x_{t-1}, Z_{t-1})\big) \\ y_{t-1}^e &= \Psi(x_{t-1}, Z_{t-1}) \end{cases} .$$

Then the sequence of functions

$$\big\{\mathbf{G}\big(t, \cdot, x_0, Z_0, \Psi)\big)\big\}_{t\in\mathbb{N}}$$

is $L_2(\Omega; \mathbb{R}^{d_{\mathbb{H}}})$-NED on $\{D_t\}_{t\in\mathbb{Z}}$ of size $-\frac{1}{2}$.

In order to manipulate the least amount of forecasts necessary, we set $r = 0$ and modify the censor map (5.14) as introduced in Example 5.1 as follows. Recall that by Remark 5.2 it suffices to consider the space of bounded continuously differentiable forecasting rules $\psi \in C_B^p(\mathbb{X} \times \mathbb{Y}^{m-1}; \mathbb{R}^{d_y})$. Choose a positive constant $0 < c_\Upsilon < 1$ and redefine (5.13) as

$$\Upsilon' : \begin{cases} C_B^1(\mathbb{X} \times \mathbb{Y}^{m-1}; \mathbb{R}^{d_y}) & \longrightarrow C_B^1(\mathbb{X} \times \mathbb{Y}^{m-1}; \mathbb{R}^{d_y}) \\ \qquad \psi & \longmapsto \Upsilon'(\psi) \end{cases} \tag{5.17}$$

by setting

$$\Upsilon'(\psi) = \begin{cases} \psi & \text{if } \|\mathrm{D}\psi\|_\infty < c_\Upsilon \\ \frac{c_\Upsilon}{\|\mathrm{D}\psi\|_\infty}\,\psi & \text{if } \|\mathrm{D}\psi\|_\infty \geq c_\Upsilon \end{cases} .$$

As before, (5.17) maps differentiable functions onto differentiable contractions with prescribed Lipschitz constant c_Υ. For each $t \in \mathbb{N}$, define

$$\Upsilon : C_B^p(\mathbb{X} \times \mathbb{Y}^{m-1}; \mathbb{R}^{d_y}) \to \mathbb{S}, \quad \psi \mapsto \mathbf{I} \circ \Upsilon'(\psi), \tag{5.18}$$

where $\mathbf{I}$ is defined in (5.5). By Lemma 5.6 of Section 5.5.1 below, the censor map (5.18) satisfies Assumption 5.2 (ii). The following proposition shows that with the help of the censor map (5.18), the hypotheses of Proposition 5.4 can be fulfilled.

Lemma 5.5. *Let the hypotheses of Assumptions 5.10 and 5.11 be satisfied. Suppose, in addition, that* $c_2 = c_3 < 1$ *and let* $r = 0$*, so that* $\Sigma = \mathbb{Y}^m$ *and* $r = 0$*. Choose the Lipschitz constant of the censor map* Υ *defined in (5.18) such that* $(m-1)c_\Upsilon + c_2 < 1$*. Then each* $\Psi = \Upsilon(\psi)$*,* $\psi \in C_B^p(\mathbb{X} \times \mathbb{Y}^{m-1}; \mathbb{R}^{d_y})$ *satisfies the hypothesis of Proposition 5.4, so that the conclusion of Proposition 5.4 holds.*

Observe that the Lipschitz constant c_Υ cannot be chosen independently of the economic law G so that a-priori information on the Lipschitz constant c_2 of G is necessary to apply Lemma 5.5. On the other hand, it is straightforward to generalize Lemma 5.5 to the case $r > 0$.

The almost-sure convergence of the learning scheme (5.9) under the more specialized assumptions of this section is now obvious as all crucial assumptions of Section 5.2 follow from Propositions 5.2 and 5.3 and Lemma 5.5 together with Theorem 5.2.

Theorem 5.3. *Let Assumptions 5.10-5.13 be fulfilled and assume, in addition, that all hypotheses of Lemma 5.5 are satisfied. Then Assumptions 5.1-5.5 and Assumption 5.7 are fulfilled and all conclusions of Theorem 5.2 hold. In particular, if* $(x_0, Z_0) \in \mathbb{D}$ *and* $\hat{\psi}_0 \in \mathbb{H}_0$ *are arbitrary initial conditions, then the sequence* $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ *of functions in* $\mathbb{H}$ *generated by the learning scheme (5.9) converges, i.e.,*

$$\lim_{t\to\infty} \|\hat{\psi}_t - \psi_\star\|_{\mathbb{H}} = 0 \quad \mathbb{P} - a.s.$$

If, in addition, $\Upsilon(\psi_\star) = \Psi_\star$*, then*

$$\lim_{t\to\infty} \|\Psi_t - \Psi_\star\|_{\mathbb{H}'} = 0 \quad \mathbb{P} - a.s.$$

Under the hypotheses of Theorem 5.3, an approximated unbiased forecasting rule is only applied, if the system under unbiased expectations is sufficiently contractive. In this case the contraction property of Theorem 4.3 in Section 4.4 of Chapter 4, guaranteeing the existence of a random fixed point, is satisfied. In other words, the hypotheses of Theorem 5.3 are sufficient to estimate and thus learn an unbiased forecasting rule of an economic random dynamical system which admits a random fixed point under rational expectations.

Concluding Remarks

Based on nonparametric estimation techniques, this chapter developed a learning scheme for nonlinear economic systems with expectations feedback. Combining the notion of an economic law with the nonparametric estimation techniques, we established conditions under which strongly consistent estimates of an unbiased no-updating rule exist. A key element of the method was to censor the forecasts in such a way that the resulting process has properties which ensure consistent estimators. Our approach opens up the possibility to select an approximation of a preferred forecasting rule on economic grounds, as soon as parameter estimates are sufficiently precise. It thus provides the missing link for an appropriate treatment of multiple solutions which in models with expectational leads will occur generically.

An open issue is the case in which the dynamics under unbiased expectations is not governed by a contraction. This requires a method which allows to remove the restriction of the Lipschitz constant of the applied forecasting rule as soon as estimators are sufficiently precise in the sense of an ε-unbiased forecasting rule introduced in Chapter 4. As in the linear case, this requires knowledge on the asymptotic behavior of the estimation procedure. Asymptotic properties such as convergence rates of Robbins-Monro procedures are known (e.g., see Chen & White (2002) and Yin & Zhu 1990), the application to the present setup, however, is left for future research.

Summarizing, this chapter confirms the lesson from the linear case. A careful distinction between four separate issues is important for plausible learning schemes for economic systems with expectations feedback. 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, finally, the success of the learning scheme in terms of strongly consistent estimates. In essence, the mathematical object to be estimated from time series data is the economic law.

5.5 Mathematical Appendix

5.5.1 Sobolev Imbedding Theorem

We begin with a brief review of all necessary properties of Sobolev spaces. All details for the case of $\mathbb{R}$-valued functions are found in Adams (1975). The extension to vector-valued functions required for this section is straightforward, see also Yosida (1978). In view of the Sobolev imbedding theorem, which will be stated below we require that, in addition, $\mathbb{X} \times \Sigma$ satisfies the *cone property.* Recall to this end that given a point $\zeta \in \mathbb{R}^{d_{\mathbb{H}}}$, an open ball B_1 with center x, and an open ball B_2 not containing ζ, the set

$$C_\zeta := B_1 \cap \left\{ \zeta + \lambda(\eta - \zeta) \in \mathbb{R}^{d_{\mathbb{H}}} \,\middle|\, \eta \in B_2,\ \lambda > 0 \right\}$$

is called a finite cone in $\mathbb{R}^{d_{\mathbb{H}}}$ with vertex ζ. The set $\mathbb{X} \times \Sigma$ has the *cone property* if there exists a finite cone C such that each point $\zeta \in \mathbb{X} \times \Sigma$ is the vertex of a finite cone $C_\zeta \subset \mathbb{X} \times \Sigma$ which is congruent to C.

For sake of completeness we define the *segment property* of a set. This property is required, for instance, to show that the space of differentiable functions with compact support is a dense subspace of a Sobolev space (e.g., Thm. 3.18, p. 54 Adams 1975). Denote by $\overline{\mathbb{X} \times \Sigma}$ the closure of $\mathbb{X} \times \Sigma$ and by $\partial(\mathbb{X} \times \Sigma)$ its boundary. $\mathbb{X} \times \Sigma$ has the *segment property* if there exists a locally finite open cover $\{U_j\}$ of $\partial(\mathbb{X} \times \Sigma)$ and a corresponding sequence $\{\eta_j\}$ nonzero vectors such that if $\zeta \in \overline{\mathbb{X} \times \Sigma} \cap U_j$ for some j, then $\zeta + t\eta_j \in \mathbb{X} \times \Sigma$ for all $0 < t < 1$. Loosely speaking, the boundary of $\mathbb{X} \times \Sigma$ must be $(d_{\mathbb{H}} - 1)$-dimensional and cannot simultaneously lie on both sides of any given part of its boundary.

If, in addition, $\mathbb{X} \times \Sigma$ is bounded, then the cone and the segment condition are implied by the simple condition that $\partial(\mathbb{X} \times \Sigma)$ is a locally Lipschitz boundary, that is, each point on the boundary has a neighborhood which can be represented by the graph of a Lipschitz continuous function. Further details may be found in Adams (1975, Chap. IV, pp. 65-67).

An $\mathbb{R}^n$-valued function $f : \mathbb{X} \to \mathbb{R}^n$ is called square integrable if

$$\|f\|_2 := \Big(\int_{\mathbb{X} \times \Sigma} \|f(\zeta)\|^2 \, \mathrm{d}\zeta \Big)^{\frac{1}{2}},$$

where $\mathrm{d}\zeta$ means integration with respect to the Lebesgue measure. Denote by $\alpha = (\alpha_1, \ldots, \alpha_{d_{\mathbb{H}}})$ a $d_{\mathbb{H}}$-tuple of nonnegative integers α_j, also referred to as *multi-index*, and set $|\alpha| = \sum_{j=1}^{d_{\mathbb{H}}} \alpha_j$. A *differential operator* D^α of order $|\alpha|$ is then defined by setting

$$\mathrm{D}^\alpha f(\zeta) := \frac{\partial^{|\alpha|}}{\partial \zeta_1^{\alpha_1} \cdots \partial \zeta_{d_{\mathbb{H}}}^{\alpha_{d_{\mathbb{H}}}}} f(\zeta).$$

Put $\mathrm{D}^{(0,\ldots,0)} f(\zeta) = f(\zeta)$ and denote the space of all p-times continuously differentiable $\mathbb{R}^n$-valued functions by $C^p(\mathbb{X} \times \Sigma; \mathbb{R}^n)$. Define a norm on this space by

$$\|f\|_p := \Big(\sum_{0 \le |\alpha| \le p} \|\mathrm{D}^\alpha f\|_2^2 \Big)^{\frac{1}{2}}. \tag{5.19}$$

The completion of the set

$$\Big\{ f \in C^p(\mathbb{X} \times \Sigma; \mathbb{R}^n) \Big| \; \|f\|_p < \infty \Big\}$$

with respect to the norm (5.19) is commonly referred to as the Sobolev space $\mathbb{W}^{p,2}(\mathbb{X} \times \Sigma; \mathbb{R}^n)$. By Theorem 3.5 on p. 47 of Adams (1975), each space $\mathbb{W}^{p,2}(\mathbb{X} \times \Sigma; \mathbb{R}^n)$, $1 \le p < \infty$, is a separable Hilbert space.

For any $j \ge 0$, consider the norm

$$\|f\|_{\infty,j} := \max_{0 \le \alpha \le j} \sup_{\zeta \in \mathbb{X} \times \Sigma} \|\mathrm{D}^\alpha f(\zeta)\| \tag{5.20}$$

for some j-times continuously differentiable function f. For notational convenience, put $\| \cdot \|_\infty \equiv \| \cdot \|_{\infty,0}$. Then the space

$$C_B^j(\mathbb{X} \times \Sigma; \mathbb{R}^n) := \Big\{ f \in C^j(\mathbb{X} \times \Sigma; \mathbb{R}^n) \Big| \; \|f\|_{\infty,j} < \infty \Big\}$$

is a Banach space when endowed with the norm $\|\cdot\|_{\infty,j}$.

By an imbedding of $\mathbb{W}^{j+p,2}(\mathbb{X} \times \Sigma; \mathbb{R}^n)$ into $C_B^j(\mathbb{X} \times \Sigma; \mathbb{R}^n)$ is meant that each $f \in \mathbb{W}^{j+p,2}(\mathbb{X} \times \Sigma; \mathbb{R}^n)$ when interpreted as a function can be redefined on a set of measure zero in $\mathbb{X} \times \Sigma$ in such a way that the modified function $\tilde{f}$ [which equals f in $\mathbb{W}^{j+p,2}(\mathbb{X} \times \Sigma; \mathbb{R}^n)$] belongs to $C_B^j(\mathbb{X} \times \Sigma; \mathbb{R}^n)$ with an upper bound for its norm. This result is known as the *Sobolev Imbedding Theorem*. The version needed in these notes is stated next.

Theorem 5.4. *Let $\mathbb{X} \times \Sigma$ be an open subset of $\mathbb{R}^{d_{\mathbb{H}}}$ which has the cone property. Suppose $2p > d_{\mathbb{H}}$. Then the imbedding $\mathbb{W}^{j+p,2}(\mathbb{X} \times \Sigma; \mathbb{R}^n) \to C_B^j(\mathbb{X} \times \Sigma; \mathbb{R}^n)$ is continuous in the sense that*

$$\|f\|_{\infty,j} \leq K \|f\|_{j+p} \tag{5.21}$$

for all $f \in \mathbb{W}^{j+p,2}(\mathbb{X} \times \Sigma; \mathbb{R}^n)$, $j \geq 0$. The constant K depends only on $\mathbb{X} \times \Sigma$ and the parameters p and j.

The proof of Theorem 5.4 and further details are found in Adams (1975, Chap. V, p. 95). With this brief review all proofs of Section 5.4 can be carried out, setting $\mathbb{H} = \mathbb{W}^{p,2}(\mathbb{X} \times \Sigma; \mathbb{R}^{d_y})$ and $\|\cdot\|_{\mathbb{H}} = \|\cdot\|_p$. We begin with a fundamental property of the censor map (5.14)

Lemma 5.6. *Let Assumption 5.2 (i) be satisfied and let $\mathbb{A} \subset \mathbb{H}$ be a subset which is bounded with respect to the norm $\|\cdot\|_{\mathbb{H}}$. Then the following holds.*

(i) The censor map as defined in (5.14) can be continuously extended to a map $\Upsilon : \mathbb{H} \to \mathbb{H}'$ which is Lipschitz continuous when restricted to $\mathbb{A}$, so that

$$\|\Upsilon(\psi) - \Upsilon(\tilde{\psi})\|_{\mathbb{H}'} \leq K_\Upsilon^{\mathbb{H}} \|\psi - \tilde{\psi}\|_{\mathbb{H}} \quad \textit{for all} \quad \psi, \tilde{\psi} \in \mathbb{A}$$

with a suitable Lipschitz constant $K_\Upsilon^{\mathbb{H}}$.

(ii) There exists a constant K_Υ such that Υ when restricted to $\mathbb{A}$ is Lipschitz continuous with respect to the supremum norm, i.e.,

$$\|\Upsilon(\psi) - \Upsilon(\tilde{\psi})\|_{\infty} \leq K_\Upsilon \|\psi - \tilde{\psi}\|_{\mathbb{H}} \quad \textit{for all} \quad \psi, \tilde{\psi} \in \mathbb{A}.$$

Proof. Define a map $F : C_B^p(\mathbb{X} \times \Sigma; \mathbb{R}^{d_y}) \to \mathbb{R}_+$ by

$$F(\psi) = \begin{cases} 1 & \text{if } \|\mathrm{D}\psi\|_\infty < c_\Upsilon, \\ \frac{c_\Upsilon}{\|\mathrm{D}\psi\|_\infty} & \text{if } \|\mathrm{D}\psi\|_\infty \geq c_\Upsilon. \end{cases}$$

with the constant $0 < c_\Upsilon < 1$ as chosen above. Then the censor map (5.14) takes the form $\Upsilon(\psi) = \mathbf{I}\big(F(\psi)\psi\big)$. We show that Υ is Lipschitz continuous with respect to the norms $\|\cdot\|_{\mathbb{H}}$ and $\|\cdot\|_{\mathbb{H}'}$ when restricted to $\|\cdot\|_{\mathbb{H}}$-bounded subsets of $C_B^p(\mathbb{X} \times \Sigma; \mathbb{R}^{d_y})$. Since $0 < F(\psi) \leq 1$ and $\|\mathbf{I}(\psi)\|_{\mathbb{H}'} \leq \|\psi\|_{\mathbb{H}}$, we have

$$\|\Upsilon(\psi) - \Upsilon(\tilde{\psi})\|_{\mathbb{H}'} \leq \|\psi - \tilde{\psi}\|_{\mathbb{H}} + |F(\psi) - F(\tilde{\psi})| \, \|\tilde{\psi}\|_{\mathbb{H}}. \tag{5.22}$$

There are essentially three cases. Υ is clearly Lipschitz continuous in the first one when $\|\mathrm{D}\psi\|_\infty$, $\|\mathrm{D}\tilde{\psi}\|_\infty < c_\Upsilon$. In the second one $\|\mathrm{D}\psi\|_\infty$, $\|\mathrm{D}\tilde{\psi}\|_\infty \geq c_\Upsilon$ and in the third one $\|\mathrm{D}\tilde{\psi}\|_\infty < c_\Upsilon \leq \|\mathrm{D}\psi\|_\infty$. In the second case,

$$|F(\psi) - F(\tilde{\psi})| \leq c_\Upsilon \left| \frac{\|\mathrm{D}\tilde{\psi}\|_\infty - \|\mathrm{D}\psi\|_\infty}{\|\mathrm{D}\tilde{\psi}\|_\infty \|\mathrm{D}\psi\|_\infty} \right| \leq \tfrac{1}{c_\Upsilon} \|\mathrm{D}\psi - \mathrm{D}\tilde{\psi}\|_\infty. \tag{5.23}$$

In the third case,

$$|F(\psi) - 1| \leq \frac{\|\mathrm{D}\psi\|_\infty - \|\mathrm{D}\tilde{\psi}\|_\infty}{\|\mathrm{D}\psi\|_\infty} \leq \tfrac{1}{c_\Upsilon} \|\mathrm{D}\psi - \mathrm{D}\tilde{\psi}\|_\infty. \tag{5.24}$$

Since $2p > d_{\mathbb{H}}$, we may apply Theorem 5.4 for $j = 0$ to (5.23) and (5.24) to obtain

$$|F(\psi) - F(\tilde{\psi})| \leq \tfrac{1}{c_\Upsilon} \|\mathrm{D}\psi - \mathrm{D}\tilde{\psi}\|_\infty \leq K'_\Upsilon \|\psi - \tilde{\psi}\|_{\mathbb{H}}$$

for some suitable constant K'_Υ. Inserting into (5.22) gives

$$\|\Upsilon(\psi) - \Upsilon(\tilde{\psi})\|_{\mathbb{H}'} \leq \Big[1 + K'_\Upsilon \|\tilde{\psi}\|_{\mathbb{H}}\Big] \|\psi - \tilde{\psi}\|_{\mathbb{H}}.$$

This shows that Υ is Lipschitz continuous on $\|\cdot\|_{\mathbb{H}}$-bounded subsets. Since $C_B^p(\mathbb{X} \times \Sigma; \mathbb{R}^{d_y})$ is a dense subset of $\mathbb{H}$, Υ can be continuously extended to $\mathbb{H}$. This proves the first assertion. The second assertion then follows again from Theorem 5.4. □

5.5.2 Proofs of Section 5.2

We will preface the proof of Proposition 5.1 by an infinite-dimensional version of Lemma D on p. 147 of Metivier & Priouret (1984). Its proof follows along the same lines and is omitted.

Lemma 5.7. *Let $\{a_t\}_{t \in \mathbb{N}}$ be a sequence of numbers which satisfies Assumption 5.3. Let $\mathbb{A}$ be some* 'index set' *and for each $\Phi \in \mathbb{A}$, $\{L_t(\Phi)\}_{t \in \mathbb{N}}$ be a sequence of $\mathbb{H}$-valued random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that*

$$\lim_{t \to \infty} \sup_{\Phi \in \mathbb{A}} \Big\| a_t \sum_{j=1}^{t} L_j(\Phi) \Big\|_{\mathbb{H}} = 0 \quad \mathbb{P} - a.s.$$

Then for each $\alpha > 0$, there exists a random variable $N(\alpha, \omega)$, such that for all $m > t > N(\alpha, \omega)$,

$$\sup_{\Phi \in \mathbb{A}} \Big\| \sum_{j=t}^{m-1} a_j L_j(\Phi) \Big\|_{\mathbb{H}} \leq \alpha \Big(1 + \sum_{j=t}^{m-1} a_j \Big) \quad \mathbb{P} - a.s.$$

Proof of Proposition 5.1. *Step 1.* Assume on the contrary that $\gamma_{\max} = \infty$, so that the sequence of estimates $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ crosses the sphere $\{\psi \in \mathbb{H} |\ \|\psi\|_{\mathbb{H}} \leq \bar{b}\}$ infinitely often. As a consequence $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ cannot converge to $\psi_\star$ so that by Assumption 5.6 (ii), there exists $\eta > 0$ such that

$$|\langle \mathbf{V}'(\hat{\psi}_t), M_\star(\hat{\psi}_t)\rangle_{\mathbb{H}}| > \eta \quad \mathbb{P}-\text{a.s. for all } \ t \in \mathbb{N}. \tag{5.25}$$

By Assumption 5.6 (iii), $(\mathbf{V}(\overline{\psi}), d_{\mathbf{V}})$ is a non-empty open interval. It follows from Assumption 5.2 (iii) and Assumption 5.8 (i) that there exists c_M such that

$$\big\|M_t(x, Z, \psi)\big\|_{\mathbb{H}} \leq c_M \quad \text{for all } \ (x, Z) \in \mathbb{D},\ \|\psi\| \leq \bar{b}. \tag{5.26}$$

For $\delta > 0$, choose t_0 large enough so that $a_t < \delta$ for all $t \geq t_0$ and $\sigma_1 \leq \sigma_2$ with $[\sigma_1, \sigma_2] \subset (V(\overline{\psi}), d_{\mathbf{V}})$. Define

$$t_1 := \min\Big\{t \geq t_0 \ \Big|\ \mathbf{V}(\hat{\psi}_t) \geq \sigma_1\Big\},$$
$$t_2 := \min\Big\{t \geq t_0 \ \Big|\ \mathbf{V}(\hat{\psi}_t) \geq \sigma_2\Big\}.$$

By our initial assumption, t_1 and t_2 are well defined and finite. t_1 describes the first entry of the sequence $\mathbf{V}(\hat{\psi}_t)$ into the interval $[\sigma_1, \sigma_2]$ after time t_0 and t_2 the first exit. The sequence $\{\hat{\psi}_t\}_{t=t_0}^{t_2-1}$ is bounded by $\bar{b}$. (If this were not the case and $\|\hat{\psi}_{t'}\|_{\mathbb{H}} > \bar{b}$ for some $t_1 \leq t' < \tau_2$, then $\mathbf{V}(\hat{\psi}_{t'}) > d_{\mathbf{V}}$, a contradiction.) The continuity of $\mathbf{V}$ and (5.26) imply that δ and hence t_0 can be chosen such that $t_2 - t_1 > 1$. For $\tau > \delta$ and $i \geq t_0$, set

$$\mathrm{n}(i, \tau) := \max\Big\{n > i \Big|\ \sum_{j=i}^{n-1} a_j \leq \tau\Big\}. \tag{5.27}$$

Then for $\tau > \delta$ small enough, we have

$$t_1 - 1 < \mathrm{n}(t_1 - 1, \tau) < t_2.$$

As a consequence, $\mathbf{V}(\hat{\psi}_{\mathrm{n}(t_1-1,\tau)}) \in [\sigma_1, \sigma_2]$. We next will show in two steps that $\mathbf{V}(\hat{\psi}_{\mathrm{n}(t_1-1,\tau)}) < \sigma_1$. This contradicts the assumption $\gamma_{\max} = \infty$ and thus proves the assertions of the proposition.

Step 2. We will next apply Lemma 5.7 several times. From Step 1. we now that $\tau > \delta$ are such that no truncations occur between $t = i \geq t_1 - 1$ and $t = \mathrm{n}(i, \tau)$. Let $n \leq \mathrm{n}(i, \tau)$. In this case

$$\hat{\psi}_n - \hat{\psi}_i = \sum_{j=i}^{n-1} a_j\, M_j(\mathbf{G}(j+1, \omega, x_0, Z_0, \Psi_0^j), \hat{\psi}_j). \tag{5.28}$$

Since $\{\hat{\psi}_t\}_{t=i}^{n}$ is bounded by $\bar{b}$, it follows from (5.26) that

$$\|\hat{\psi}_n - \hat{\psi}_i\|_{\mathbb{H}} \leq \sum_{j=i}^{n-1} a_j c_M \leq c_M \tau. \tag{5.29}$$

For each $j > i$, write

$$\Psi_{0i}^j := (\Psi_0, \ldots, \Psi_i, \underbrace{\Psi_i, \ldots, \Psi_i}_{(j-i)-\text{times}}) \tag{5.30}$$

for a series of forecasting rules which is constant after the i-th entry. Decompose the sum of the r.h.s. of (5.28) into

$$\hat{\psi}_n - \hat{\psi}_i = S_o(i,n) + \sum_{j=i}^{n-1} a_j\, M_\star(\hat{\psi}_i), \tag{5.31}$$

where $S_o(i,n) = S_1(i,n) + S_2(i,n) + S_3(i,n)$ with

$$S_1(i,n) := \sum_{j=i}^{n-1} a_j \big[M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_0^j),\hat{\psi}_j) \tag{5.32}$$

$$-M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j),\hat{\psi}_j)\big],$$

$$S_2(i,n) := \sum_{j=i}^{n-1} a_j \big[M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j),\hat{\psi}_j) \tag{5.33}$$

$$-M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j),\hat{\psi}_i)\big],$$

$$S_3(i,n) := \sum_{j=i}^{n-1} a_j \big[M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j),\hat{\psi}_i) - M_\star(\hat{\psi}_i)\big]. \tag{5.34}$$

The first term (5.32) takes account of the expectations feedback. By Assumption 5.9, for any $\epsilon > 0$, there exists $\delta' = \delta'(\epsilon)$ such that

$$\|\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_0^j) - \mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j)\| \leq \epsilon.$$

for all $j = i, \ldots, n-1$, provided that $\tau \leq \frac{\delta'}{c_M}$. Choose $\alpha > 0$. By Assumption 5.8 (ii) and Lemma 5.7, there exists $N_1(\alpha,\omega) > t_0$ such that for each $i > N_1(\alpha,\omega)$ and each $n \leq \mathrm{n}(i,\tau)$,

$$\|S_1(i,n)\|_{\mathbb{H}} \leq \sum_{j=i}^{n-1} a_j\, m_j(\hat{\psi}_j)\epsilon \leq \overline{m}\epsilon\,\tau \quad \mathbb{P}-\text{a.s.} \tag{5.35}$$

Concerning the second term (5.33), write for simplicity of notation

$$\big(x_{j+1}(\omega), Z_{j+1}(\omega)\big) = \mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j) \quad \text{for} \quad j \geq i.$$

Assumption 5.8 (iii) and (5.29) then imply $\mathbb{P}$-a.s.

$$\|S_2(i,n)\|_{\mathbb{H}} \;\leq\; \sum_{j=i}^{n-1} a_j\; h_{j+1}\big(x_{j+1}(\omega), Z_{j+1}(\omega)\big)\,\|\hat{\psi}_j - \hat{\psi}_i\|_{\mathbb{H}}$$

$$\leq \sum_{j=i}^{n-1} a_j \Big[h_{j+1}\big(x_{j+1}(\omega), Z_{j+1}(\omega)\big) - \mathbb{E}\big[h_{j+1}\big(x_{j+1}(\cdot), Z_{j+1}(\cdot)\big)\big]\Big]\|\hat{\psi}_j - \hat{\psi}_i\|_{\mathbb{H}}$$

$$+ \sum_{j=i}^{n-1} a_j \mathbb{E}\big[h_{j+1}\big(x_{j+1}(\cdot), Z_{j+1}(\cdot)\big)\big]\|\hat{\psi}_j - \hat{\psi}_i\|_{\mathbb{H}}.$$

By Lemma 5.7, there exists $N_2(\alpha,\omega) > N_1(\alpha,\omega)$ such that for each $i > N_2(\alpha,\omega)$ and each $n \leq \mathrm{n}(i,\tau)$,

$$\|S_2(i,n)\|_{\mathbb{H}} \leq \alpha c_M(1+\tau)\tau + \overline{h}c_M\tau^2 \quad \mathbb{P}-\text{a.s.} \tag{5.36}$$

Concerning the third term (5.34), Assumption 5.8 (iv) and Lemma 5.7 imply the existence of $N_3(\alpha,\omega) > N_2(\alpha,\omega)$ such that for each $i > N_3(\alpha,\omega)$ and each $n \leq \mathrm{n}(i,\tau)$,

$$\|S_3(i,n)\|_{\mathbb{H}} \leq \alpha(1+\tau) \quad \mathbb{P}-\text{a.s.} \tag{5.37}$$

Combining the estimates (5.35)-(5.37), we get

$$\|S_o(i,n)\|_{\mathbb{H}} \leq \alpha(1+\tau+c_M\tau) + \overline{m}\epsilon\,\tau + (\alpha+\overline{h})c_M\tau^2 \quad \mathbb{P}-\text{a.s.} \tag{5.38}$$

Step 3. The Taylor expansion of $\mathbf{V}$ at $\hat{\psi}_{t_1-1}$ gives

$$\begin{aligned}\mathbf{V}(\hat{\psi}_{\mathrm{n}(t_1-1,\tau)}) =& \mathbf{V}(\hat{\psi}_{t_1-1}) + \langle \mathbf{V}'(\hat{\psi}_{t_1-1}), [\hat{\psi}_{\mathrm{n}(t_1-1,\tau)} - \hat{\psi}_{t_1-1}]\rangle_{\mathbb{H}} \\ &+ \mathcal{R}_2(\hat{\psi}_{t_1-1}, \hat{\psi}_{\mathrm{n}(t_1-1,\tau)} - \hat{\psi}_{t_1-1}),\end{aligned} \tag{5.39}$$

where

$$\mathcal{R}_2(\psi,\varphi) = \int_0^1 (1-s)\mathrm{D}^2(\psi+s\varphi)(\varphi,\varphi)\mathrm{d}s, \quad \psi,\varphi\in\mathbb{H},$$

e.g., see Lang (1968, Chap. XVI) or Berger (1977).[4] In view of (5.29), there exists κ_2 so that

$$\begin{aligned}|\mathcal{R}_2(\hat{\psi}_{t_1-1}, \hat{\psi}_{\mathrm{n}(t_1-1,\tau)} - \hat{\psi}_{t_1-1})| &\leq \kappa_2\|\hat{\psi}_{\mathrm{n}(t_1-1,\tau)} - \hat{\psi}_{t_1-1}\|^2_{\mathbb{H}}\int_0^1(1-s)\mathrm{d}s \\ &\leq \kappa_2 c_M^2\tau^2\end{aligned} \tag{5.40}$$

Using (5.31), we can decompose the second term in (5.39) into

$$\langle \mathbf{V}'(\hat{\psi}_{t_1-1}), [\hat{\psi}_{\mathrm{n}(t_1-1,\tau)} - \hat{\psi}_{t_1-1}]\rangle_{\mathbb{H}} = \mathcal{I} + \sum_{j=t_1-1}^{\mathrm{n}(t_1-1,\tau)-1} a_j\langle \mathbf{V}'(\hat{\psi}_{t-1}), M_\star(\hat{\psi}_{t_1-1})\rangle_{\mathbb{H}}$$

---

[4] Here $\mathrm{D}^2\mathbf{V}(\psi)(\varphi_1,\varphi_2)$ denotes the second Fréchet derivative at $\psi\in\mathbb{H}$ applied to $(\varphi_1,\varphi_2)\in\mathbb{H}\times\mathbb{H}$.

with

$$\mathcal{I} := \langle \mathbf{V}'(\hat{\psi}_{t_1-1}), S_o(t_1-1, \mathrm{n}(t_1-1,\tau)) \rangle_{\mathbb{H}}.$$

Using (5.38), we have

$$|\mathcal{I}| \leq \kappa_1 \big[\alpha(1+\tau+c_M\tau) + \epsilon\overline{m}\tau + (\alpha+\overline{h})c_M\tau^2\big] \tag{5.41}$$

with

$$\kappa_1 := \sup\big\{\|\mathbf{V}'(\psi)\|_{\mathbb{H}} \;\big|\; \|\psi\|_{\mathbb{H}} \leq \overline{b}\big\}.$$

Observe that by definition of $\mathrm{n}(t_1-1,\tau)$ in (5.27),

$$\tau - a_{\mathrm{n}(t_1-1,\tau)-1} \leq \sum_{j=t_1-1}^{\mathrm{n}(t_1-1,\tau)-1} a_j \leq \tau.$$

Hence, inserting (5.40) and (5.41) into (5.39), we get

$$\begin{aligned}\mathbf{V}(\hat{\psi}_{\mathrm{n}(t_1-1,\tau)}) \leq\; & \sigma_1 + |\mathcal{I}| - \eta[\tau - a_{\mathrm{n}(t_1-1,\tau)-1}] + \kappa_2 c_M^2\tau^2 \\ \leq\; & \sigma_1 + \alpha\kappa_1[1+\tau+c_M\tau] + \eta\, a_{\mathrm{n}(t_1-1,\tau)-1} \\ & + \Big[(\kappa_2 c_M^2 + (\alpha+\overline{h})\kappa_1 c_M)\tau - \eta + \epsilon\overline{m}\kappa_1)\Big]\tau\end{aligned}$$

for η as defined in (5.25). Let

$$\epsilon < \tfrac{\eta}{\overline{m}\kappa_1} \quad \text{and} \quad \tau < \min\Big\{\tfrac{\eta-\epsilon\overline{m}\kappa_1}{(\kappa_2 c_M + (\alpha+\overline{h})\kappa_1)c_M}, \tfrac{\delta'}{c_M}\Big\}$$

and choose α and δ such that

$$\alpha\kappa_1[1+\tau+c_M\tau] + \delta\eta < \Big[\eta - \epsilon\overline{m}\kappa_1 - (\kappa_2 c_M + (\alpha+\overline{h})\kappa_1)c_M\tau\Big]\tau. \tag{5.42}$$

By our initial assumption we may choose t_0 so large that $a_t < \delta$ for all $t \geq t_0$. Hence (5.42) implies $\mathbf{V}(\hat{\psi}_{\mathrm{n}(t_1-1,\tau)}) < \sigma_1$ which proves our contradiction stated at the end of Step 1. □

Proof of Theorem 5.2. The proof of the proposition relies on a repeated application of Lemma 5.7.
Step 1. Recall that by Assumption 5.2 (iii), the sequence

$$\Big\{\mathbf{G}(t+1, \omega, x_0, Z_0, \Psi_0^t)\Big\}_{t\in\mathbb{N}} \tag{5.43}$$

with $\Psi_t = \Upsilon_t(\hat{\psi})$, $t \in \mathbb{N}$ remains bounded in some set $\mathbb{D}$. By Proposition 5.1 there exists a random integer $N_0(\omega)$ from where on no truncation of the learning algorithm occurs such that

$$\hat{\psi}_{t+1} = \hat{\psi}_t + a_t M_t\big(\mathbf{G}(t+1,\omega,x_0,Z_0,\Psi_0^t), \hat{\psi}_t\big) \quad \text{for all} \;\; t \geq N_0(\omega) \tag{5.44}$$

and a constant $b_{\mathbb{H}}$ so that $\|\hat{\psi}_t\|_{\mathbb{H}} \leq b_{\mathbb{H}}$ for all $t \geq N_0(\omega)$. The two boundedness properties (5.43) and (5.44) together with the boundedness property of M_t stated in Assumption 5.8 (i) imply the existence of a constant $c_M > 0$ such that

$$\big\|M_t\big(\mathbf{G}(t+1,\omega,x_0,Z_0,\Psi_0^t),\hat{\psi}_t\big)\big\|_{\mathbb{H}} \leq c_M \quad \mathbb{P}-\text{a.s.} \tag{5.45}$$

for all $t \geq N_0(\omega)$.

Step 2. The idea is to show that Assumption 5.7 holds, as this ensures almost sure convergence by Theorem 5.1. For arbitrary $t < n$, set

$$S(t,n) := \sum_{j=t}^{n-1} a_j \big[M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_0^j),\hat{\psi}_j) - M_\star(\hat{\psi}_j)\big]. \tag{5.46}$$

In order to verify Assumption 5.7, it suffices to show that for arbitrary but fixed $T > 0$,

$$\lim_{t\to\infty} \sup_{t<n\leq \mathrm{n}(t,T)} \big\|S(t,n)\big\|_{\mathbb{H}} = 0 \quad \mathbb{P}-\text{a.s.}, \tag{5.47}$$

where for each $t \in \mathbb{N}$,

$$\mathrm{n}(t,T) := \max\Big\{i > t \Big| \sum_{j=t}^{i-1} a_j \leq T\Big\}. \tag{5.48}$$

We will apply Lemma 5.7 several times to prove (5.47). By Assumption 5.5, for each $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ so that

$$\|M_\star(\hat{\psi}_i) - M_\star(\hat{\psi}_j)\|_{\mathbb{H}} < \epsilon \quad \text{whenever} \quad \|\hat{\psi}_i - \hat{\psi}_j\|_{\mathbb{H}} < \delta. \tag{5.49}$$

Moreover, by Assumption 5.9, there exists $\delta' = \delta'(\epsilon)$ so that (5.16) holds. Fix an arbitrary $\alpha \in (0,1)$ such that $\sqrt{\alpha}\,T c_M < \min\{\delta,\delta'\}$. Set $\tau = \sqrt{\alpha}\,T$ and define recursively

$$i_0 = t,\ i_1 = \mathrm{n}(i_0,\tau),\ldots,\ i_r = \mathrm{n}(i_{r-1},\tau),\ldots \tag{5.50}$$

Using the fact that by Step 1 no truncations occur from $N_0(\omega)$ on, it follows from (5.45) that for each $i_0 = t \geq N_0(\omega)$, each i_r, and each $i \in (i_r, i_{r+1}]$,

$$\begin{aligned}\|\hat{\psi}_i - \hat{\psi}_{i_r}\|_{\mathbb{H}} &\leq \sum_{j=i_r}^{i-1} a_j \big\|M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_0^j),\hat{\psi}_j)\big\|_{\mathbb{H}} \\ &\leq \sum_{j=i_r}^{i-1} a_j\, c_M \;\leq\; \tau\, c_M < \min\{\delta,\delta'\} \quad \mathbb{P}-\text{a.s.}\end{aligned} \tag{5.51}$$

Step 3. For each $j > i$, write

$$\Psi_{0i}^{j} := (\Psi_0, \ldots, \Psi_i, \underbrace{\Psi_i, \ldots, \Psi_i}_{(j-i)-\text{times}}) \tag{5.52}$$

for a series of forecasting rules which is constant after the i-th entry. Now let $i < i'$ be arbitrary and decompose the sum in (5.46) into four terms

$$S(i,i') := S_1(i,i') + S_2(i,i') + S_3(i,i') + S_4(i,i')$$

with and

$$S_1(i,i') := \sum_{j=i}^{i'-1} a_j \big[M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_0^j),\hat{\psi}_j) - M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j),\hat{\psi}_j)\big], \tag{5.53}$$

$$S_2(i,i') := \sum_{j=i}^{i'-1} a_j \big[M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j),\hat{\psi}_j) - M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j),\hat{\psi}_i)\big], \tag{5.54}$$

and

$$S_3(i,i') := \sum_{j=i}^{i'-1} a_j \big[M_j(\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j),\hat{\psi}_i) - M_\star(\hat{\psi}_i)\big], \tag{5.55}$$

$$S_4(i,i') := \sum_{j=i}^{i'-1} a_j \big[M_\star(\hat{\psi}_i) - M_\star(\hat{\psi}_j)\big]. \tag{5.56}$$

The first term (5.53) takes account of the expectations feedback. Assumption 5.9 together with (5.51) implies for any $i_r \le j < i_{r+1}$,

$$\|\mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_0^j) - \mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i_r}^j)\| \le \epsilon.$$

By Assumption 5.8 (ii) and Lemma 5.7, there exists $N_1(\alpha,\omega) > N_0(\alpha,\omega)$ such that for each $i_0 = t > N_1(\alpha,\omega)$, each i_r and each $i \in (i_r, i_{r+1}]$,

$$\|S_1(i_r,i)\|_{\mathbb{H}} \le \sum_{j=i_r}^{i-1} a_j\, m_j(\hat{\psi}_j)\epsilon \le \overline{m}\tau\epsilon \quad \mathbb{P}-\text{a.s.} \tag{5.57}$$

Concerning the second term (5.54), write for simplicity of notation

$$\big(x_{j+1}(\omega), Z_{j+1}(\omega)\big) = \mathbf{G}(j+1,\omega,x_0,Z_0,\Psi_{0i}^j) \quad \text{for} \quad j \ge i.$$

Assumption 5.8 (iii) and (5.51) then imply $\mathbb{P}$-a.s.

$$\|S_2(i_r,i)\|_{\mathbb{H}} \;\le\; \sum_{j=i_r}^{i-1} a_j\, h_{j+1}\big(x_{j+1}(\omega), Z_{j+1}(\omega)\big)\,\|\hat\psi_j - \hat\psi_{i_r}\|_{\mathbb{H}}$$

$$\le \sum_{j=i_r}^{i-1} a_j\Big[h_{j+1}\big(x_{j+1}(\omega), Z_{j+1}(\omega)\big) - \mathbb{E}\big[h_{j+1}\big(x_{j+1}(\cdot), Z_{j+1}(\cdot)\big)\big]\Big]\|\hat\psi_j - \hat\psi_{i_r}\|_{\mathbb{H}}$$

$$+ \sum_{j=i_r}^{i-1} a_j\mathbb{E}\big[h_{j+1}\big(x_{j+1}(\cdot), Z_{j+1}(\cdot)\big)\big]\|\hat\psi_j - \hat\psi_{i_r}\|_{\mathbb{H}}.$$

By Lemma 5.7, there exists $N_2(\alpha,\omega) > N_1(\alpha,\omega)$ such that for each $i_0 = t > N_2(\alpha,\omega)$, each i_r, and each $i \in (i_r, i_{r+1}]$,

$$\|S_2(i_r,i)\|_{\mathbb{H}} \le \alpha(1+\tau)\tau c_M + \tau^2\overline{h}c_M \quad \mathbb{P}-\text{a.s.} \tag{5.58}$$

Concerning the third term (5.55), Assumption 5.8 (iv) and Lemma 5.7 imply the existence of $N_3(\alpha,\omega) > N_2(\alpha,\omega)$ such that for each $i_0 = t > N_3(\alpha,\omega)$, each i_r, and each $i \in (i_r, i_{r+1}]$,

$$\|S_3(i_r,i)\|_{\mathbb{H}} \le \alpha(1+\tau) \quad \mathbb{P}-\text{a.s.} \tag{5.59}$$

Concerning the fourth term (5.56), (5.51) together with (5.49) imply

$$\|S_4(i_r,i)\|_{\mathbb{H}} \le \sum_{j=i_r}^{i-1} a_j\,\|M_\star(\hat\psi_i) - M_\star(\hat\psi_j)\|_{\mathbb{H}} \le \sum_{j=i_r}^{i-1} a_j\,\epsilon \le \tau\epsilon. \tag{5.60}$$

Step 4. Combining the estimates (5.60)–(5.57), there exists constants β_1 and β_2 such that for all $i_r > i_0 = t > N_3(\alpha,\omega)$,

$$\sup_{i_r < i \le i_{r+1}} \big\|S(i_r,i)\big\|_{\mathbb{H}} \le \sqrt{\alpha}\,\epsilon\beta_1 + \alpha\beta_2 \quad \mathbb{P}-\text{a.s.}$$

It follows from the definition of $\mathrm{n}(t,T)$ that for all $i_r > i_0 = t > N_3(\alpha,\omega)$,

$$\sup_{t<n\le \mathrm{n}(t,T)} \big\|S(t,n)\big\|_{\mathbb{H}} \le \Big(1 + \tfrac{2}{\sqrt{\alpha}}\Big)\,(\sqrt{\alpha}\,\epsilon\beta_1 + \alpha\beta_2)$$
$$= (2+\sqrt{\alpha})\epsilon\beta_1 + (\alpha + 2\sqrt{\alpha})\beta_2.$$

Since ϵ and α can be made arbitrarily small, this proves (5.47) and the conclusion follows from Theorem 5.1. □

5.5.3 Proofs of Section 5.4

Proof of Lemma 5.4. Let $(x_0, Z_0), (\tilde x_0, \tilde Z_0) \in \mathbb{X} \times \Sigma$, $\omega \in \Omega$ be arbitrary but fixed and $\{\Psi_i\}_{i\in\mathbb{N}}$ and $\{\tilde\Psi_i\}_{i\in\mathbb{N}}$ be two arbitrary sequences of forecasting rules in $\mathbb{S}$. Let

$$y_t^e = \Psi_t(x_t, Z_t) \quad \text{and} \quad \tilde{y}_t^e = \tilde{\Psi}_t(\tilde{x}_t, \tilde{Z}_t).$$

By assumption

$$\|x_t - \tilde{x}_t\| \leq c_2\|x_{t-1} - \tilde{x}_{t-1}\| + c_3\|y_{t-1}^e - \tilde{y}_{t-1}^e\|$$

and by induction on t,

$$\|x_t - \tilde{x}_t\| \leq c_2^t \|x_0 - \tilde{x}_0\| + c_3 \sum_{i=0}^{t-1} c_2^i \|y_{t-i}^e - \tilde{y}_{t-i}^e\|.$$

The assertion then follows from the definition of the maximum norm $\|\cdot\|_\infty$. □

Proof of Proposition 5.2. Let $\delta > 0$ be arbitrary and $\{\hat{\psi}_t\}_{t\in\mathbb{N}}$ be a sequence of forecasting rules so that

$$\|\hat{\psi}_t - \hat{\psi}_0\|_{\mathbb{H}} < \delta \quad \text{for all} \;\; t \in \mathbb{N}.$$

Put $\Psi_t = \Upsilon(\hat{\psi}_t)$ for each $t \in \mathbb{N}$ and let

$$(x_t, Z_t) = \mathbf{G}(t, \omega, x_0, Z_0, \Psi_0^{t-1}) \quad \text{and} \quad (\tilde{x}_t, \tilde{Z}_t) = \mathbf{G}(t, \omega, x_0, Z_0, \Psi_0),$$

where

$$Z_t = (y_{t-1}^e, x_{t-1}, \ldots, y_{t-r}^e, x_{t-r}) \quad \text{and} \quad \tilde{Z}_t = (\tilde{y}_{t-1}^e, \tilde{x}_{t-1}, \ldots, \tilde{y}_{t-r}^e, \tilde{x}_{t-r}).$$

By Lemma 5.4 and Lemma 5.6,

$$\begin{aligned}
\|x_t - \tilde{x}_t\|^2 + \|y_t^e - \tilde{y}_t^e\|^2 &\leq \Big(c_3 \sum_{i=0}^{t-1} c_2^i \|\Psi_{t-i} - \Psi_0\|_\infty\Big)^2 + \|\Psi_t - \Psi_0\|_\infty^2 \\
&\leq \Big(c_3 K_\Upsilon \sum_{i=0}^{t-1} c_2^i \|\hat{\psi}_{t-i} - \hat{\psi}_0\|_{\mathbb{H}}\Big)^2 + K_\Upsilon^2 \|\hat{\psi}_t - \hat{\psi}_0\|_{\mathbb{H}}^2 \\
&\leq \Big(1 + \tfrac{c_3}{1-c_2}\Big)^2 (K_\Upsilon \delta)^2.
\end{aligned}$$

For arbitrary $\epsilon > 0$, choose $\delta := \epsilon(1 - c_2)/[K_\Upsilon(1 - c_2 + c_3)\sqrt{1 + r}]$. Then

$$\big\|\mathbf{G}(t, \omega, x_0, Z_0, \Psi_0^{t-1}) - \mathbf{G}(t, \omega, x_0, Z_0, \Psi_0)\big\| < \epsilon$$

and Assumption 5.9 is satisfied. □

Proof of Proposition 5.3. *Step 1.* Let $(x_0, Z_0) \in \mathbb{X} \times \Sigma$, $\omega \in \Omega$, and a differentiable $\psi \in \mathbb{H}$ with $\|\psi\|_{\mathbb{H}} \leq b_{\mathbb{H}}$ be arbitrary. For ease of notation, write

$$\big(x_t(\omega), Z_t(\omega)\big) = \mathbf{G}\big(t, \omega, x_0, Z_0, \Upsilon(\psi)\big), \quad t \in \mathbb{N}$$

and for arbitrary $t < n$, set

$$S(t,n) := \sum_{j=t}^{n-1} a_j \left[M_j(x_{j+1}(\omega), Z_{j+1}(\omega), \psi) - M_\star(\psi)\right]. \tag{5.61}$$

In order to verify Assumption 5.8 (iv), it suffices to show that for arbitrary but fixed $T > 0$,

$$\lim_{t\to\infty} \sup_{t<n\le \mathrm{n}(t,T)} \big\|S(t,n)\big\|_{\mathbb{H}} = 0 \quad \mathbb{P}-\text{a.s.}, \tag{5.62}$$

where $\mathrm{n}(t,T)$ is defined in (5.48). To this end decompose (5.61) into

$$S(t,n) = S_1(t,n) + S_2(t,n),$$

where

$$S_1(t,n) = \sum_{j=t}^{n-1} a_j \Big[M_j(x_{j+1}(\omega), Z_{j+1}(\omega), \psi) \tag{5.63}$$

$$-\mathbb{E}\big[M_j(x_{j+1}(\cdot), Z_{j+1}(\cdot), \psi)\big]\Big],$$

$$S_2(t,n) = \sum_{j=t}^{n-1} a_j \Big[\mathbb{E}\big[M_j(x_{j+1}(\cdot), Z_{j+1}(\cdot), \psi)\big] - M_\star(\psi)\Big]. \tag{5.64}$$

Step 2. By assumption

$$\big\{\mathbf{G}\big(t, \cdot, x_0, Z_0, \Upsilon(\psi)\big)\big\}_{t\in\mathbb{N}}$$

is a sequence of bounded functions which is $L_2(\Omega;\mathbb{R}^{d_{\mathbb{H}}})$-NED on $\{D_t\}_{t\in\mathbb{Z}}$ of size $-\frac{1}{2}$, where $\mathcal{A}_{k_1}^{k_2} = \sigma(D_t|\ k_1 \le t \le k_2)$, $k_1 < k_2$ as before. By the minimum-mean-squared-error property and Assumption 5.15 (ii), we have

$$\big\|M_t(x_{t+1}(\cdot), Z_{t+1}(\cdot), \psi) - \mathbb{E}\big[M_t(x_{t+1}(\cdot), Z_{t+1}(\cdot), \psi)|\mathcal{A}_{t-k}^{t+k}\big]\big\|_{L_2}$$

$$\le \big\|M_t(x_{t+1}(\cdot), Z_{t+1}(\cdot), \psi) - M_t\big(\mathbb{E}\big[\big(x_{t+1}(\cdot), Z_{t+1}(\cdot)\big)|\mathcal{A}_{t-k}^{t+k}\big], \Psi\big)\big\|_{L_2}$$

$$\le c_{4,t}\,\big\|\big(x_{t+1}(\cdot), Z_{t+1}(\cdot)\big) - \mathbb{E}\big[\big(x_{t+1}(\cdot), Z_{t+1}(\cdot)\big)|\mathcal{A}_{t-k}^{t+k}\big]\big\|_{L_2}.$$

Hence for any fixed $\psi \in \mathbb{H}$ with $\|\psi\|_{\mathbb{H}} \le b_{\mathbb{H}}$, $\big\{M_t\big(\mathbf{G}\big(t, \cdot, x_0, Z_0, \Upsilon(\psi)\big), \psi\big)\big\}_{t\in\mathbb{N}}$ is a sequence of functions which is $L_2(\Omega;\mathbb{H})$-NED on $\{D_t\}_{t\in\mathbb{Z}}$ of size $-1/2$. Now Lemma 5.3 implies that

$$\Big\{\Big(M_t(x_{t+1}(\cdot), Z_{t+1}(\cdot), \psi) - \mathbb{E}\big[M_t(x_{t+1}(\cdot), Z_{t+1}(\cdot), \psi)\big]\Big),\ \mathcal{A}_t\Big\} \tag{5.65}$$

with $\mathcal{A}_t = \sigma(D_k| -\infty < k \le t)$ is an $L_2(\Omega;\mathbb{H})$ mixingale of size $-1/2$ with magnitude indices $d_t = O(c_{4,t})$ and decay parameters $\delta_t = O(k^{-\frac{1}{2}})$. It follows from Lemma 5.2 together with Lemma 5.7 that for each $\alpha > 0$, there exists $N_1(\alpha,\omega)$ such that

$$\sup_{t<n\le \mathrm{n}(t,T)} \|S_1(t,n)\|_{\mathbb{H}} \le \alpha(1+T) \quad \mathbb{P}-\text{a.s.}$$

for all $t > N_1(\alpha,\omega)$.

Step 3. Assumption 5.15 (ii) implies that for arbitrary $\alpha > 0$, there exists $N_2(\alpha,\omega) > N_1(\alpha,\omega)$ so that

$$\big\|\mathbb{E}\big[M_t(x_{t+1}(\cdot), Z_{t+1}(\cdot), \psi)\big] - M_\star(\psi)\big\|_{\mathbb{H}} < \alpha$$

for all $t \ge N_2(\alpha,\omega)$ and hence

$$\sup_{t<n\le \mathrm{n}(t,T)} \Big\|S_2(t,n)\Big\|_{\mathbb{H}} \le \alpha T \quad \mathbb{P}-\text{a.s.}$$

for all $t \ge N_2(\alpha,\omega)$. As a consequence

$$\sup_{t<n\le \mathrm{n}(t,T)} \big\|S(t,n)\big\|_{\mathbb{H}} \le \alpha(2+T) \quad \mathbb{P}-\text{a.s.}$$

for all $t > N_2(\alpha,\omega)$. This establishes (5.62) and thus Assumption 5.8 (iv) is satisfied. □

Proof of Proposition 5.4. For simplicity of notation, write

$$\mathbb{E}_{t-k}^{t+k}[\,\cdot\,] = \mathbb{E}[\,\cdot\,|\mathcal{A}_{t-k}^{t+k}], \quad \text{where } \mathcal{A}_{k_1}^{k_2} = \sigma(D_s|\ k_1 \le s \le k_2),\ k_1 < k_2.$$

By Assumption 5.10 (ii) on the exogenous $\mathbb{R}^{d_\xi}$-valued noise process $\{\xi_t\}_{\mathbb{N}}$, there exist $\lambda > \frac{1}{2}$ and $a_\xi > 0$ such that

$$\kappa_\xi(k) := \sup_{t\in\mathbb{N}} \|\xi_t - \mathbb{E}_{t-k}^{t+k}[\xi_t]\|_{L_2} \le a_\xi k^{-\lambda}, \quad k \in \mathbb{N}. \tag{5.66}$$

By construction, $\{\mathbf{G}(t,\cdot,x_0,Z_0,\Psi)\}_{t\in\mathbb{N}}$ is a sequence of $\mathbb{R}^{d_{\mathbb{H}}}$-valued bounded functions on $(\Omega,\mathcal{F},\mathbb{P})$. Thus for each $k \in \mathbb{N}$,

$$\kappa_G(k) := \sup_{t\in\mathbb{N}} \big\|\mathbf{G}(t,\cdot,x_0,Z_0,\Psi) - \mathbb{E}_{t-k}^{t+k}[\mathbf{G}(t,\cdot,x_0,Z_0,\Psi)]\big\|_{L_2} < \infty.$$

In order to establish near-epoch dependency of $\{\mathbf{G}(t,\cdot,x_0,Z_0,\Psi)\}_{t\in\mathbb{N}}$ in the sense of Definition 5.4, we will show in two steps that $\kappa_G(k) = O(k^{-\lambda})$ with $\lambda > \frac{1}{2}$ as given in (5.66).

Step 1. Let $(x_0,Z_0) \in \mathbb{X}\times\Sigma$, $\omega \in \Omega$ be arbitrary, and $\Psi \in \mathbb{S}$ be a forecasting rule such that

$$\kappa_y(k) := \sup_{t\in\mathbb{N}} \left\| y_t^e - \mathbb{E}_{t-k}^{t+k}[y_t^e] \right\|_{L_2} \le a_y k^{-\lambda}, \quad k \in \mathbb{N},$$

where

$$\begin{cases} x_t &= G\big(\xi_t, x_{t-1}, \Psi(x_{t-1}, Z_{t-1})\big) \\ y_{t-1}^e &= \Psi(x_{t-1}, Z_{t-1}) \end{cases} . \tag{5.67}$$

We show that the sequence of random variables $\{x_t\}_{t\in\mathbb{N}}$ generated by (5.67) is $L_2(\Omega;\mathbb{R}^{d_x})$-NED on $\{D_t\}_{t\in\mathbb{Z}}$ of size $-\frac{1}{2}$ so that

$$\kappa_x(k) := \sup_{t\in\mathbb{N}} \left\| x_t - \mathbb{E}_{t-k}^{t+k}[x_t] \right\|_{L_2} \le a_x k^{-\lambda}, \quad k \in \mathbb{N} \tag{5.68}$$

for some suitable constant $a_x > 0$. The minimum-squared-error property of the conditional expectations operator and the triangle equality yield

$$\begin{aligned} &\left\| x_t - \mathbb{E}_{t-k}^{t+k}[x_t] \right\|_{L_2} \\ &\le \left\| G(\xi_t, x_{t-1}, y_{t-1}^e) - G\big(\mathbb{E}_{t-k}^{t+k}[\xi_t], \mathbb{E}_{t-k}^{t+k}[x_{t-1}], \mathbb{E}_{t-k}^{t+k}[y_{t-1}^e]\big) \right\|_{L_2} \\ &\le c_1 \left\| \xi_t - \mathbb{E}_{t-k}^{t+k}[\xi_t] \right\|_{L_2} + c_2 \left\| x_{t-1} - \mathbb{E}_{t-k}^{t+k}[x_{t-1}] \right\|_{L_2} \\ &\quad + c_3 \left\| y_{t-1}^e - \mathbb{E}_{t-k}^{t+k}[y_{t-1}^e] \right\|_{L_2} \\ &\le c_1 \left\| \xi_t - \mathbb{E}_{t-k}^{t+k}[\xi_t] \right\|_{L_2} + c_2 \left\| x_{t-1} - \mathbb{E}_{t-k}^{t+k-2i}[x_{t-1}] \right\|_{L_2} \\ &\quad + c_3 \left\| y_{t-1}^e - \mathbb{E}_{t-k}^{t+k-2i}[y_{t-1}^e] \right\|_{L_2} \end{aligned}$$

Taking suprema over t, this implies for each $k \in \mathbb{N}$,

$$\kappa_x(k) \le c_2 \kappa_x(k-1) + c_3 \kappa_y(k-1) + c_1 \kappa_\xi(k).$$

We use induction over k to establish (5.68). Since $c_2 < 1$, there exist $\delta > 0$ with $c_2(1+\delta) < 1$ and k_0 such that

$$\left(\tfrac{k}{k-r}\right)^\lambda < 1 + \delta \quad \text{for all} \quad k > k_0.$$

Set

$$a_x = \max\left\{ \kappa_x(k_0) k_0^\lambda,\ \tfrac{c_3(1+\delta)a_y + c_1 a_\xi}{1 - c_2(1+\delta)} \right\} \tag{5.69}$$

and observe that

$$\kappa_x(k_0) \le a_x k_0^{-\lambda}.$$

Suppose now that for some $k > k_0$,

$$\kappa_x(k) \le a_x\, k^{-\lambda}.$$

Then

$$\kappa_x(k+1) \le \Big[c_2 a_x \left(\tfrac{k+1}{k}\right)^\lambda + c_3 a_y \left(\tfrac{k+1}{k}\right)^\lambda + c_1 a_\xi\Big](k+1)^{-\lambda}$$
$$\le [c_2(1+\delta)a_x + c_3(1+\delta)a_y + c_1 a_\xi](k+1)^{-\lambda} \le a_x(k+1)^{-\lambda}.$$

Hence $\kappa_x(k) = O(k^{-\lambda})$ for all $k \in \mathbb{N}$.

Step 2. Since

$$(x_t, Z_t) = (x_t, y^e_{t-1}, x_{t-1}, \dots, y^e_{t-r}, x_{t-r}),$$

we have

$$\big\|(x_t, Z_t) - \mathbb{E}^{t+k}_{t-k}\big[(x_t, Z_t)\big]\big\|_{L_2}$$
$$\le \sum_{i=0}^{r} \big\|x_{t-i} - \mathbb{E}^{t+k}_{t-k}[x_{t-i}]\big\|_{L_2} + \sum_{i=1}^{r} \big\|y^e_{t-i} - \mathbb{E}^{t+k}_{t-k}[y^e_{t-i}]\big\|_{L_2}$$
$$\le \sum_{i=0}^{r} \big\|x_{t-i} - \mathbb{E}^{t+k-2i}_{t-k}[x_{t-i}]\big\|_{L_2} + \sum_{i=1}^{r} \big\|y^e_{t-i} - \mathbb{E}^{t+k-2i}_{t-k}[y^e_{t-i}]\big\|_{L_2}.$$

Taking the supremum over t, this implies

$$\kappa_G(k) \le (a_x + a_y) \sum_{i=0}^{n} (k-i)^{-\lambda} = (a_x + a_y)\Big[\sum_{i=0}^{r} \left(\tfrac{k}{k-i}\right)^\lambda\Big] k^{-\lambda} = O(k^{-\lambda}).$$

This completes the proof of Proposition 5.4. □

Proof of Lemma 5.5. *Step 1.* We first prove the following lemma which is variant of Proposition 4.4 in Kuan & White (1994).

Lemma 5.8. *Assume that the hypotheses of Assumption 5.10 are satisfied. Let $H : \Xi \times \mathbb{Y}^n \to \mathbb{Y}$, $\mathbb{Y} \subset \mathbb{R}^{d_y}$ be a Lipschitz continuous map such that*

$$\|H(\xi, \zeta_1, \dots, \zeta_n) - H(\tilde\xi, \tilde\zeta_1, \dots, \tilde\zeta_n)\| \le c_H^{(1)} \sum_{i=1}^{n} \|\zeta_i - \tilde\zeta_i\| + c_H^{(2)} \|\xi - \tilde\xi\|$$

for all $\xi, \tilde\xi \in \Xi$, $\zeta_i, \tilde\zeta_i \in \mathbb{Y}$, $i = 1, \dots, n$. Suppose $n c_H^{(1)} < 1$. Then the sequence of random variables $\{\zeta_t\}_{t \in \mathbb{N}}$ generated by

$$\zeta_t = H(\xi_t, \zeta_{t-1}, \dots, \zeta_{t-n}), \quad t \in \mathbb{N}$$

is $L_2(\Omega; \mathbb{R}^{d_y})$-NED on $\{D_t\}_{t \in \mathbb{Z}}$ of size $-\frac{1}{2}$. In particular, for each $\delta > 0$ there exists k_0 such that

$$\kappa^n_\zeta(k) := \sup_{t \in \mathbb{N}} \big\|(\zeta_{t-1}, \dots, \zeta_{t-n}) - \mathbb{E}^{t+k}_{t-k}\big[(\zeta_{t-1}, \dots, \zeta_{t-n})\big]\big\|_{L_2} \le a_y k^{-\lambda} \quad (5.70)$$

for all $k > k_0$, where

$$a_y := \max\Big\{\kappa_\zeta^n(k_0)k_0^\lambda, \ldots, \kappa_\zeta^n(k_0-n+1)(k_0-n+1)^\lambda, \\ \tfrac{nc_H^{(2)}(1+\delta)}{1-nc_H^{(1)}(1+\delta)}a_\xi\Big\} \tag{5.71}$$

with λ *and* a_ξ *as given in (5.66).*

Proof of Lemma 5.8. The minimum-squared-error property of the conditional expectations operator and the triangle equality yield

$$\begin{aligned}
&\big\|\zeta_t - \mathbb{E}_{t-k}^{t+k}[\zeta_t]\big\|_{L_2} \\
&\le \big\|H(\xi_t,\zeta_{t-1},\ldots,\zeta_{t-n}) - H\big(\mathbb{E}_{t-k}^{t+k}[\xi_t],\mathbb{E}_{t-k}^{t+k}[\zeta_{t-1}],\ldots,\mathbb{E}_{t-k}^{t+k}[\zeta_{t-n}]\big)\big\|_{L_2} \\
&\le c_H^{(1)}\sum_{i=1}^n \big\|\zeta_{t-i} - \mathbb{E}_{t-k}^{t+k}[\zeta_{t-i}]\big\|_{L_2} + c_H^{(2)}\big\|\xi_t - \mathbb{E}_{t-k}^{t+k}[\xi_t]\big\|_{L_2} \\
&\le c_H^{(1)}\sum_{i=1}^n \big\|\zeta_{t-i} - \mathbb{E}_{t-k}^{t+k-2i}[\zeta_{t-i}]\big\|_{L_2} + c_H^{(2)}\big\|\xi_t - \mathbb{E}_{t-k}^{t+k}[\xi_t]\big\|_{L_2}.
\end{aligned}$$

Taking suprema over t, this implies for each $k \in \mathbb{N}$,

$$\kappa_\zeta(k) := \sup_{t\in\mathbb{N}} \|\zeta_t - \mathbb{E}_{t-k}^{t+k}[\zeta_t]\|_{L_2} \;\le\; c_H^{(1)}\sum_{i=1}^n \kappa_\zeta(k-i) \;+\; c_H^{(2)}\kappa_\xi(k).$$

An induction argument over k is used next to establish (5.70). Choose $\delta > 0$ such that $(1+\delta)n\,c_H^{(1)} < 1$ and k_0 such that

$$\Big(\tfrac{k}{k-n}\Big)^\lambda < 1+\delta \quad \text{for all} \quad k > k_0.$$

Set

$$b = \max\Big\{\kappa_\zeta(k_0)k_0^\lambda, \ldots, \kappa_\zeta(k_0-n+1)(k_0-n+1)^\lambda, \tfrac{c_H^{(2)}}{1-nc_H^{(1)}(1+\delta)}a_\xi\Big\}$$

and observe that

$$\kappa_\zeta(k_0-i) \le b(k_0-i)^{-\lambda}, \quad i = 0,\ldots,n-1.$$

Suppose now that for some $k > k_0+1$,

$$\kappa_\zeta(k-i) \le b\,(k-i)^{-\lambda}, \quad i = 1,\ldots,n.$$

Then

$$\begin{aligned}
\kappa_\zeta(k) &\le c_H^{(1)}\sum_{i=1}^n \kappa_\zeta(k_0-i) \;+\; c_H^{(2)}\kappa_\xi(k) \;\le\; \Big[c_H^{(1)} b\sum_{i=1}^n \Big(\tfrac{k}{k-i}\Big)^\lambda + c_H^{(2)}a_\xi\Big]k^{-\lambda} \\
&\le [b\,nc_H^{(1)}(1+\delta) + c_H^{(2)}a_\xi]k^{-\lambda} \;\le\; bk^{-\lambda}.
\end{aligned}$$

This establishes near-epoch dependency of $\{\zeta_t\}_{t\in\mathbb{N}}$.

As regards the second assertion, we have

$$\big\|(\zeta_{t-1},\ldots,\zeta_{t-n}) - \mathbb{E}_{t-k}^{t+k}\big[(\zeta_{t-1},\ldots,\zeta_{t-n})\big]\big\|_{L_2} \le \sum_{i=1}^{n} \big\|\zeta_{t-i} - \mathbb{E}_{t-k}^{t+k}[\zeta_{t-i}]\big\|_{L_2}$$

$$\le \sum_{i=1}^{n} \big\|\zeta_{t-i} - \mathbb{E}_{t-k}^{t+k-2i}[\zeta_{t-i}]\big\|_{L_2}$$

Taking the supremum over t, this implies

$$\kappa_\zeta^n(k) \le b\sum_{i=1}^{n}(k-i)^{-\lambda} \;=\; b\Big[\sum_{i=1}^{n}\Big(\tfrac{k}{k-i}\Big)^{\lambda}\Big]\,k^{-\lambda} \;=\; (1+\delta)nb\,k^{-\lambda}.$$

Setting $a_y := (1+\delta)nb$, this completes the proof of Lemma 5.8.

Step 2. In view of Remark 5.2, the forecasting rule for the m-th forecast is of the form

$$y^e_{t,t+m} = \Psi^{(m)}(x_t, y^e_{t-1,t+1},\ldots,y^e_{t-1,t-1+m}).$$

Assume for a moment that $\{x_t\}_{t\in\mathbb{N}}$ is near-epoch dependent so that (5.68) holds. We may then apply Lemma 5.8 with $n=m-1$, $c_H^{(1)} = c_H^{(2)} = c_\Upsilon$, and $\{\xi_t\}_{t\in\mathbb{N}}$ being replaced by $\{x_t\}_{t\in\mathbb{N}}$ to conclude that

$$\kappa_y(k) = \sup_{t\in\mathbb{N}}\big\|y^e_t - \mathbb{E}_{t-k}^{t+k}[y^e_t]\big\|_{L_2} \le a_y k^{-\lambda}, \quad k\in\mathbb{N},$$

where

$$a_y = \max\Big\{\kappa_y(k_0)k_0^\lambda,\ldots,\kappa_y(k_0-n+2)(k_0-n+2)^\lambda,\ \tfrac{(m-1)c_\Upsilon(1+\delta)}{1-(m-1)c_\Upsilon(1+\delta)}a_x\Big\} \tag{5.72}$$

with λ and a_x as given in (5.69). However, since $\{x_t\}_{t\in\mathbb{N}}$ and $\{y^e_t\}_{t\in\mathbb{N}}$ are interdependent processes generated by (5.67), a_x and a_y cannot be chosen independently. In fact, conditions (5.69) and (5.72) are two equations for two coefficients a_x and a_y which must hold simultaneously. It is straightforward to find a_x and a_y which fulfill conditions (5.69) and (5.72) if

$$\tfrac{(m-1)c_\Upsilon(1+\delta)}{1-(m-1)c_\Upsilon(1+\delta)}\,\tfrac{c_2(1+\delta)}{1-c_2(1+\delta)} < 1$$

or, equivalently, if $(m-1)c_\Upsilon + c_2 < \frac{1}{1+\delta}$. Since δ may be chosen arbitrarily small, the latter condition is satisfied. This establishes near-epoch dependency of $\{\mathbf{G}(t,\cdot,x_0,Z_0,\Psi)\}_{t\in\mathbb{N}}$ in the sense of Definition 5.4. □

6

Stochastic Exchange Economies

As a first application of the results obtained in Chapters 2 and 5, we discuss a standard version of the overlapping generations model with pure exchange which can be found in any macroeconomic textbook, e.g., see Azariadis (1993). The problem of learning the rational expectations equilibria when endowments are random has been treated by many authors, among which are Chen & White (1998) and Kelly & Shorish (2000). We will strengthen these results by showing that perfect foresight is possible despite of the fact that endowments are random. Generalizing ideas from Wenzelburger (2002) for the deterministic version of the model, we will introduce a learning scheme which generates strongly consistent nonparametric estimates of a so-called perfect forecasting rule which by definition generates perfect foresight orbits. We will show that the required structural knowledge is comprised in the *error function* which arises naturally when distinguishing between the basic market mechanism of the economy and a *forecasting rule* according to which agents form expectations. The zero contour set of an error function determines the so-called perfect forecasting rules which generate perfect foresight equilibria. Exploiting the geometry of the error function similar to Wenzelburger (2002), we develop an algorithm which generates an arbitrarily precise approximation of a perfect forecasting rule using historical data only. This algorithm converges globally for all initial conditions under fairly general conditions.

6.1 A Model of Pure Exchange

Consider a standard version of the overlapping generations model with one non-storable commodity per period and fiat money as the only store of value between periods. There will be neither growth of the population nor production. Given the usual assumption of price taking behavior of all generations, young agents need to transfer purchasing power from the first to the second period of their lives. To abstract from heterogenous beliefs, let p_t^e denote the common forecast for the future price of the consumption good in period $t+1$

on which all members of the young generation in t base their decisions. We assume that a forecasting agency is in charge of issuing the forecasts and that this agency knows the basic market mechanism without information concerning the households' savings behavior.

Initial endowments are $w_t^{(1)}$ for young and $w_t^{(2)}$ for old consumers and are measured in units of the consumption good. They are assumed to be generated by a stationary stochastic process $\{w_t\}_{t\in\mathbb{N}}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values $w_t = (w_t^{(1)}, w_t^{(2)})$ in a compact cube $\mathbb{W} \times \mathbb{W}$ with $\mathbb{W} = [\underline{w}, \overline{w}]$. At the beginning of each period t, a young agent observes her current endowment $w_t^{(1)}$. The only uncertainty is over her future endowment $w_{t+1}^{(2)}$ when old and the price level p_{t+1} prevailing in period $t+1$. Consumers are endowed with a von-Neumann-Morgenstern utility function $u : \mathbb{R}_+^2 \to \mathbb{R}$ over consumption $c^{(1)}$ and $c^{(2)}$ when young and old, respectively. u is assumed to be bounded, strictly monotonically increasing, strictly concave and twice differentiable. The budget constraints are

$$c^{(1)} = w_t^{(1)} - s, \quad c^{(2)} = w^{(2)} + \frac{p}{q} s,$$

where $q \in \mathbb{R}_+$ denotes the possible future price of the consumption good and $w^{(2)}$ the future endowment. Both are uncertain when deciding on how much to save. Let $\mu(p_t^e, \cdot) \in \mathrm{Prob}(\mathbb{R}_+ \times \mathbb{W})$ denote the subjective joint distribution of p_{t+1} and $w_{t+1}^{(2)}$, which is parameterized in subjective mean values p_t^e. Given the initial endowment of goods $w_t^{(1)}$ and the forecast p_t^e for the unknown price level p_{t+1}, the solution of the consumer's maximization problem is given by a savings function $S : \mathbb{R}_+^3 \to \mathbb{R}_+$, defined by

$$S(w_t^{(1)}, p, p_t^e) := \underset{0 \le s \le w_t^{(1)}}{\operatorname{argmax}} \int_{\mathbb{R}_+ \times \mathbb{W}} u\left(w_t^{(1)} - s, w^{(2)} + \tfrac{p}{q} s\right) \mu(p_t^e, dq, dw^{(2)}), \tag{6.1}$$

such that $S(w_t^{(1)}, p, p_t^e)$ describes real savings of young consumers in period t at the goods price p.

Market clearing on the goods market in any period t requires that the real savings of young consumers, which defines the amount of consumption goods supplied to the market, has to be equal to the demand of the old generation which is equal to the real purchasing power of their money received in exchange for consumption goods in the previous period. If $\bar{m}$ denotes the total stock of money in the economy, a market clearing price in period t is defined to be that price p_t for which the excess demand of period t is equal to zero, i.e.,

$$\Phi_{\mathrm{ex}}(w_t^{(1)}, p_t, p_t^e) := \frac{\bar{m}}{p_t} - S\left(w_t^{(1)}, p_t, p_t^e\right) = 0. \tag{6.2}$$

Assume now that there exists an open subset $\mathbb{Y} \subset \mathbb{R}_+$ and a map $G : \mathbb{W} \times \mathbb{Y} \to \mathbb{R}_+$ such that

$$\Phi_{\rm ex}(w^{(1)}, G(w^{(1)}, p^e), p^e) = 0 \quad \text{for all} \quad (w^{(1)}, p^e) \in \mathbb{W} \times \mathbb{Y}.$$

If $\Phi_{\rm ex}$ is sufficiently regular and satisfies the hypotheses of the implicit function theorem, a set $\mathbb{Y}$ together with a map G will exists. The market-clearing price p_t of period t is then uniquely determined by the *temporary equilibrium map*

$$G : \mathbb{W} \times \mathbb{Y} \to \mathbb{R}_+, \quad p_t = G(w_t^{(1)}, p_t^e). \tag{6.3}$$

The map G given in (6.3) includes a two-step-ahead forecast and for this reason defines an economic law with an expectational lead in the sense of Chapter 2. Since apart from the stochastic endowments only expected prices enter G as an argument, the economic law (6.3) has the structure of a Cobweb model. From a sequential view point, the forecast p_t^e for the price p_{t+1} has to be determined before the actual trading takes place in period t and thus prior to the realization of p_t. This feature is illustrated in Figure 6.1.

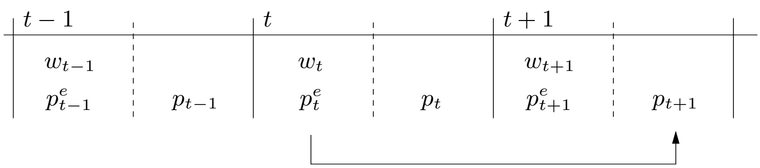

Fig. 6.1. Time-line of price formation.

It follows from (6.3) that the market-clearing price p_t is a deterministic function of the endowment $w_t^{(1)}$ and the price forecast p_t^e. For this reason, we will assume that consumers probability distribution is of the form $\mu = \delta_{p_t^e} \times \nu$, where $\delta_{p_t^e}$ is Dirac measure and ν is the probability distribution of $w_{t+1}^{(2)}$. As a consequence, the consumer's savings function may assumed to be of the functional form

$$S\left(w_t^{(1)}, \tfrac{p_t^e}{p}\right) := \underset{0 \le s \le w_t^{(1)}}{\operatorname{argmax}} \int_{\mathbb{W}} u\left(w_t^{(1)} - s, w^{(2)} + \tfrac{p}{p_t^e} s\right) \nu(\mathrm{d}w^{(2)}). \tag{6.4}$$

Remark 6.1. *With the notation of Chapter 2 we would have written* $p_{t-1}^e \equiv p_{t-1,t+1}^e$ *instead of* p_t^e. *For the exchange model, however, the latter notation is economically more appealing.*

6.2 The Existence of Perfect Forecasting Rules

The discussion of the sequential trading structure in Section 6.1 revealed that perfect foresight may be possible despite the fact that endowments are generated by an exogenous stochastic process. Due to the particular information structure of the exchange economy, the mean error function associated with the economic law (6.3) becomes a deterministic error function of the form introduced in Böhm & Wenzelburger (2004). In order to address the question whether forecasting rules exists that generate perfect foresight along all orbits (trajectories) of the economy, we therefore consider this deterministic error function. Analogously to the notion of an unbiased forecasting rule, a forecasting rule is called perfect if it generates perfect foresight orbits. Perfect forecasting rules are thus special cases of unbiased forecasting rules. (Compare with Remark 4.3, Chapter 2.)

For an arbitrary period, let p^e_{old} and p^e_{new} denote the forecasts of the old and the young generation, respectively. The forecast error of the old generation is then given by the *error function* $e_G : \mathbb{W} \times \mathbb{Y}^2 \to \mathbb{R}$, defined by

$$e_G(w^{(1)}, p^e_{\text{old}}, p^e_{\text{new}}) := G(w^{(1)}, p^e_{\text{new}}) - p^e_{\text{old}}. \tag{6.5}$$

As in Chapter 2, the function e_F describes all possible forecast errors, independently of what forecasting rule or learning scheme has been used to obtain the forecasts. The zero-contour set of e_F describes all triplets $(w^{(1)}, p^e_{\text{old}}, p^e_{\text{new}})$ for which perfect foresight obtains. A perfect forecasting rule must therefore be a function $p^e_{\text{new}} = \Psi_\star(w^{(1)}, p^e_{\text{old}})$ defined on a suitable subset $\mathbb{U} \subset \mathbb{W} \times \mathbb{Y}$ such that

$$G\big(w^{(1)}, \Psi_\star(w^{(1)}, p^e_{\text{old}})\big) = p^e_{\text{old}}. \tag{6.6}$$

The economic implication for a perfect forecasting rule is that the new forecast p^e_{new} has to be chosen such that the old forecast p^e_{old} becomes correct.

The following proposition shows that for the OLG exchange model forecasting rules which are perfect locally on some subset $\mathbb{U}$ exist under relatively mild conditions, whereas globally perfect forecasting rules exist only under very restrictive assumptions. Let

$$\mathrm{E}_S(w^{(1)}, \theta) := \frac{\mathrm{D}_2 S(w^{(1)}, \theta)\theta}{S(w^{(1)}, \theta)}$$

denote the elasticity of the savings function with respect to expected inflation factors θ and $S^{-1}(w^{(1)}, \cdot)$, $w^{(1)} \in \mathbb{W}$ denote the inverse of $S(w^{(1)}, \cdot)$, whenever this inverse is well defined.

Proposition 6.1. *Let S be continuously differentiable, and assume that for each $w^{(1)} \in \mathbb{W}$,*

$$(i) \quad \bar{p}_w := \frac{\bar{m}}{S(w^{(1)}, 1)} \in \mathbb{Y} \quad \textit{and} \quad (ii) \quad \mathrm{D}_2 S(w^{(1)}, 1) \neq 0.$$

Then there exists an open subset $\mathbb{U} \subset \mathbb{W} \times \mathbb{Y}$ and a forecasting rule $\Psi_\star$, given by

$$\Psi_\star(w^{(1)}, p^e) = p^e\, S^{-1}\left(w^{(1)}, \frac{\bar{m}}{p^e}\right), \quad (w^{(1)}, p^e) \in \mathbb{U} \tag{6.7}$$

which is locally perfect on $\mathbb{U}$.

Proof. Let $w^{(1)} \in \mathbb{W}$ be arbitrary but fixed. Then

$$\Phi_{\text{ex}}(w^{(1)}, \bar{p}_w, \bar{p}_w) = \frac{\bar{m}}{\bar{p}_w} - S(w^{(1)}, 1) = 0,$$

where Φ_{ex} is the excess demand function defined in (6.2). Since $\mathrm{D}_2 S(w^{(1)}, 1) \neq 0$, $\mathrm{D}_3 \Phi_{\text{ex}}(w^{(1)}, \bar{p}_w, \bar{p}_w) \neq 0$. Invoking the Implicit Function Theorem, this yields the existence of an open set $U(w^{(1)}) \subset \mathbb{Y}$ such that

$$\Phi_{\text{ex}}\big(w^{(1)}, p^e, \Psi_\star(w^{(1)}, p^e)\big) = 0, \quad p^e \in U(w^{(1)}),$$

where $\Psi_\star$ is given in (6.7). The definition of the economic law G defined in (6.3), then shows

$$G\big(w^{(1)}, \Psi_\star(w^{(1)}, p^e)\big) = p^e, \quad p^e \in U(w^{(1)}).$$

Since $w^{(1)}$ was arbitrary, the proof is completed by setting

$$\mathbb{U} = \{(w^{(1)}, p^e) \in \mathbb{W} \times \mathbb{Y} |\ p^e \in U(w^{(1)})\}.$$

□

Corollary 6.1. *Let youthful and old-age consumption be normal goods. Then* $\mathrm{D}_2 \Psi_\star \gtreqless 0$ *on* $\mathbb{U}$ *if and only if* $\mathrm{D}_2 S \lesseqgtr 0$ *on* $\mathbb{U}$.

Proof. We have

$$\mathrm{D}_2 \Psi_\star(w^{(1)}, p^e) = S^{-1}\left(w^{(1)}, \tfrac{\bar{m}}{p^e}\right)\left[1 - \frac{1}{\mathrm{E}_S\big(w^{(1)}, S^{-1}(w^{(1)}, \frac{\bar{m}}{p^e})\big)}\right]$$

as long as $\mathrm{D}_2 S \neq 0$. The normality assumption together with the Slutsky conditions imply that $\mathrm{E}_S(w^{(1)}, \theta) < 1$. This completes the proof.

□

Proposition 6.1 assures that forecasting rules which are locally perfect will generically exist. The resulting dynamics under perfect foresight will be investigated in the next section.

6.3 Perfect Foresight Dynamics

Proposition 6.1 states that perfect forecasting rules depend exclusively on previous forecasts and endowments but not on observed prices. Geometrically, the graph of a locally perfect forecasting rule $\Psi_\star$ is contained in the zero contour of the error function. Under perfect foresight in period $t \in \mathbb{N}$, one has

$$\begin{aligned} p_t &= G\big(w_t^{(1)}, p_t^e\big) = p_{t-1}^e, \\ p_t^e &= \Psi_\star(w_t^{(1)}, p_{t-1}^e), \end{aligned} \tag{6.8}$$

where $\Psi_\star$ is the perfect forecasting rule defined in (6.7) and $w_t^{(1)} \in [\underline{w}, \overline{w}]$ is random. Eq. (6.8) reveals that the dynamics under perfect foresight, if it exist, will essentially be governed by a dynamics on forecasts alone.

Since savings must never exceed initial endowments and $w_t^{(1)}$ is random, it follows from (6.2) and (6.7) that the map $\Psi_\star$ is not defined for $p_t < \bar{m}/\underline{w}$. Hence, each $\mathbb{U}$ must be a proper subset of $\mathbb{W} \times \mathbb{R}_+$, even if $\mathbb{Y} = \mathbb{R}_+$. For this reason it is impossible to define a perfect foresight dynamics globally on $\mathbb{R}_+$. In order to obtain a well-defined perfect foresight dynamics, we seek a subset $\mathbb{V} \subset \mathbb{R}_+$ with $\mathbb{W} \times \mathbb{V} \subset \mathbb{U}$ which is forward-invariant under $\Psi_\star$ in the sense that $\Psi_\star : \mathbb{W} \times \mathbb{V} \longrightarrow \mathbb{V}$. Then each orbit of (6.8) that starts in some $p_0^e \in \mathbb{V}$ will stay in $\mathbb{V}$ for all times t.

Proposition 6.2. *Let youthful and old-age consumption be normal goods, let all hypotheses of Proposition 6.1 be satisfied, and assume that* $\mathrm{D}_2 S > 0$. *Then the compact interval* $\mathbb{V} = [\underline{p}, \overline{p}] \subset \mathbb{R}_{++}$ *is forward-invariant under* $\Psi_\star$, *if*

$$\overline{p}\, S\left(\overline{w}, \frac{\underline{p}}{\overline{p}}\right) \le \bar{m} \le \underline{p}\, S\left(\underline{w}, \frac{\overline{p}}{\underline{p}}\right).$$

Proof. We know from Corollary 6.1 that $\Psi_\star$ is monotonically decreasing in price forecasts, i.e. $\mathrm{D}_2\Psi_\star < 0$. Moreover, since both goods are normal we have $\mathrm{D}_1 S > 0$. This implies $\mathrm{D}_1\Psi_\star > 0$. Therefore, a compact interval $[\underline{p}, \overline{p}] \subset \mathbb{R}_{++}$ is forward-invariant under $\Psi_\star$, if

$$\Psi_\star(\underline{w}, \overline{p}) \ge \underline{p} \quad \text{and} \quad \Psi_\star(\overline{w}, \underline{p}) \le \overline{p}$$

By the monotonicity of S and the definition of $\Psi_\star$, these conditions are equivalent to

$$S\left(\underline{w}, \frac{\overline{p}}{\underline{p}}\right) \ge \frac{\bar{m}}{\underline{p}} \quad \text{and} \quad S\left(\overline{w}, \frac{\underline{p}}{\overline{p}}\right) \le \frac{\bar{m}}{\overline{p}}.$$

□

If the hypotheses of Proposition 6.2 hold such that a forward-invariant set $\mathbb{V} = [\underline{p}, \overline{p}]$ exists, then one has perfect foresight along any orbit starting in $\mathbb{V}$, that is, $p_t = p_{t-1}^e$ for all $t \in \mathbb{N}$, provided that $p_0^e \in \mathbb{V}$. It follows from (6.8) that the resulting dynamics can be reduced to a one-dimensional one.

In view of Assumption 4.1, Chapter 4 we assume that $\{w_t\}_{t\in\mathbb{Z}}$ has a representation as an ergodic metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$ with

$$w_t(\omega) = w(\vartheta^t\omega), \quad \omega \in \Omega,\ t \in \mathbb{Z}. \tag{6.9}$$

Using the representation (6.9), we obtain a time-one map $\phi(1,\cdot) : \Omega \times \mathbb{V} \to \mathbb{V}$, given by

$$p_{t+1} = \phi\big(1, \vartheta^t\omega, p_t\big) := \Psi_\star(w^{(1)}(\vartheta^t\omega), p_t), \tag{6.10}$$

of a random dynamical system over $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$. The time-one map (6.10) describes the price dynamics under perfect foresight of the system which

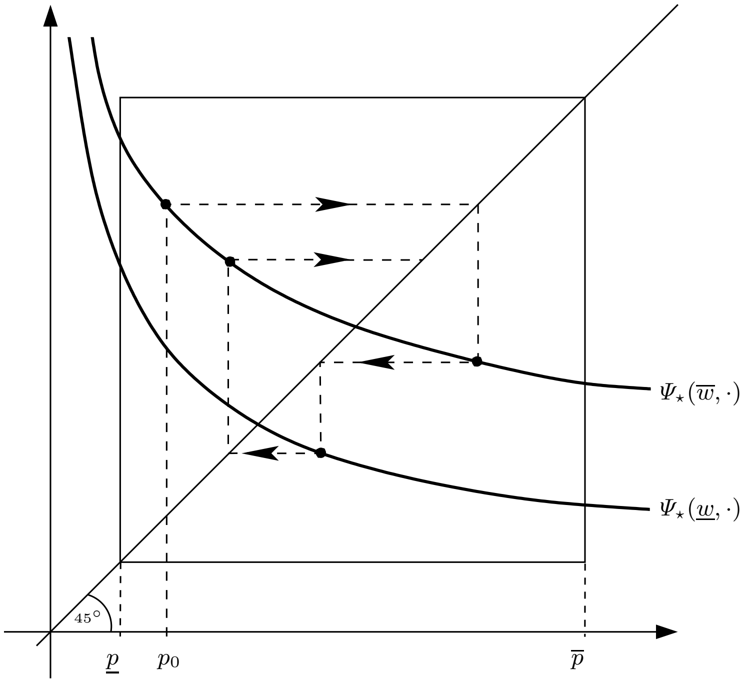

Fig. 6.2. Orbit with perfect foresight.

is well defined for all initial price levels $p_0 \in \mathbb{V} = [\underline{p}, \overline{p}]$. This situation is shown in Figure 6.2 which depicts a typical orbit of the system (6.10).

We conclude this section by showing that under certain conditions, the long-run behavior is determined by a stationary stochastic process induced by a random fixed point of the system (6.10).

Theorem 6.1. *Let the hypotheses of Proposition 6.1 be satisfied and assume that the savings function S is differentiable with respect to both arguments and* $\mathrm{D}_2 S > 0$. *Assume, in addition, that the following conditions hold:*

(i) There exits a compact set $\mathbb{V} = [\underline{p}, \overline{p}]$ which is forward-invariant under $\Psi_\star$ such that $\Psi_\star : \mathbb{W} \times \mathbb{V} \to \mathbb{V}$.
(ii) There exists a random variable η with $\mathbb{E}[\eta] < 0$, such that

$$\sup\Big\{\log\left|\mathrm{D}_2\Psi_\star\big(w^{(1)}(\omega), p\big)\right| \;\Big|\; p \in \mathbb{V}\Big\} \le \eta(\omega), \quad \omega \in \Omega.$$

Then there exists a unique nontrivial random fixed point $p^\star : \Omega \to \mathbb{R}_+$ which is asymptotically stable on $\mathbb{V}$ in the sense of Definition 4.6. Moreover, $p^\star$ is measurable with respect to the sub-σ-algebra of the past $\mathcal{F}^- := \sigma(\phi(t, \vartheta^{-T}\cdot, p) \mid 0 \le t \le T,\ p \in \mathbb{R}_+) \subset \mathcal{F}$, where ϕ denotes the random dynamical system generated by $\Psi_\star$.

The proof is an application of Theorem 4.3 in Section 4.4 of Chapter 4 and given in Section 6.5 below. It is a modification of the proof of Theorem 4.2 in Schenk-Hoppé & Schmalfuß (1998). Condition (ii) is an average contractivity condition. It is straightforward to see that (ii) is satisfied, if

(ii') $$|\mathrm{D}_2\Psi_\star(w^{(1)},p)| < 1 \quad \text{for all} \quad (w^{(1)},p) \in \mathbb{W} \times \mathbb{V}.$$

Invoking Corollary 6.1, Condition (ii') is satisfied at the deterministic fixed points

$$\bar{p}_w = \frac{\bar{m}}{S(w^{(1)},1)}, \quad w^{(1)} \in \mathbb{W}$$

of the maps $\Psi_\star(w^{(1)},\cdot)$, $w^{(1)} \in \mathbb{W}$ if $\mathrm{E}_S(w^{(1)},1) > \frac{1}{2}$. It thus provides a natural extension of the classical stability condition for deterministic pure exchange models, cf. Azariadis (1993).

6.4 Adaptive Learning of Perfect Forecasting Rules

Consider a forecasting agency which is in charge of forecasting the future evolution of the economy and hence has to decide on p^e. We assume here that *all* young consumers in period t share the belief in p^e_t issued by the forecasting agency prior to their savings decisions. This excludes heterogenous beliefs and strategic behavior of consumers. In order to be credible for households, we assume that the forecasting agency itself has no strategic interest other than issuing forecasts as precise as possible. The informational constraints faced by our forecasting agency can be described as follows.

Suppose that the agency in some arbitrary period τ has observed past prices $\{p_t\}_{t=0}^{\tau-1}$ and knows forecasts $\{p^e_t\}_{t=0}^{\tau-1}$ corresponding to these prices. Recall that p_τ is not available prior to the decision on p^e_τ. Let the agency be aware of the basic market mechanism of the economy, that is, the basic structure of the economic law given by (6.3) but not its specific parameterization. Hence, neither the preferences nor the savings behavior of young consumers are known to the agency. The forecasting agency is assumed to use the concept of the error function without its exact functional specification and for this reason is boundedly rational in the sense of Sargent (1993).

The goal of a forecasting agency now is to estimate a perfect forecasting rule $\Psi_\star$. Given the temporary equilibrium map (6.3), observe first that in an arbitrary period τ,

$$p_t = G(w^{(1)}_t, p^e_t)$$

for all $\{(w^{(1}_t, p_t, p^e_t)\}_{t=0}^{\tau-1}$, no matter what forecasting rule has been applied to generate the time series. If a locally perfect forecasting rule $\Psi_\star$ exists, it follows from (6.6) that the data points $\{(w^{(1}_t, p_t, p^e_t)\}_{t=0}^{\tau-1}$ must also satisfy

$$p^e_t = \Psi_\star(w^{(1)}_t, p_t) \tag{6.11}$$

in period τ, provided that $(w_t^{(1)}, p_t) \in \mathbb{U}$. As $w_t^{(1)}$, p_t, and p_t^e are observable quantities, the idea is to estimate $\Psi_\star$ nonparametrically from this data.

With the notation of Chapter 5, let $\mathbb{X} = \mathbb{W} \times \mathbb{Y}$ with $\Sigma = \mathbb{Y} \subset \mathbb{R}_+$ and set

$$x_t := (w_t^{(1)}, p_t) \in \mathbb{W} \times \mathbb{Y} \quad \text{and} \quad Z_t := p_t^e \in \mathbb{Y}.$$

Let $0 < c_\Upsilon < 1$ be some constant, $\mathbb{Y}' \subset \mathbb{Y}$ be a compact interval not containing zero, and $\pi_{\mathbb{Y}'} : \mathbb{R} \to \mathbb{Y}'$ be a projection onto $\mathbb{Y}'$. As in Section 5.5.1 of Chapter 5, denote by $C_B^1(\mathbb{X}; \mathbb{R})$ the space differentiable functions with bounded derivatives. Define the censor map as

$$\Upsilon : \begin{cases} C_B^1(\mathbb{X}; \mathbb{R}) & \longrightarrow C_B^1(\mathbb{X}; \mathbb{R}) \\ \Psi & \longmapsto \Upsilon(\Psi) \end{cases}, \tag{6.12}$$

where

$$\Upsilon(\Psi) = \begin{cases} \pi_{\mathbb{Y}'} \circ \Psi & \text{if } \|\mathrm{D}_2\Psi\|_\infty < c_\Upsilon \\ \pi_{\mathbb{Y}'} \circ \left(\frac{c_\Upsilon}{\|\mathrm{D}_2\Psi\|_\infty} \Psi \right) & \text{if } \|\mathrm{D}_2\Psi\|_\infty \geq c_\Upsilon \end{cases}$$

and

$$\|\mathrm{D}_2\Psi\|_\infty := \sup\Big\{ |\mathrm{D}_2\Psi(w^{(1)}, p)| \;\Big|\; (w^{(1)}, p) \in \mathbb{W} \times \mathbb{Y} \Big\}.$$

The role of the censor map Υ is first to confine the system to a compact set $\mathbb{Y}'$ at all stages and second to generate a near-epoch dependent series of forecasts as required by the convergence result of Theorem 5.3 in Chapter 5. The latter property follows from Lemma 5.8 of Chapter 5, noticing that for each $\Psi \in C_B^1(\mathbb{X}; \mathbb{R})$ and each $w^{(1)} \in \mathbb{W}$, the map $\Upsilon(\Psi)(w^{(1)}, \cdot)$ is a contraction.

Since $\mathbb{X} \subset \mathbb{R}_+^2$, we choose the Sobolev space $\mathbb{H} = \mathbb{W}^{2,2}(\mathbb{X}; \mathbb{R})$ as introduced in Section 5.5.1 of Chapter 5 to be the underlying Hilbert space of our learning scheme. Recall that by Theorem 5.4, $\mathbb{H}$ can then be continuously imbedded into the space of continuous and bounded functions $C_B(\mathbb{X}, \mathbb{R})$. For arbitrary initial conditions $(w_0^{(1)}, p_0, p_0^e) \in \mathbb{X} \times \mathbb{Y}$ and $\hat{\Psi}_0 \in C_B^1(\mathbb{X}; \mathbb{R})$, our learning scheme is defined by

$$\begin{cases} p_t = G\big(w_t^{(1)}, p_t^e\big) \\ p_t^e = \Psi_t(w_t^{(1)}, p_{t-1}^e) \\ \Psi_t = \Upsilon(\hat{\Psi}_t) \\ \hat{\Psi}_t = \hat{\Psi}_{t-1} + \frac{1}{t-1} M_{t-1}(w_{t-1}^{(1)}, p_{t-1}, p_{t-1}^e, \hat{\Psi}_{t-1}) \end{cases} \tag{6.13}$$

with $t \in \mathbb{N}$. Here the updating functions $\{M_t\}_{t \in \mathbb{N}}$ are assumed to satisfy Assumption 5.15 of Chapter 5. For simplicity we suppressed the truncation device introduced in Section 5.1, knowing that under suitable conditions no truncations of the estimates will occur after a sufficient amount of time has elapsed. The next theorem shows under which conditions the learning scheme (6.13) succeeds in finding the perfect forecasting rule $\Psi_\star$.

Theorem 6.2. *Consider an exchange economy as described above and assume that the following hypotheses are satisfied:*

(i) The stationary process $\{w_t^{(1)}\}_{t\in\mathbb{Z}}$ *is* $L_2(\Omega;\mathbb{R})$*-NED on* $\{D_t\}_{t\in\mathbb{Z}}$ *of size* $-\frac{1}{2}$*, where* $\{D_t\}_{t\in\mathbb{Z}}$ *is a sequence which is either uniformly mixing with* α_k *of size* -1 *or strongly mixing with* ϕ_k *of size* $-\frac{1}{2}$ *in the sense of Definition 5.3, Chapter 5.*
(ii) The economic law G *when restricted to* $\mathbb{Y}'$ *is Lipschitz continuous with respect to both variables.*
(iii) The conditions of Theorem 6.1 hold such that there exists a differentiable perfect forecasting rule $\Psi_\star \in \mathbb{H}$ *together with a forward-invariant set* $\mathbb{V} \subset \mathbb{Y}'$*. Moreover,* $\|\mathrm{D}_2\Psi_\star\|_\infty \le c_\Upsilon < 1$*.*
(iv) The sequence $\{M_t\}_{t\in\mathbb{N}}$ *of updating functions satisfies Assumptions 5.12-5.15 of Chapter 5.*

Let $\hat{\Psi}_0 \in C^1_B(\mathbb{X};\mathbb{R})$*. Then the sequence of forecasting rules* $\{\Psi_t\}_{t\in\mathbb{N}}$ *obtained from (6.13) converges to* $\Psi_\star$ $\mathbb{P}$*-a.s., that is,*

$$\lim_{t\to\infty} \|\Psi_t - \Psi_\star\|_{\mathbb{H}} = 0 \quad \mathbb{P}-a.s.$$

The proof of Theorem 6.2 is an application of Theorem 5.3, Chapter 5 and is found in Section 6.5 below.

Concluding Remarks

The fact that perfect foresight is possible in a random environment of an exchange economy seems to have been overlooked in the literature. This chapter demonstrates that forecasting rules which generate perfect foresight exist under standard assumptions. Using the methodology of Chapter 5, these so-called perfect forecasting rules can successfully be estimated from historical data by means of nonparametric estimations. The censoring of the approximated forecasting rules assures that the system is kept stable at all stages of the estimation procedure and that the resulting process satisfies the necessary stochastic properties. It should be noted that the method required neither the existence of a perfect forecasting rule nor its contraction property.

6.5 Mathematical Appendix

Proof of Theorem 6.1. Notice first that by construction $\mathbb{E}[w^{(1)}(\cdot)] < \infty$. We verify the assumptions of Theorem 4.3, given in Chapter 4.
Step 1. Let $\mathbb{V} = [\underline{p}, \overline{p}]$. Then $\mathbb{V}(\omega) \equiv \mathbb{V}, \omega \in \Omega$ is a random set and the set of all tempered random variables $\mathcal{V}$ associated with $\mathbb{V}$ is nonempty since it contains constant functions with values in $[\underline{p}, \overline{p}]$.

Step 2. Check of Assumption (i). Condition (i) implies that $\mathbb{V}$ is forward invariant under $\Psi_\star$. Since it is bounded by $\overline{p}$, the map $\omega \mapsto \Psi_\star\big(\xi(\vartheta^{-1}\omega), g(\vartheta^{-1}\omega)\big)$ is tempered, because it is bounded by a tempered (constant) map. Hence Assumption (i) is satisfied.
Step 3. Condition (ii) is a mere restatement of Assumption (ii) in Theorem 4.3.
Step 4. Check of Assumption (iii). Let $\{\phi\big(t, \vartheta^{-t}\omega, g(\vartheta^{-1}\omega)\big)\}_{t\in\mathbb{N}}$ be a Cauchy sequence for some $g \in \mathcal{V}$ and all ω, where ϕ denotes the random dynamical system generated by the time-one map $\Psi_\star$. Then its limit $\lim_{t\to\infty} \phi\big(t, \vartheta^{-t}\omega, g\big(\vartheta^{-1}\omega)\big)$ is in $\mathbb{V}$ for all ω, because $\mathbb{V}$ is forward invariant and complete. It remains to show that this limit is contained in $\mathcal{V}$. Since $\lim_{t\to\infty} \phi\big(t, \vartheta^{-t}\omega, g\big(\vartheta^{-1}\omega)\big)$ is bounded by the constant $\overline{p}$, it is bounded by a tempered random variable and hence must itself be tempered.

It follows from steps 1-4 and Theorem 4.3 that there exists a unique random fixed point $p^\star \in \mathcal{V}$ whose domain of attraction contains the set $\mathbb{V} = [\underline{p}, \overline{p}]$.
Step 5. Measurability with respect to the past. Since $\mathcal{V}$ contains all constant functions $g(\omega) = p$, $p \in [\underline{p}, \overline{p}]$ and $p^\star$ is attracting on $[\underline{p}, \overline{p}]$, the stationarity of ϑ implies that $\phi(T, \vartheta^{-T}\omega, p)$ converges a.s. to $p^\star(\omega)$ as $T \to \infty$. Hence, $p^\star$ is measurable with respect to the sub-σ-algebra $\sigma\big(\phi(t, \vartheta^{-T}\cdot, p)|\ 0 \le t \le T,\ p \in [\underline{p}, \overline{p}]\big) \subset \mathcal{F}$.

□

Proof of Theorem 6.2. All Assumptions 5.10-5.15 of Chapter 5 are satisfied, so that Theorem 5.3 can be applied. Recall to this end that the censor map is a variant of the one introduced in Example 5.1, Chapter 5. □

7

Heterogeneous Beliefs in a Financial Market

In this chapter the theoretical concepts of the first chapters are applied to a financial market with interacting agents. In recent years agent-based models have increasingly attracted attention in focusing on heterogeneous market participants as a central building block of a descriptive theory of financial markets. It is a commonly accepted view that agents in financial markets trade, since either preferences, beliefs or wealth positions differ. Their attitudes towards risk and their beliefs about the future development of prices constitute the key influence on the determination of market prices. Behavioral assumptions concerning agents' ability to foresee the future evolution of a market play a central role in modeling agent-based financial markets. As traders, in general, will not share a common belief about future prices, the evolution of actual market prices cannot be self-confirming for all of them. This is the reason why the issue of updating beliefs and learning in asset markets is of particular importance.

Following an approach of Böhm, Deutscher & Wenzelburger (2000) and Böhm & Chiarella (2005), we consider a model with heterogeneous myopic market participants characterized by preferences and subjective beliefs. Such a structure, for instance, arises naturally in models with overlapping generations of consumers. Consumers face investment possibilities in one risk-free and $K \geq 1$ assets. They have no direct access to the financial market and instead choose among *'professional'* financial mediators who carry out their transactions. We distinguish between two levels of rationality in the model. Each financial mediator, who may be thought of as a stylized fund manager or broker, is boundedly rational in the sense of Sargent (1993) and is characterized by a forecasting technology which she uses to form expectations about future asset prices. On the contrary, consumers are boundedly rational in the sense of Simon (1982) and base their choice of a mediator on a performance indicator that measures the success of past investment strategies of the mediator. The imitating behavior of consumers is modeled by a standard discrete-choice approach as found in Anderson, de Palma & Thisse (1992). It constitutes the evolutionary feature of the model which is common in many models including Brock & Hommes (1997a, 1998), Chiarella & He (2002, 2003a,b) or Föllmer, Horst & Kirman (2005), Horst (2005) and many others. Given consumers' preferences and the subjective beliefs of mediators, market-clearing prices at each date are determined by a temporary equilibrium map. Combined with forecasting rules that model the way in which mediators update beliefs, this yields a time-one map of a dynamical system in a random environment of the form discussed in Chapter 2 in which expectations feed back into the actual evolution of asset prices.

The results of the present chapter are refinements of results in Wenzelburger (2004b). Extending the concept of an unbiased forecasting rule of Chapter 2, we introduce perfect forecasting rules for second moments. These together with unbiased forecasting rules, also referred to as perfect forecasting rules for first moments, provide correct first and second moments of the price process conditional on all available information. In this sense, these perfect forecasting rules generate rational expectations for a particular mediator such as a fundamentalist while other market participants may have non-rational beliefs. We show that this concept derives naturally from the aggregate demand function of heterogeneous non-rational investors such as chartists and noise traders. Perfect forecasting rules for first and second moments are required to obtain *mean-variance efficient portfolios* for fundamentalists in the sense of classical CAPM theory.

The natural question how a boundedly rational financial mediator can learn perfect forecasting rules from historical data is addressed. The main informational constraint encountered by a mediator stems from the fact that neither the fraction of consumers choosing a particular mediator nor the beliefs of the mediators themselves are observable quantities. We will show that this missing information can be retrieved by estimating the excess demand function of all market participants. This will, in general, become a highly nonlinear problem. However, in the case when consumers are endowed with linear mean-variance preferences and chartists use simple linear forecasting rules, the main nonlinearity consists of the discrete-choice model which governs the switching behavior of consumers. Since rigorous techniques for estimating a discrete choice model exist (e.g., see Judge, Griffiths, Hill, Lütkepohl & Lee 1985, Chap. 18), it turns out that the problem reduces to estimating a stochastic linear functional relationship. Using the methods introduced in Chapter 3, an adaptive learning scheme is developed which approximates perfect forecasting rules from historical data. Necessary and sufficient conditions for the convergence of the learning scheme are discussed.

7.1 The Model

Consider an overlapping generations model with a finite number of types $h = 1, \dots, H$ of young households, where each type consists of a total number of $N^{(h)}$ consumers, respectively. Each young household of type h lives for two periods, receives an initial endowment $e^{(h)} > 0$ of a non-storable commodity in the first period of life, and does not consume. In order to transfer wealth to the second period of his life, a consumer will choose a portfolio of $K + 1 \in \mathbb{N}$ retradeable assets whose proceeds he will consume. He receives no additional endowment in the second period of his life, so that his total consumption is equal to the return on the investment of his initial endowment when young.

There is one risk-free asset which has a constant real rate of return $R = 1 + r > 0$ given exogenously, i.e., investing one unit of the consumption good in the risk-free asset yields R units of the consumption good in the subsequent period. The other K assets correspond to shares of firms whose production activities induce a stochastic process of dividends which are distributed to the shareholders. The total amount of retradeable risky assets per capita of young households is $\bar{x} \in \mathbb{R}^K_+$.

Young consumers maximize expected utility over future consumption with respect to a von-Neumann-Morgenstern utility function and subjective expectations for future asset prices. Let $y \in \mathbb{R}$ denote the amount of the consumption good invested in the risk-free asset and $x \in \mathbb{R}^K$ be the vector of shares purchased at the *ex-dividend* price vector $p \in \mathbb{R}^K_+$ (in units of the consumption good). Abstracting from short-sale constraints, the budget equation of a consumer of type h is $e^{(h)} = y + p^\top x$. Let $d \in \mathbb{D}$ denote a dividend payment, which is randomly drawn from some subset $\mathbb{D} \subset \mathbb{R}^K_+$, $p' \in \mathbb{R}^K_+$ be the future price vector of the risky assets, and $q = p' + d$ denote the corresponding *cum-dividend* price vector. Since there is no youthful consumption, the future wealth of consumer h associated with the portfolio $x \in \mathbb{R}^K$ of risky assets is

$$w^{(h)}(x, p, q) = R\,\underbrace{(e^{(h)} - p^\top x)}_{\text{risk-free investment}} \quad + \quad \underbrace{q^\top x.}_{\text{equity return}}$$

Assuming that each consumer treats the price p at which he purchases as a parameter, the remaining uncertainty of the portfolio return rests with the cum-dividend price q of shares.

Households have no direct assess to a forecasting technology and select a mediator who solves their individual investment problems using subjective probability distributions. There is a finite number $i = 0, \dots, I$ of financial mediators characterized by subjective beliefs regarding the future cum-dividend price of the assets. Their beliefs are described by subjective probability distributions $\nu^{(i)} \in \text{Prob}(\mathbb{R}^K_+)$, $i = 0, \dots, I$, with $\text{Prob}(\mathbb{R}^K_+)$ denoting the set of all Borelian probability measures on $\mathbb{R}^K_+$. For simplicity, we abstract from intermediation costs for households and suppress the constant $e^{(h)}$ whenever possible. The asset demand of a young consumer of type h based on the subjective belief $\nu^{(i)}$ provided by mediator i is defined by an optimal portfolio choice which maximizes expected utility of future wealth

$$\phi^{(h)}(\nu^{(i)}, p) := \operatorname*{argmax}_{x \in \mathbb{R}^K} \int_{\mathbb{R}^K_+} u^{(h)}\big(w^{(h)}(x,p,q)\big)\nu^{(i)}(dq). \tag{7.1}$$

For the remainder of this chapter we will assume that the von-Neumann-Morgenstern utility functions $u^{(h)}$, $h = 1, \ldots, H$ and the subjective probability measures $\nu^{(i)}$, $i = 0, \ldots, I$, are such that a unique solution (7.1) to the consumer's decision problem exists. In particular, we assume that each distribution $\nu^{(i)}$, $i = 0, \ldots, I$, is such that none of the assets is redundant. Existence and uniqueness can also be guaranteed by suitably restricting the class of utility functions and the class probability measures. This will be done in Definition 7.1 later in this chapter.[1]

We will express all aggregate quantities in terms of per capita of young households. Denote by

$$\overline{\eta}^{(h)} := \frac{N^{(h)}}{\sum_{h=1}^{H} N^{(h)}}$$

the fraction of households of type h and $\eta_t^{(hi)} \in [0, \overline{\eta}^{(h)}]$ be the fraction of households of type h employing mediator i in period t so that $\sum_{i=0}^{I} \eta_t^{(hi)} = \overline{\eta}^{(h)}$. Then

$$W_t^{(i)} = \sum_{h=1}^{H} \eta_t^{(hi)}\, e^{(h)} \tag{7.2}$$

is the amount of resources per capita which mediator i receives from all young households before trading takes place in that period. Her earnings from dividend and interest payments from the portfolio per capita $(x_{t-1}^{(i)}, y_{t-1}^{(i)}) \in \mathbb{R}^K \times \mathbb{R}$ obtained after trading in period $t-1$ are $d_t^\top x_{t-1}^{(i)} + r y_{t-1}^{(i)}$. Since aggregate repayment obligations to old households are $(p + d_t)^\top x_{t-1}^{(i)} + R y_{t-1}^{(i)}$, her budget constraint in period t reads $W_t^{(i)} = p^\top x^{(i)} + y^{(i)}$. Denote by $\eta_t^{(i)} := (\eta_t^{(1i)}, \ldots, \eta_t^{(Hi)}) \in [0,1]^H$ the distribution of households employing mediator i in period t and $\nu_t^{(i)}$ denote mediator i's subjective probability distribution for the future cum-dividend price $q_{t+1} = p_{t+1} + d_{t+1}$ in period t. Then the aggregate demand function for risky assets per capita of all households which employ i in period t is

$$x^{(i)} = \Phi^{(i)}(\eta_t^{(i)}, \nu_t^{(i)}, p) := \sum_{h=1}^{H} \eta_t^{(hi)}\, \phi^{(h)}(\nu_t^{(i)}, p), \quad p \in \mathbb{R}^K_+. \tag{7.3}$$

[1]For further details we refer to the literature, e.g., see Grandmont (1982), Ingersoll (1987), and Pliska (1997).

Finally, denote by $\xi_t \in \mathbb{R}^K$ a random quantity of assets per capita which noise traders hold after trading in period t.[2] Following Böhm, Deutscher & Wenzelburger (2000), the market-clearing asset price vector in period t is defined by a solution $p_t \in \mathbb{R}^K_+$ to

$$\sum_{i=0}^{I} \left[\Phi^{(i)}(\eta_t^{(i)}, \nu_t^{(i)}, p) - x_{t-1}^{(i)}\right] + \xi_t - \xi_{t-1} \stackrel{!}{=} 0. \tag{7.4}$$

If for a fixed distribution of households $\eta_t^{(i)}$ and fixed beliefs $\nu_t^{(i)}$ the (per-capita) aggregate excess demand for assets (7.4) is globally invertible with respect to p, then there exists a unique ex-dividend price p_t which clears the asset market. This implies that the market-clearing asset price vector is a deterministic function of individual characteristics of young consumers and mediators.

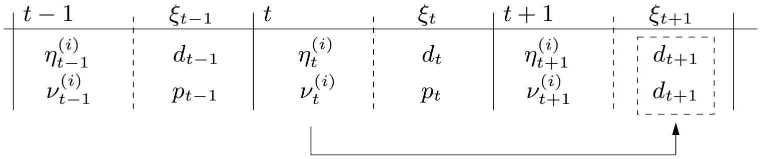

Fig. 7.1. Time-line of price formation.

In order to define a price map, note that the total amount of assets per capita $\bar{x}$ must be equal to the sum of previous positions $\sum_{i=0}^{I} x_{t-1}^{(i)} + \xi_{t-1}$. Let the market shares $(\eta_t^{(i)})_{i=0}^{I} \in \Delta$ with a suitably defined simplex $\Delta \subset [0,1]^{HI}$ and the subjective beliefs $(\nu_t^{(i)})_{i=0}^{I} \in \mathrm{Prob}(\mathbb{R}^K_+)^I$ of all the mediators at date t be given. Then under the above invertibility condition, the asset price vector which clears the market in period t is uniquely determined by the *temporary equilibrium map* $G : \mathbb{R}^K \times \Delta \times \mathrm{Prob}(\mathbb{R}^K_+)^I \to \mathbb{R}^K_+$,

$$p_t = G\big(\xi_t, (\eta_t^{(i)}, \nu_t^{(i)})_{i=0}^{I}\big). \tag{7.5}$$

The map (7.5) is an economic law in the sense of Chapter 2 extended to subjective and heterogeneous probability distributions. Observe that G does not contain the price itself as an argument, such that the map G is of *Cobweb type*. The dating of the expectations $\nu_t^{(i)}$ relative to the actual price p_t reveals that the economic law (7.5) has an expectational lead, that is, expectations are with respect to the realization of prices one period ahead of the map G. This observation is illustrated in Fig. 7.1. The functional form of (7.5) reflects

[2]Noise traders are thought of speculators whose transactions are not captured by a standard microeconomic decision model. Alternative interpretations as given in De Long, Shleifer, Summers & Waldmann (1990, p. 709) apply as well.

the fact that asset prices are essentially determined by subjective beliefs of the financial mediators and the way in which households choose among mediators.

The decision of a household is based on the performance of a mediator. The general idea is that the market shares of mediators $(\eta_t^{(i)})_{i=0}^I$ depend on the empirical distribution of the resulting price process and possibly other information. There are numerous ways of modeling this type of evolutionary behavior. A convenient way of formalizing this behavioral assumption is by means of a discrete choice approach (see Anderson, de Palma & Thisse 1992) and will be postponed to Section 7.4. For the next three sections, it suffices to assume that the market shares $\{(\eta_t^{(0)}, \dots, \eta_t^{(I)})\}_{t \in \mathbb{N}}$ are driven by some endogenous process which depends on empirical observations of the past.

To describe the evolution of the asset prices, we need to specify the probabilistic assumptions on the exogenous noise and the exogenous dividend process.

Assumption 7.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$ an increasing family of sub-σ-algebras of $\mathcal{F}$. Then we assume the following.*

(i) The dividend payments are described by a $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$-adapted stochastic process $\{d_t\}_{t \in \mathbb{N}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{D} \subset \mathbb{R}_+^K$. The process is predictable *meaning each d_t, $t \in \mathbb{N}$ is $\mathcal{F}_{t-1}$ measurable.*

(ii) The noise traders' transactions are governed by a $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$-adapted stochastic process $\{\xi_t\}_{t \in \mathbb{N}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^K$ which is uncorrelated with the dividend process $\{d_t\}_{t \in \mathbb{N}}$ defined in (*i*).

Given the price law (7.5) for ex-dividend prices and Assumption 7.1 on the two exogenous processes, the cum-dividend price in period t is

$$q_t = G\big(\xi_t, (\eta_t^{(i)}, \nu_t^{(i)})_{i=0}^I\big) + d_t, \quad t \in \mathbb{N}. \tag{7.6}$$

Assumption 7.1 states that the dividend payment d_t of period t is known whereas the noise trader transaction ξ_t is unknown before trading in period t. Mathematically this means that d_t is assumed to be measurable with respect to $\mathcal{F}_{t-1}$, while ξ_t is assumed to be measurable with respect to $\mathcal{F}_t$. Since the subjective beliefs $(\nu_t^{(i)})_{i=0}^I$ and the market shares $(\eta_t^{(i)})_{i=0}^I$ must be set prior to trading in period t, they are based on information observable up to that time and thus must be $\mathcal{F}_{t-1}$ measurable as well. This implies that the uncertainty in ex-dividend prices and in the traded quantities of assets rests solely on the behavior of the noise traders. In other words, the randomness of ex-dividend prices is essentially due to the stochastic nature of noise-trader behavior.

Remark 7.1. *The basic structure of the model follows Böhm, Deutscher & Wenzelburger (2000) and Böhm & Chiarella (2005). It should be emphasized that the underlying OLG structure is not essential for this setup. A generalization to multiperiod planning horizons is provided by Hillebrand (2003).*

7.2 Heterogeneous Beliefs

This section is concerned with unbiased forecasting rules that generate correct first moments of the price process along orbits of the system. The problem is central in agent-based financial markets, since investors by their very nature have heterogeneous beliefs which they will not share. It has long been recognized that the two requirements of market clearing in all periods and rational expectations at all times can often not be fulfilled simultaneously. As indicated in Chapter 2 (see also Böhm & Wenzelburger 2002), the evaluation of a forecasting rule can be carried out on several levels. For the purpose of the present model, we restrict our analysis to the comparison of the first two moments of subjective and true distributions, i.e., on the conditional mean values and the conditional (variance-) covariance matrices.

To simplify the analysis, assume that each mediator $i = 0, \dots, I$ picks his subjective distribution $\nu^{(i)}$ from a fixed family of probability distributions parameterized in subjective mean values $q^{(i)}$ and subjective covariance matrices $V^{(i)}$ such that $\nu^{(i)} \equiv \nu_{q^{(i)},V^{(i)}}$. Abstract from second moments beliefs and identify each subjective distribution $\nu_t^{(i)}$ at a particular date t with the corresponding subjective mean value $q_t^{(i)}$ for the future cum-dividend price $q_{t+1} = p_{t+1} + d_{t+1}$. Let j be an arbitrary mediator and

$$q_t^{(-j)} = (q_t^{(0)}, \dots, q_t^{(j-1)}, q_t^{(j+1)}, \dots, q_t^{(I)}) \in \mathbb{R}_+^{KI}$$

denote the respective subjective means of all mediators $i \neq j$. By abuse of notation, mediator j's forecast error on the price q_t prevailing at date t is

$$q_t - q_{t-1}^{(j)} = G\big(\xi_t, (\eta_t^{(i)})_{i=0}^I, q_t^{(j)}, q_t^{(-j)}\big) + d_t - q_{t-1}^{(j)}, \tag{7.7}$$

where the map G is given in (7.5) and $q_{t-1}^{(j)}$ denotes the forecast for q_t made at date $t-1$. The functional form in (7.7) reveals that the forecast error for mediator j depends crucially on the current forecasts of all mediators.

Assume for a moment that the market shares $(\eta_t^{(i)})_{i=0}^I$ as well as the forecasts $q_t^{(-j)}$ of the other mediators are known to mediator j. Since d_t was assumed to be $\mathcal{F}_{t-1}$ measurable, all uncertainty rests on the random transaction ξ_t of the noise-traders. Let $\mathbb{E}_{t-1}$ denote the expectations operator with respect to $\mathcal{F}_{t-1}$. In view of Chapter 2, all possible mean forecast errors on cum-dividend prices for mediator j at a particular date t can be described by a *mean error function*. In the context of this model, this mean error function is a map $\mathcal{E}_t^{(j)} : \Delta \times \mathbb{R}_+^{K(I+1)} \times \mathbb{D} \times \mathbb{R}_+^K \longrightarrow \mathbb{R}^K$, defined by

$$\begin{aligned} &\mathcal{E}_t^{(j)}\big((\eta_t^{(i)})_{i=0}^I, q_t^{(j)}, q_t^{(-j)}, d_t, q_{t-1}^{(j)}\big) \\ &\qquad := \mathbb{E}_{t-1}\Big[G\big(\xi_t, (\eta_t^{(i)})_{i=0}^I, q_t^{(j)}, q_t^{(-j)}\big)\Big] + d_t - q_{t-1}^{(j)}, \end{aligned} \tag{7.8}$$

which describes all possible mean errors between actual cum-dividend prices and cum-dividend forecasts for mediator j conditional on information available prior to trading in period t. An *unbiased forecasting rule* for mediator j

at date t, if it exists, is a map $\psi_t^{(j)} : \Delta \times \mathbb{R}_+^{KI} \times \mathbb{D} \times \mathbb{R}_+^K \to \mathbb{R}_+^K$ with

$$q_t^{(j)} = \psi_t^{(j)}\big((\eta_t^{(i)})_{i=0}^I, q_t^{(-j)}, d_t, q_{t-1}^{(j)}\big) \tag{7.9}$$

such that

$$\mathcal{E}_t^{(j)}\big((\eta_t^{(i)})_{i=0}^I, \psi_t^{(j)}\big((\eta_t^{(i)})_{i=0}^I, q_t^{(-j)}, d_t, q_{t-1}^{(j)}\big), q_t^{(-j)}, d_t, q_{t-1}^{(j)}\big) = 0$$

identically on $\Delta \times \mathbb{R}_+^{K(I+1)} \times \mathbb{D} \times \mathbb{R}_+^K$. As shown in Chapter 4, Section 4.2, an unbiased forecasting rule (7.9) exists if the error function (7.8) satisfies the conditions of a global implicit function theorem. Recall that the implicit function theorem stipulates the particular functional form of (7.9).

We will show below that an unbiased forecasting rule exists in the case in which all households have linear mean-variance preferences. An unbiased forecasting rule for ex-dividend prices is obtained from (7.9) by setting $p_t^{(j)} = q_t^{(j)} - d_{t+1}$ as a forecast for p_{t+1}. It then follows from (7.8) that all forecast errors on ex-dividend prices for mediator j vanish identically, that is, $\mathbb{E}_t[p_{t+1}] - p_t^{(j)} = 0$. In the sequel, an unbiased forecasting rule will also be referred to as a *perfect forecasting rule for first moments.*

The underlying economic intuition of an unbiased forecasting rule is as follows. By abuse of notation, write $\Phi^{(i)}(\eta^{(i)}, \nu^{(i)}, p) \equiv \Phi^{(i)}(\eta^{(i)}, q^{(i)}, p)$ for the respective aggregate demand functions (7.3), where we identify $\nu^{(i)}$ with $q^{(i)}$ as before. Assume, in addition, that for each $p \in \mathbb{R}_+^K$ and each $\eta^{(j)}$ the aggregate demand $\Phi^{(j)}(\eta^{(j)}, q^{(j)}, p)$ of mediator j is invertible with respect to the price forecasts $q^{(j)}$. Let $q_{t-1}^{(j)} \in \mathbb{R}_+^K$ and $(\eta_t^{(i)})_{i=0}^I$ be arbitrary and p_t denote the market-clearing price in period t. The market-clearing condition (7.4) then implies that ex post, an unbiased forecasting rule (7.9) for j yielding $q_t^{(j)}$ must satisfy

$$q_t^{(j)} = \Phi^{(j)-1}\Big(\eta_t^{(j)}, p_t, \bar{x} - \xi_t - \sum_{i \neq j} \Phi^{(i)}(\eta_t^{(i)}, q_t^{(i)}, p_t)\Big). \tag{7.10}$$

The functional relationship (7.10) demonstrates that the essential unknown quantity for any mediator j is the excess demand of all other market participants. For any mediator, the existence of an unbiased forecasting rule thus depends exclusively on the fundamentals of the market mechanism, the subjective beliefs of all mediators who take part in the market, and on the way in which households choose their mediators. Consequently, any mediator with rational expectations has to be able to replicate the demand behavior of his fellow mediators. The main informational constraint is the fact that neither the market shares of a mediator nor her beliefs are observable quantities.

The notion of a *perfect forecasting rule for second moments*, i.e., a forecasting rule which generates correct covariances matrices of the cum-dividend prices conditioned on the available information, could be derived using an analogous reasoning. We will postpone the discussion of such a rule to Section 7.5 when a more concrete example is available.

7.3 Risk Premia and Reference Portfolios

In order to evaluate the performance of a mediator and thus the quality of households' investment decisions, we analyze the risk premia obtained by mediators. We seek a portfolio which in the presence of diverse beliefs is *(mean-variance) efficient* in the classical sense of CAPM. Naturally, such a portfolio must take into account discrepancies in beliefs. Throughout this section we assume that each conditional covariance matrix $\mathbb{V}_t[q_{t+1}]$, $t \in \mathbb{N}$ is positive definite and hence invertible. For each $t \in \mathbb{N}$ define now the *reference portfolio* of period t by

$$x_t^{\text{ref}} = \mathbb{V}_t[q_{t+1}]^{-1}[\mathbb{E}_t[q_{t+1}] - Rp_t]. \tag{7.11}$$

The reference portfolio x_t^{ref} of period t is a 'fictitious' portfolio in the sense that it is not necessarily held by a trader. It is determined after the realization of asset prices in period t. It is convenient to think of an investor who holds the reference portfolio after fictitious trading in period t. Such a fictitious investor may be an outside observer with linear mean-variance preferences with risk aversion 1 and rational expectations about future asset prices. Investing one unit of the consumption good into the risk-free and risky assets, gives the portfolio $(x_t^{\text{ref}}, 1 - p_t^\top x_t^{\text{ref}}) \in \mathbb{R}^K \times \mathbb{R}$. If $r = R - 1$ is the interest rate payed for the risk-free asset, then the return of the portfolio (7.11) obtained after a (fictitious) selling at prices in period $t+1$ is

$$R_{t+1}^{\text{ref}} = r + [q_{t+1} - Rp_t]^\top x_t^{\text{ref}}. \tag{7.12}$$

The conditional variance $\mathbb{V}_t[R_{t+1}^{\text{ref}}]$ of R_{t+1}^{ref} is given by

$$\begin{aligned} \mathbb{V}_t[R_{t+1}^{\text{ref}}] &= x_t^{\text{ref}\top}\mathbb{V}_t[q_{t+1}]x_t^{\text{ref}} \qquad (7.13) \\ &= \big[\mathbb{E}_t[q_{t+1}] - Rp_t\big]^\top \mathbb{V}_t[q_{t+1}]^{-1}\big[\mathbb{E}_t[q_{t+1}] - Rp_t\big]. \end{aligned}$$

This implies that the *(conditional) risk premium* of the reference portfolio satisfies

$$\mathbb{E}_t[R_{t+1}^{\text{ref}}] - r = \mathbb{V}_t[R_{t+1}^{\text{ref}}]. \tag{7.14}$$

The reference portfolio of a financial market is thus characterized by the property that its risk premium, also referred to as *mean equity premium*, is equal to its variance. For this reason, the risk premium (7.14) of the reference portfolio is always non-negative.

Let $x_t^{(i)}$ denote the portfolio held by mediator i in period t after investing $W_t^{(i)}$ given in (7.2). Then the realized return $R_{t+1}^{(i)}$ from selling $x_t^{(i)}$ in period $t+1$ is

$$R_{t+1}^{(i)} = r + \tfrac{1}{W_t^{(i)}}[q_{t+1} - Rp_t]^\top x_t^{(i)}. \tag{7.15}$$

The conditional covariance $\mathbb{C}\text{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\text{ref}}]$ between the return of mediator i's portfolio $x_t^{(i)}$ and the return of the reference portfolio (7.11) is

$$\mathbb{C}\mathrm{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\mathrm{ref}}] = \tfrac{1}{W_t^{(i)}}\, x_t^{(i)\top} \mathbb{V}_t[q_{t+1}] x_t^{\mathrm{ref}}. \tag{7.16}$$

Inserting (7.11) into (7.16), we see that the risk premium of mediator i's portfolio takes the form

$$\mathbb{E}_t[R_{t+1}^{(i)}] - r = \mathbb{C}\mathrm{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\mathrm{ref}}]. \tag{7.17}$$

Thus the risk premium of mediator i is equal to the covariance between i's return and the return of the reference portfolio which, contrary to that of the reference portfolio, may well be negative. Combining (7.17) with (7.14) yields the following theorem.

Theorem 7.1. *Let i be an arbitrary mediator and assume that preferences and beliefs of all market participants are such that asset markets clear in each period $t \in \mathbb{N}$ at prices p_t. Moreover, assume that $\mathbb{V}_t[q_{t+1}]$ is positive definite for all $t \in \mathbb{N}$ such that the reference portfolio is well defined. Then for each $t \in \mathbb{N}$,*

$$\mathbb{E}_t[R_{t+1}^{(i)}] - r = \frac{\mathbb{C}\mathrm{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\mathrm{ref}}]}{\mathbb{V}_t[R_{t+1}^{\mathrm{ref}}]} \big[\mathbb{E}_t[R_{t+1}^{\mathrm{ref}}] - r\big].$$

Theorem 7.1 states that the risk premium (7.17) of any mediator i can only be higher than the risk premium of the reference portfolio (7.11) at the expense of higher risk, i.e. $\mathbb{V}_t[R_{t+1}^{(i)}] \geq \mathbb{V}_t[R_{t+1}^{\mathrm{ref}}]$. Using the inequality

$$\mathbb{C}\mathrm{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\mathrm{ref}}] \leq \sqrt{\mathbb{V}_t[R_{t+1}^{(i)}]}\sqrt{\mathbb{V}_t[R_{t+1}^{\mathrm{ref}}]},$$

Theorem 7.1 implies that for each mediator $i = 0, \dots, I$, the *Sharpe ratios* (conditional on information at date t) satisfy

$$\frac{\mathbb{E}_t[R_{t+1}^{(i)}] - r}{\sqrt{\mathbb{V}_t[R_{t+1}^{(i)}]}} \leq \frac{\mathbb{E}_t[R_{t+1}^{\mathrm{ref}}] - r}{\sqrt{\mathbb{V}_t[R_{t+1}^{\mathrm{ref}}]}} = \sqrt{\mathbb{V}_t[R_{t+1}^{\mathrm{ref}}]} \quad \text{for all times } \ t. \tag{7.18}$$

As can be seen from (7.13), the upper bound in (7.18) is essentially determined by the endogenous price process. Using (7.15), the Sharpe ratios (7.18) of mediator i's returns take the form

$$\frac{\mathbb{E}_t[R_{t+1}^{(i)}] - r}{\sqrt{\mathbb{V}_t[R_{t+1}^{(i)}]}} = \frac{\big[\mathbb{E}_t[q_{t+1}] - Rp_t\big]^\top x_t^{(i)}}{\sqrt{x_t^{(i)\top}\mathbb{V}_t[q_{t+1}]x_t^{(i)}}}, \tag{7.19}$$

showing that these are invariant under rescaling of the portfolio $x_t^{(i)}$. Any portfolio $\tilde{x}_t^{(i)} = \lambda\, x_t^{(i)}$ with $\lambda > 0$ will have the same Sharpe ratio as $x_t^{(i)}$ so that the size of a portfolio has no influence on the realized Sharpe ratios of a mediator.

Theorem 7.1 is a generalization of the famous *security market line result* (e.g., see LeRoy & Werner (2001) or Pliska 1997) to asset markets with heterogenous beliefs and preferences which are not necessarily of the mean-variance type. It demonstrates that in a world of heterogeneous investors, perfect forecasting rules for first and second moments, if they exists, generate mean-variance efficient portfolios in the sense of classical CAPM theory. Observe that the market-clearing condition in the proof of the theorem is not used so that the theorem easily generalizes to asset markets that do not necessarily clear such as electronic stock markets, cf. Li & Wenzelburger (2005).

7.4 Mean-Variance Preferences

Starting with the work by Markowitz (1952) and Tobin (1958), mean-variance preferences provide a popular parameterization of preferences under risk. In the linear case, it is shown in Böhm & Chiarella (2005) that the resulting aggregate demand function for shares is explicitly invertible. This allows for an explicit functional form of the market-clearing prices which greatly facilitates the analysis of adaptive learning. The relationship between mean-variance preferences and expected utility theory is given in the following definition.

Definition 7.1. *Let* $U : \mathbb{R} \times \mathbb{R}_+ \longrightarrow \mathbb{R}$ *denote a strictly concave function which is strictly increasing in its first component and strictly decreasing in its second. U is said to represent mean-variance preferences if there exists a von-Neumann-Morgenstern utility function* $u : \mathbb{R} \longrightarrow \mathbb{R}$ *and a class of probability measures* $\mathcal{P} \subset \mathrm{Prob}(\mathbb{R})$ *such that*

$$\int_{\mathbb{R}} u(w')\tau(dw') = U\big(\mathbb{E}_\tau[w], \sqrt{\mathbb{V}_\tau[w]}\big), \quad \textit{for all } \tau \in \mathcal{P}.$$

Here, $\mathbb{E}_\tau$ *and* $\mathbb{V}_\tau$ *denote expected value and variance with respect to* τ*, respectively.*

As before, identify each subjective probability distribution $\nu_t^{(i)}$ on $\mathbb{R}^K$ with its respective parameterization $(q_t^{(i)}, V_t^{(i)})$ and let

$$w^{(h)}(x, p, q) = Re^{(h)} + (q - Rp)^\top x$$

denote future wealth of household h associated with the portfolio $x \in \mathbb{R}^K$.[3] Then based on the belief $(q_t^{(i)}, V_t^{(i)})$, the subjective expected wealth of the portfolio $x \in \mathbb{R}^K$ and its subjective variance conditional on information available prior to trading in period t are

[3]The fact that the probability distributions are now defined on the whole of $\mathbb{R}^K$ becomes a subtle technical necessity when linking expected utility theory with mean-variance preferences as done in Definition 7.1.

$$\mathbb{E}_t^{(i)}[w^{(h)}(x,p,\cdot)] = Re^{(h)} + (q_t^{(i)} - Rp)^\top x$$

and

$$\mathbb{V}_t^{(i)}[w^{(h)}(x,p,\cdot)] = x^\top V_t^{(i)} x,$$

respectively. Assume for the remainder of this chapter that each household h is characterized by linear mean-variance preferences, such that

$$U^{(h)}\left(\mathbb{E}_t^{(i)}[w^{(h)}(x,p,\cdot)], \sqrt{\mathbb{V}_t^{(i)}[w^{(h)}(x,p,\cdot)]}\right)$$
$$= \mathbb{E}_t^{(i)}[w^{(h)}(x,p,\cdot)] - \tfrac{\alpha^{(h)}}{2}\mathbb{V}_t^{(i)}[w^{(h)}(x,p,\cdot)],$$

describes the expected utility associated with the portfolio $x \in \mathbb{R}^K$ in period t, where $\alpha^{(h)} > 0$ measures risk aversion.

Given the market shares $\eta_t^{(hi)}$ of all mediators, the (per-capita) aggregate demand (7.3) of all households h which employ i is given by

$$\Phi^{(i)}(a_t^{(i)}, q_t^{(i)}, V_t^{(i)}, p) = a_t^{(i)} V_t^{(i)-1}[q_t^{(i)} - Rp], \tag{7.20}$$

where

$$a_t^{(i)} := \sum_{h=1}^{H} \frac{\eta_t^{(hi)}}{\alpha^{(h)}}, \quad i = 0, \ldots, I \tag{7.21}$$

denotes the risk-adjusted market share of the mediator i, $i = 0, \ldots, I$. Set

$$A_t := \left(\sum_{i=0}^{I} a_t^{(i)} V_t^{(i)-1}\right)^{-1} \tag{7.22}$$

and note that A_t is well defined and positive definite due to the positive definiteness of all covariance matrices $V_t^{(i)}$. From the market-clearing condition (7.4), one immediately obtains an explicit functional form of the ex-dividend price law (7.5) which for arbitrary beliefs $(q_t^{(i)}, V_t^{(i)})_{i=0}^I$ takes the form

$$p_t = G\big(\xi_t, (a_t^{(i)}, q_t^{(i)}, V_t^{(i)})_{i=0}^I\big) := \frac{1}{R}\left(\sum_{i=0}^{I} A_t^{(i)}\, q_t^{(i)} \; - A_t(\bar{x} - \xi_t)\right) \tag{7.23}$$

with

$$A_t^{(i)} := a_t^{(i)} A_t V_t^{(i)-1} = \begin{cases} 0 & \text{if } a_t^{(i)} = 0, \\ \left[I_K + \displaystyle\sum_{j \neq i} \frac{a_t^{(j)}}{a_t^{(i)}} V_t^{(i)}\, V_t^{(j)-1} \right]^{-1} & \text{if } a_t^{(i)} > 0. \end{cases} \tag{7.24}$$

All coefficient matrices (7.22) and (7.24) in the price law (7.23) are determined by subjective covariances matrices $V_t^{(i)}$ and risk-adjusted distributions of households (7.21) and hence are $\mathcal{F}_{t-1}$ measurable.

The evolution of the market shares is modeled using a LOGIT model, see Anderson, de Palma & Thisse (1992).[4] As before, let $\eta_t^{(hi)}$ denote the fraction of households h which employ mediator i and $r = R - 1$ be the interest rate payed for the risk-free asset. The return obtained by mediator i after trading in period t is

$$R_t^{(i)} = r + \frac{a_{t-1}^{(i)}}{W_{t-1}^{(i)}} [q_t - Rp_{t-1}]^\top V_{t-1}^{(i)-1} [q_{t-1}^{(i)} - Rp_{t-1}]. \tag{7.25}$$

For each i, the sample mean $\hat{\mu}_t^{(i)}$ and the sample standard deviation $\hat{\sigma}_t^{(i)}$ of the time series $\{R_s^{(i)}\}_{s=0}^t$ are given by

$$\begin{aligned} \hat{\mu}_t^{(i)} &:= \frac{1}{t+1} \sum_{s=0}^t R_s^{(i)}, \\ \hat{\sigma}_t^{(i)} &:= \left[\frac{1}{t+1} \sum_{s=0}^t (R_s^{(i)} - \hat{\mu}_t^{(i)})^2 \right]^{\frac{1}{2}} = \left[\frac{1}{t+1} \sum_{s=0}^t R_s^{(i)2} - \hat{\mu}_t^{(i)2} \right]^{\frac{1}{2}}, \end{aligned} \tag{7.26}$$

where $\hat{\mu}_{-1}^{(i)} \geq 0$ and $\hat{\sigma}_{-1}^{(i)} > 0$. Using the empirical moments (7.26), an estimator for the *Sharpe ratio* associated with the realized returns (7.25) of mediator i is given by $(\hat{\mu}_t^{(i)} - r)/\hat{\sigma}_t^{(i)}$. Recall that the fraction of households of type h is $\overline{\eta}^{(h)}$. The fraction $\eta_t^{(hj)}$ of households of type h which employs mediator j in period t is now assumed to be determined by the *discrete-choice probability*

$$\eta_t^{(hj)} := \overline{\eta}^{(h)} \frac{\exp\left(\beta^{(h)} \left(\hat{\mu}_{t-1}^{(j)} - r\right)/\hat{\sigma}_{t-1}^{(j)}\right)}{\sum_{i=0}^I \exp\left(\beta^{(h)} \left(\hat{\mu}_{t-1}^{(i)} - r\right)/\hat{\sigma}_{t-1}^{(i)}\right)}, \quad t \geq 0. \tag{7.27}$$

The parameter $\beta^{(h)}$ appearing in the modified *logit function* (7.27) describes the *intensity of choice* of a household of type h, that is, how fast a typical consumer of type h will switch to a different mediator.

The reference portfolio takes the following form. Inserting (7.23) into (7.11) and using $I_K = \sum_{i=0}^I A_t^{(i)}$ for all times t, we obtain

$$x_t^{\mathrm{ref}} = \mathbb{V}_t[q_{t+1}]^{-1} A_t(\bar{x} - \xi_t) + \sum_{i=0}^I \mathbb{V}_t[q_{t+1}]^{-1} A_t^{(i)} \left[\mathbb{E}_t[q_{t+1}] - q_t^{(i)}\right]. \tag{7.28}$$

Formula (7.28) allows us to interpret the reference portfolio (7.11) as a *'modified'* market portfolio that accounts for discrepancies in incorrect beliefs. Indeed, if all mediators had correct first and second conditional moments of the

[4] There are numerous ways of modeling the switching behavior of boundedly rational households, see Brock & Hommes (1998). We will follow the intuition that boundedly rational consumers with mean-variance preferences would prefer portfolios with the highest empirical Sharpe ratio because such portfolios promise to be mean-variance efficient in the sense of the classical CAPM theory.

price process, that is, $\mathbb{E}_t[q_{t+1}] = q_t^{(i)}$ and $\mathbb{V}_t[q_{t+1}] = V_t^{(i)}$ for all $i = 0, \dots, I$, then

$$x_t^{\mathrm{ref}} = \frac{1}{\sum_{h=1}^{H} \frac{\overline{\eta}^{(h)}}{\alpha^{(h)}}} \, (\bar{x} - \xi_t).$$

This shows that the reference portfolio (7.11) is collinear to the (per-capita) market portfolio $\bar{x}$ if all mediators have rational expectations and noise traders are absent.

Remark 7.2. *The price law (7.23) as well as the resulting price process would be the same for consumers with infinite lives who maximize wealth myopically. The setup can thus be seen as an extension of Brock & Hommes (1997a, 1998) and Chiarella & He (2002) to H types of households, I mediators, and K risky assets. A generalization of this setup to multiperiod planning horizons is provided in Hillebrand & Wenzelburger (2006a,b).*

7.5 Perfect Forecasting Rules

In this section we investigate the existence of forecasting rules which generate rational expectations on cum-dividend prices for mediator 0 in the sense that the first two moments of the future asset prices are correctly predicted. In view of mediators who seek mean-variance efficient portfolios, Section 7.3 showed that it suffices to analyze the case in which the first two moments of subjective and true distributions, i.e., the conditional mean values and the conditional covariance matrices, coincide. For brevity we will use the term rational expectations to describe the situation in which a boundedly rational mediator is able to correctly predict the first two moments of the price process conditional on all available information, whereas other market participants may have non-rational beliefs.

Consider mediator 0 and assume without loss of generality that $a_t^{(0)} > 0$ for all times $t \in \mathbb{N}$ throughout this section. This can always be guaranteed by setting $\eta_0^{(h'0)} > 0$ and $\beta^{(h')} = 0$ for some household h' in the logit functions (7.27). On the other hand $a_t^{(0)} = 0$ imposes no restriction on the rationality of 0 because it simply means that no household employs mediator 0 such that her beliefs do not feed back into the price process. In this case there is no existence problem and the perfect forecasting rules simply coincide with the first two moments of the corresponding price process. We will first address the existence of perfect forecasting rules for first moments and then the existence of perfect forecasting rules for second moments.

7.5.1 Perfect Forecasting Rules for First Moments

Recalling that d_t is assumed to be $\mathcal{F}_{t-1}$ measurable, the expected cum-dividend price conditioned on information available at date t associated with the price law (7.23) is

$$\mathbb{E}_{t-1}[q_t] = \frac{1}{R}\left[\sum_{i=1}^{I} A_t^{(i)} q_t^{(i)} + A_t^{(0)} q_t^{(0)} - A_t(\bar{x} - \mathbb{E}_{t-1}[\xi_t])\right] + d_t. \tag{7.29}$$

The condition that the forecast errors for mediator 0 vanish in the mean is

$$\mathbb{E}_{t-1}[q_t - q_{t-1}^{(0)}] = 0 \tag{7.30}$$

for all times t. Inserting (7.29) into (7.30) and rearranging, yields an explicit expression for the new forecast $q_t^{(0)}$, given by

$$q_t^{(0)} = A_t^{(0)-1}\left[R(q_{t-1}^{(0)} - d_t) - \sum_{i=1}^{I} A_t^{(i)} q_t^{(i)} + A_t(\bar{x} - \mathbb{E}_{t-1}[\xi_t])\right]. \tag{7.31}$$

Replacing the coefficients in (7.31) with (7.24), the *unbiased forecasting rule* for mediator 0 takes the form

$$\begin{aligned} q_t^{(0)} &= \psi^{(0)}\big(\mathbb{E}_{t-1}[\xi_t], d_t, (a_t^{(i)}, q_t^{(i)}, V_t^{(i)})_{i=1}^{I}, q_{t-1}^{(0)}\big) \qquad (7.32)\\ &:= R(q_{t-1}^{(0)} - d_t) \\ &\qquad - a_t^{(0)-1} V_t^{(0)}\left[\sum_{i=1}^{I} a_t^{(i)} V_t^{(i)-1}\big[q_t^{(i)} - R(q_{t-1}^{(0)} - d_t) + \mathbb{E}_{t-1}[\xi_t] - \bar{x}\big]\right]. \end{aligned}$$

The functional form of the unbiased forecasting rule (7.32), also referred to as a *perfect forecasting rule for first moments*, is closely related to that of (7.10) in Section 7.2. The quantity in (7.32) which a priori is unknown to mediator 0 is the term in the large brackets. Since by (7.30) $\mathbb{E}_{t-1}[p_t] = q_{t-1}^{(0)} - d_t$, this term describes precisely the expected excess demand of all mediators $i > 0$ including the expected demand of the noise traders $\mathbb{E}_{t-1}[\xi_t]$. The functional form (7.32) of the unbiased forecasting rule thus reflects the intuition that the precision of forecasts requires precise knowledge of the investment behavior of all market participants.

7.5.2 Perfect Forecasting Rules for Second Moments

While in the case of linear mean-variance preferences perfect forecasting rules for first moments always exist, stricter requirements for forecasting rules which generate correct second moments of the price process are needed. Such rules will be referred to as *perfect forecasting rules for second moments.* Observe that the subjective covariance matrix $V_t^{(0)}$ in the expression for (7.32) was until now arbitrary. We will now try to determine this matrix as to provide correct second moments of the price process.

Since by Assumption 7.1 (ii) ξ_t and d_t are uncorrelated, the conditional covariance matrix of the cum-dividend prices are obtained from the price law (7.23) and (7.29), yielding

$$\mathbb{V}_{t-1}[q_t] = \mathbb{V}_{t-1}\left[G\big(\xi_t, (a_t^{(i)}, q_t^{(i)}, V_t^{(i)})_{i=0}^I\big)\right] = \frac{1}{R^2} A_t \mathbb{V}_{t-1}[\xi_t] A_t. \qquad (7.33)$$

Recall that $A_t = \left(\sum_{i=0}^I a_t^{(i)} V_t^{(i)-1}\right)^{-1}$ is positive definite such that $\mathbb{V}_{t-1}[q_t]$ is invertible if $\mathbb{V}_{t-1}[\xi_t]$ is non-degenerate. This implies that the variance and the covariance of the price process receives an expectations feedback from the subjective covariance matrices $(V_t^{(i)})_{i=0}^I$ as long as there is noise in the system, i.e., $\mathbb{V}_{t-1}[\xi_t]$ is different from the zero matrix. The volatility in the price process (7.33) is therefore exclusively generated by the noise-trader behavior, the switching behavior of households, and the subjective covariance matrices of the mediators.

The condition that mediator 0's forecast errors for second moments (7.33) of the price process vanish for all times t is

$$\mathbb{V}_{t-1}[q_t] - V_{t-1}^{(0)} = \frac{1}{R^2} A_t \mathbb{V}_{t-1}[\xi_t] A_t - V_{t-1}^{(0)} \overset{!}{=} 0. \qquad (7.34)$$

The idea for a perfect forecasting rule for second moments is analogous to the idea for perfect rules for first moments: choose a new forecast $V_t^{(0)}$ such that (7.34) holds and the old forecast $V_{t-1}^{(0)}$ becomes correct. Inserting (7.22) into (7.34) yields the necessary condition

$$\left(\sum_{i=0}^I a_t^{(i)} V_t^{(i)-1}\right)^{-1} \mathbb{V}_{t-1}[\xi_t] \left(\sum_{i=0}^I a_t^{(i)} V_t^{(i)-1}\right)^{-1} = R^2\, V_{t-1}^{(0)} \qquad (7.35)$$

for the subjective covariance matrix $V_t^{(0)}$. As a consequence, perfect forecasting rules for second moments are determined by symmetric positive definite solutions $V_t^{(0)}$ to (7.35).

The existence proof is divided into two steps. First find a symmetric positive definite matrix Π_t that solves the equation $\Pi_t^{-1}\mathbb{V}_{t-1}[\xi_t]\Pi_t^{-1} = R^2 V_{t-1}^{(0)}$. Then set

$$V_t^{(0)} = a_t^{(0)} \left(\Pi_t - \sum_{i=1}^I a_t^{(i)} V_t^{(i)-1}\right)^{-1}.$$

If this matrix is positive definite, then the desired second moment is found. In order to state the conditions under which this is the case and a perfect forecasting rule for second moments exists, we define the square root of a symmetric positive definite matrix. Recall that any symmetric and positive definite $K \times K$ matrix B can be diagonalized so that $B = O^\top \operatorname{diag}(\lambda_1, \dots, \lambda_K) O$ for real eigenvalues $\lambda_1, \dots, \lambda_K > 0$ and for some orthogonal matrix O, i.e. $O^\top O = I_K$. The square root $\sqrt{B}$ of B is a symmetric positive definite matrix, defined by $\sqrt{B} := O^\top \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_K}) O$, such that $B = \sqrt{B}\sqrt{B}$. We summarize these findings in the following proposition. All technical details of the proof are found in the mathematical appendix, Section 7.8.

Proposition 7.1. *Let the hypotheses of Assumption 7.1 be satisfied and suppose that each* $\Lambda_{t-1} := \mathbb{V}_{t-1}[\xi_t]$, $t \in \mathbb{N}$ *is positive definite. Set*

$$\Pi_t = \frac{1}{R}\sqrt{\Lambda_{t-1}}\sqrt{\left(\sqrt{\Lambda_{t-1}}\,V_{t-1}^{(0)}\,\sqrt{\Lambda_{t-1}}\right)^{-1}}\sqrt{\Lambda_{t-1}}, \quad t \in \mathbb{N}. \tag{7.36}$$

Then the forecasting rule for second moments $\varphi^{(0)}$, *given by*

$$\begin{aligned} V_t^{(0)} &= \varphi^{(0)}\big(\Lambda_{t-1}, (a_t^{(i)}, V_t^{(i)})_{i=1}^{I}, a_t^{(0)}, V_{t-1}^{(0)}\big) \\ &:= a_t^{(0)}\left(\Pi_t - \sum_{i=1}^{I} a_t^{(i)} V_t^{(i)-1}\right)^{-1}, \end{aligned} \tag{7.37}$$

provides correct second moments of the price process at date t in the sense that (7.34) holds whenever $\Pi_t - \sum_{i=1}^{I} a_t^{(i)} V_t^{(i)-1}$ *is positive definite and* $a_t^{(0)} > 0$.

A special case occurs when noise traders induce zero correlation between different assets and $\mathbb{V}_{t-1}[\xi_t] \equiv \sigma_\xi^2 I_K$. Then Π_t reduces to $\Pi_t = \frac{\sigma_\xi}{R}\sqrt{V_{t-1}^{(0)-1}}$. If the r.h.s. of (7.37) is neither well defined nor positive definite, the resulting expression on the left defines no covariance matrix such that a forecasting rule (7.37) which is perfect for all times t may fail to exist. Hence, additional requirements which guarantee that (7.37) is well defined for all periods are needed.

There are two special cases in which perfect forecasting rules for second moments exists. In the first one, all mediators agree upon subjective second moments. In the second case, all mediators $i > 0$ believe in constant covariance matrices. Lemma 7.1 establishes existence of perfect forecasting rules for second moments in the first case, Lemma 7.2 in the second case. Both proofs are found in the mathematical appendix, Section 7.8.

Lemma 7.1. *Under the hypotheses of Proposition 7.1, suppose* $V_t^{(i)} \equiv V_t^{(0)}$ *for all* $i = 1, \ldots, I$ *and all times* t. *Then the forecasting rule (7.37) takes the form*

$$V_t^{(0)} = \left(R\sum_{h=1}^{H}\frac{\overline{\eta}^{(h)}}{\alpha^{(h)}}\right)\sqrt{\Lambda_{t-1}^{-1}}\sqrt{\left(\sqrt{\Lambda_{t-1}}\,V_{t-1}^{(0)}\,\sqrt{\Lambda_{t-1}}\right)}\sqrt{\Lambda_{t-1}^{-1}}.$$

If, in addition, $\mathbb{V}_{t-1}[\xi_t] \equiv \Lambda$ *is constant over time, then the constant rule*

$$V_t^{(0)} \equiv \left(R\sum_{h=1}^{H}\frac{\overline{\eta}^{(h)}}{\alpha^{(h)}}\right)^2 \Lambda^{-1} \quad \textit{for all } \; t \in \mathbb{N}$$

is perfect for second moments as well.

Lemma 7.2. *Under the hypotheses of Assumption 7.1, let* $\mathbb{V}_{t-1}[\xi_t] \equiv \sigma_\xi^2 I_K$ *with* $\sigma_\xi > 0$. *Assume that the following hypotheses are satisfied.*

(i) All mediators $i > 0$ *use constant covariance matrices* $V_t^{(i)} \equiv V^{(i)}$ *such that* $V^{(i)-1} = O^\top \operatorname{diag}(\lambda^{(i1)}, \ldots, \lambda^{(iK)})\ O$, $i = 1, \ldots, I$ *for some orthogonal* $K \times K$ *matrix* O.
(ii) The eigenvalues $\lambda^{(i1)}, \ldots, \lambda^{(iK)}$, $i = 1, \ldots, I$ *given in* (i) *satisfy*

$$\left(\sum_{h=1}^{H} \frac{\overline{\eta}^{(h)}}{\alpha^{(h)}} \right) \sqrt{\sum_{i=1}^{I} \lambda^{(ik)}} < \frac{\sigma_\xi}{2R}, \quad k = 1, \ldots, K.$$

(iii) The initial forecast is $V_0^{(0)} := O^\top \operatorname{diag}(\lambda_0^{(01)-1}, \ldots, \lambda_0^{(0K)-1})O$, *where* O *is given in* (i) *and* $\lambda_0^{(01)}, \ldots, \lambda_0^{(0K)} > 0$.

If the initial eigenvalues $\lambda_0^{(01)}, \ldots, \lambda_0^{(0K)}$ *are suitably large, then the forecasting rule (7.37) is perfect for second moments whenever* $a_t^{(0)} > 0$.

Observe that perfect forecasting rules for second moments are not necessarily uniquely determined. This is illustrated by Lemma 7.1. Mathematically this is due to the natural appearance of multiplicities in matrix polynomial equations, cf. Gohberg, Lancaster & Rodman (1982). Of course, this multiplicity will already appear in the one-asset case $K = 1$ when (7.34) reduces to a scalar quadratic equation.

Setting aside the existence problem, economically the main informational constraint for applying the forecasting rule (7.37) is again the fact that neither the market shares of a mediator nor the subjective covariances matrices of the other mediators are observable quantities. We will show in Section 7.7 how this missing information may be estimated from historical data, such that a suitable approximation of (7.37) may be used whenever it is well defined.

7.6 Dynamics Under Rational Expectations

In this section we discuss the dynamics of asset prices when, under the hypotheses of Section 7.5, mediator 0 has rational expectations in the sense that first moments of cum-dividend prices are correctly predicted for all times t and second moments whenever possible. Assuming that each mediator $i > 0$ uses some forecasting rules $\psi^{(i)}$ and $\varphi^{(i)}$, the prices under rational expectations for group 0 are determined by

$$\begin{cases} q_t &= q_{t-1}^{(0)} + \frac{1}{R} A_t(\xi_t - \mathbb{E}_{t-1}[\xi_t]), \\ q_t^{(i)} &= \psi^{(i)}(q_{t-1}, \dots, q_{t-L}), \quad i = 1, \dots, I, \\ V_t^{(i)} &= \varphi^{(i)}(q_{t-1}, \dots, q_{t-L}), \quad i = 1, \dots, I, \\ q_t^{(0)} &= \psi^{(0)}\big(\mathbb{E}_{t-1}[\xi_t], d_t, (a_t^{(i)}, q_t^{(i)}, V_t^{(i)})_{i=1}^I, q_{t-1}^{(0)}\big), \\ V_t^{(0)} &= \varphi^{(0)}\big(\mathbb{V}_{t-1}[\xi_t], (a_t^{(i)}, V_t^{(i)})_{i=1}^I, V_{t-1}^{(0)}\big), \end{cases} \tag{7.38}$$

where $\varphi^{(0)}$ coincides with (7.37) whenever (7.37) is well defined. The cum-dividend price q_t in (7.38) is essentially determined by the forecasting rule $\psi^{(0)}$ of mediator 0 which takes the form

$$\begin{aligned} q_t^{(0)} = R\left[I_K + \textstyle\sum_{i=0}^I \frac{a_t^{(i)}}{a_t^{(0)}} V_t^{(0)} V_t^{(i)-1}\right] q_{t-1}^{(0)} - \textstyle\sum_{i=1}^I \frac{a_t^{(i)}}{a_t^{(0)}} V_t^{(0)} V_t^{(i)-1} q_t^{(i)} \\ - a_t^{(0)-1} V_t^{(0)} \left[R \textstyle\sum_{i=0}^I a_t^{(i)} V_t^{(i)-1} d_t + \mathbb{E}_{t-1}[\xi_t] - \bar{x}\right]. \end{aligned} \tag{7.39}$$

The map (7.38) together with the logit functions (7.27) with (7.21) defines a time-one map of a dynamical system in a random environment which describes the evolution of prices. The map (7.38) will, in general, be non-linear because nonlinearities may enter the system through forecasting rules $\varphi^{(i)}$ and $\psi^{(i)}$, $i > 0$. Since time appears as an explicit argument in the discrete choice model, the map (7.38) is not a random dynamical system in the strict sense of Arnold (1998).

It is beyond the scope of this chapter to conduct a stochastic analysis of the system (7.38) in the spirit of Föllmer, Horst & Kirman (2005) and Horst (2005) or to present a dynamic investigation paralleling numerous works initiated by Brock & Hommes (1998). The interested reader is referred to Horst & Wenzelburger (2005). For the remainder of this section, we instead discuss two cases of a stochastic linear benchmark dynamics. Assume for this purpose that market shares are constant over time, that is $a_t^{(i)} \equiv a^{(i)}$, $i = 0, 1$. This case, for instance, obtains if the intensity of choice parameters in (7.27) are all set to zero and no household switches between mediators.[5]

Case I. Assume that there are only two mediators $i = 0, 1$. Let mediator 1 be a chartist who uses the simple *technical trading rule*

$$q_t^{(1)} = \psi^{(1)}(q_{t-1}, \dots, q_{t-L}) := \sum_{l=1}^{L} D^{(l)} q_{t-l} \tag{7.40}$$

as a forecasting rule, where $D^{(l)} = \mathrm{diag}(\delta_1^{(l)}, \dots, \delta_K^{(l)})$, $l = 1, \dots, L$ are diagonal matrices whose non-zero entries are weighted trends of asset prices,

[5] Numerical simulations of the model indicate that the market shares settle relatively fast to constant values, at least for sufficiently small parameters of the intensity of choice. As switching becomes negligible in the long run, the system then is asymptotically linear, cf. Böhm & Wenzelburger (2005).

respectively. In addition, let all subjective covariance matrices be constant over time, i.e. $V_t^{(i)} \equiv V^{(i)}$, $i = 0, 1$. Then, under the hypotheses of Lemma 7.2, mediator 0 may have correct second moments which are constant over time, since no switching occurs. If $Q_t = (q_t^{(0)}, q_t, \ldots, q_{t+1-L}) \in \mathbb{R}_+^{K(L+1)}$ denotes the vector consisting of the last price forecast of mediator 0 and past L realized prices, then (7.38) defines an ARMAX process

$$Q_t = \mathbf{A} Q_{t-1} + \mathbf{b}_t \tag{7.41}$$

with a constant coefficient block matrix $\mathbf{A}$ and a vector-valued random variable $\mathbf{b}_t$. More precisely, setting

$$\mathbf{A}_0 := R(I_K + \tfrac{a^{(1)}}{a^{(0)}} V^{(0)} V^{(1)-1}), \quad \mathbf{A}_l := \tfrac{-a^{(1)}}{a^{(0)}} V^{(0)} V^{(1)-1} D^{(l)}, \quad l = 1, \ldots, L,$$

we have

$$\mathbf{A} := \begin{pmatrix} \mathbf{A}_0 & \cdots & \cdots & \mathbf{A}_L \\ I_K & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I_K & 0 \end{pmatrix}, \tag{7.42}$$

whereas

$$\mathbf{b}_t = \begin{pmatrix} \frac{1}{a^{(0)}} V^{(0)} \left[\bar{x} - \mathbb{E}_{t-1}[\xi_t] - R[a^{(0)} V^{(0)-1} + a^{(1)} V^{(1)-1}] \, d_t \right] \\ \frac{1}{R} \left[a^{(0)} V^{(0)-1} + a^{(1)} V^{(1)-1} \right]^{-1} (\xi_t - \mathbb{E}_{t-1}[\xi_t]) \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{7.43}$$

In view of Chapter 4, assume that each of the two process $\{\xi_t\}_{t \in \mathbb{N}}$ and $\{d_t\}_{t \in \mathbb{N}}$ can be extended to processes on $\mathbb{Z}$ and represented by an ergodic metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t \in \mathbb{Z}})$. Then the stochastic process (7.43) can be written as $\mathbf{b}_t(\omega) = \mathbf{b}(\vartheta^t \omega)$ with some measurable map $\mathbf{b} : \Omega \to \mathbb{R}^{K(L+1)}$ and (7.41) becomes an *affine random dynamical system* in the sense of Arnold (1998, Chap. 5). A stationary solution to the stochastic difference equation (7.41) is given by a so-called[6] *random fixed point* $Q^\star(\omega)$ which is defined by the random variable

$$Q^\star(\omega) = \sum_{s=0}^{\infty} \mathbf{A}^s \mathbf{b}(\vartheta^{-(s+1)} \omega), \quad \omega \in \Omega. \tag{7.44}$$

[6] In classical terminology, a random fixed point of (7.41) corresponds to a *'steady state solution'* of a linear stochastic system, cf. Hannan & Deistler (1988, Chap. 1). An asymptotically stable random fixed point with rational expectations is also a special case of a stochastic consistent expectations equilibria in the sense of Hommes, Sorger & Wagner (2004), see Chapter 4, Section 4.4.

Under certain conditions given in the next proposition, the random fixed point $Q^\star(\omega)$ is said to be globally asymptotically stable, implying that $\mathbb{P}$-almost all solutions to equation (7.41) will eventually behave like the stationary stochastic process $\{Q^\star(\vartheta^t\omega)\}_{t\in\mathbb{N}}$ induced by (7.44). Here, the stationarity follows from the stationarity assumption for the exogenous noise processes.

Proposition 7.2. *Assume that the following hypotheses are satisfied.*

(i) $\{\xi_t\}_{t\in\mathbb{N}}$ *and* $\{d_t\}_{t\in\mathbb{N}}$ *can be represented by an ergodic metric dynamical system* $(\Omega, \mathcal{F}, \mathbb{P}, \{\vartheta^t\}_{t\in\mathbb{Z}})$*, where* $\mathbb{E}[\xi_t] \equiv \overline{\xi} \in \mathbb{R}^K$ *and* $\mathbb{E}[d_t] \equiv \overline{d} \in \mathbb{R}^K_+$.
(ii) $V_t^{(i)} \equiv V^{(i)}$ *and* $a_t^{(i)} \equiv a^{(i)} > 0$*,* $i = 0, 1$.
(iii) $\det(\mathbf{A} - \lambda\mathbf{I}) \neq 0$ *for all* $\lambda \in \mathbb{C}$ *with* $|\lambda| \geq 1$*, that is, all eigenvalues of* $\mathbf{A}$ *given in (7.42) lie inside the unit disk.*

Then the random fixed point $Q^\star(\omega)$ *defined in (7.44) is well defined, uniquely determined, and globally asymptotically stable. The mean price level* $\overline{q} \in \mathbb{R}^K_+$ *is the uniquely determined solution of the linear equation*

$$(\mathbf{I} - \mathbf{A})\overline{Q} = \mathbb{E}[\mathbf{b}],$$

where $\overline{Q} := (\overline{q}, \ldots, \overline{q}) \in \mathbb{R}^{K(L+1)}$.

Proof. The stability property follows from Arnold (1998, Corollary 5.6.6). The second statement is due to the fact that $\mathbf{I} - \mathbf{A}$ is invertible such that $(\mathbf{I} - \mathbf{A})\mathbb{E}[Q^\star] = \mathbb{E}[\mathbf{b}]$ has a unique solution of the form $\mathbb{E}[Q^\star] = (\overline{q}, \ldots, \overline{q})$.

Case II. Another special case arises when the two mediators agree on the subjective covariance matrices, such that $V_t^{(0)} \equiv V_t^{(1)}$ for all times t. This implies that all $A_t^{(i)}$, $i = 0, 1$ are time-invariant diagonal matrices and the map (7.39) takes the form

$$\begin{aligned} q_t^{(0)} = R\left(1 + \tfrac{a^{(1)}}{a^{(0)}}\right) q_{t-1}^{(0)} - \tfrac{a^{(1)}}{a^{(0)}} \textstyle\sum_{l=1}^{L} D^{(l)} q_{t-l} \\ + a^{(0)-1} V_t^{(0)}(\bar{x} - \mathbb{E}_{t-1}[\xi_t]) - R\left(1 + \tfrac{a^{(1)}}{a^{(0)}}\right) d_t. \end{aligned} \tag{7.45}$$

Since all $D^{(l)}$, $l = 1, \ldots, L$ are diagonal matrices, (7.45) decouples into K scalar equations which are linked by an inhomogeneous term via $V_t^{(0)}$. The stability properties of (7.38) are now obtained from an analysis of each scalar equation for each asset $k = 1, \ldots, K$.

Proposition 7.3. *Assume that the following hypotheses are satisfied.*

(i) Hypothesis (*i*) *of Proposition 7.2 holds, where, in addition,* $\mathbb{V}_{t-1}[\xi_t] \equiv \Lambda$ *is constant over time.*
(ii) $a_t^{(i)} \equiv a^{(i)} > 0$*,* $i = 0, 1$ *and* $V_t^{(0)} \equiv V_t^{(1)}$ *for all* $t \in \mathbb{N}$*. Moreover, each* $V_t^{(0)}$*,* $t \in \mathbb{N}$ *is obtained from a constant perfect forecasting rule for second moments as given by Lemma 7.1, such that* $V_t^{(0)} \equiv \left(R \sum_{h=1}^{H} \frac{\overline{\eta}^{(h)}}{\alpha^{(h)}}\right)^2 \Lambda^{-1}$.

(iii) Let $1 \leq k \leq K$ be arbitrary but fixed. Assume that all zeros of the characteristic polynomial χ_k associated with (7.45), which is given by

$$\chi_k(\lambda) = \lambda^{L+1} - R\left(1 + \tfrac{a^{(1)}}{a^{(0)}}\right)\lambda^L + \tfrac{a^{(1)}}{a^{(0)}} \sum_{l=1}^{L} \delta_k^{(l)} \lambda^{L-l},$$

lie inside the unit circle.

Then the price process for the k-th asset admits a globally asymptotically stable random fixed point. The mean price level of the k-th asset $\overline{q}_k$ is given by

$$\overline{q}_k = \frac{\frac{y_k}{a^{(0)}+a^{(1)}} - R\overline{d}_k}{\frac{a^{(1)}}{a^{(0)}+a^{(1)}}\left(\sum_{l=1}^{L} \delta_k^{(l)} - 1\right) - R - 1},$$

where y_k is the k-th entry of the vector $y = \left(R\sum_{h=1}^{H} \frac{\overline{\eta}^{(h)}}{\alpha^{(h)}}\right)^2 \Lambda^{-1}(\bar{x} - \overline{\xi}) \in \mathbb{R}^K$ and $\overline{d}_k$ is the mean value of the dividend payment of the k-th asset.

The proof of Proposition 7.3 is analogous to that of Proposition 7.2. Since the sum $a^{(0)} + a^{(1)} = \sum_{h=1}^{H} \frac{\overline{\eta}^{(h)}}{\alpha^{(h)}}$ is constant, Proposition 7.3 shows in particular that an asset price decreases on average with an increasing market share of the chartists $a^{(1)}$. From Jury's test (see Elaydi 1996, p. 181) it is known that a necessary condition for the zeros of χ_k to lie inside the unit circle is

$$\chi_k(1) > 0 \qquad \text{and} \qquad (-1)^{L+1}\chi_k(-1) > 0.$$

If $\delta_k^{(1)} = \ldots = \delta_k^{(L)}$ are equal to some δ_k, then a routine calculation shows that this is the case precisely when

$$\tfrac{a^{(1)}}{a^{(0)}}(L\delta_k - R) > R - 1.$$

Both of the above cases describe scenarios with a stable price process along which mediator 0 has rational expectations in the above sense, whereas the beliefs of mediator 1, in general, will be non-rational and, in fact, could be arbitrarily wrong. As soon as the coefficient matrix $\mathbf{A}$ has eigenvalues outside the unit disk, the system (7.41) is unstable and bubble paths will emerge. It follows immediately from Böhm & Chiarella (2005, Thm. 3.2) that such an unstable situation occurs for $R > 1$ and $a^{(1)}$ sufficiently close to 0. The somewhat puzzling conjecture for the general nonlinear case is that the presence of sufficiently many non-rational mediators such as chartists may have a stabilizing effect on the price process. Numerical simulations of the model with varying group sizes of households seem to indicate that for a robust set of parameters, the chartists have non-zero market shares, cf. Böhm & Wenzelburger (2005) and Horst & Wenzelburger (2005).

7.7 Adaptive Learning With Heterogeneous Beliefs

The discussion in Section 7.5 revealed that in order to apply perfect forecasting rules, mediator 0 has to know the excess demand function of all market participants. In the general case, this function will be nonlinear and its estimation requires nonparametric methods provided in Chapter 5. However, in the simple case in which the non-rational mediators $i > 0$ use simple linear forecasting rules, the estimation problem reduces to a linear one, provided that the switching behavior of households is correctly anticipated. We will first outline the general idea and then return to the particularly simple example with two mediators.

7.7.1 The General Idea

In this section, we introduce a general principle of estimating an unbiased forecasting rule from historical data. Consider to this end the aggregate (per-capita) excess demand function (7.46) of all mediators except mediator 0. Using the notation of Section 7.1, let

$$\Phi\big(\xi_t, (\eta_t^{(i)}, q_t^{(i)}, V_t^{(i)})_{i=1}^I, p\big) := \sum_{i=1}^{I} \Phi^{(i)}(\eta_t^{(i)}, q_t^{(i)}, V_t^{(i)}, p) + \xi_t - \bar{x} \qquad (7.46)$$

denote the aggregate excess demand of all mediators except mediator 0. Assume that each mediator $i > 0$ uses forecasting rules $\psi^{(i)}$ and $\varphi^{(i)}$ of the functional form

$$q_t^{(i)} = \psi^{(i)}(q_{t-1}, \dots, q_{t-L}) \quad \text{and} \quad V_t^{(i)} = \varphi^{(i)}(q_{t-1}, \dots, q_{t-L}).$$

Inserting these forecasting rules into (7.46) and suppressing the respective functional expressions for notational convenience, the excess demand (7.46) takes the form

$$\Phi\big(\xi_t, (\eta_t^{(i)})_{i=1}^I, q_{t-1}, \dots, q_{t-L}, p\big). \qquad (7.47)$$

Let $x_t^{(0)}$ denote the portfolio held by mediator 0 after trading in period t. Then the market-clearing condition (7.4) implies that

$$-x_t^{(0)} = \Phi\big(\xi_t, (\eta_t^{(i)})_{i=1}^I, q_{t-1}, \dots, q_{t-L}, p_t\big) \quad \text{for each} \quad t \in \mathbb{N}. \qquad (7.48)$$

Observe that the prices and the portfolios $x_t^{(0)}$ appearing in (7.48) are observable quantities, whereas the market shares $(\eta_t^{(i)})_{i=1}^I$ are not observable for mediator 0. Moreover, the concrete functional form (7.47) including the involved forecasting rules are unknown to mediator 0.

Assume for a moment that a good approximation for the discrete choice model (7.27) with unknown intensity of choice parameters has been developed along with a model for the exogenous noise process $\{\xi_t\}_{t\in\mathbb{N}}$. Suppose that these models yield sufficiently good estimates $\hat{\eta}_t^{(i)}$ for $\eta_t^{(i)}$, $i = 1, \dots, I$ and $\hat{\xi}_t$ for ξ_t, respectively. Then the following adaptive estimation procedure may be applied. At an arbitrary date t, first estimate the functional relationship (7.48) from past observed or estimated data $\{x_s^{(0)}, \xi_s, (\hat{\eta}_s^{(i)})_{i=1}^I, p_s, q_s\}_{s=0}^{t-1}$. This yields an approximation $\widehat{\Phi}_t$ of (7.47) such that

$$\widehat{\Phi}_t\big(\hat{\xi}_t, (\hat{\eta}_t^{(i)})_{i=1}^I, q_{t-1}, \dots, q_{t-L}, p\big) \tag{7.49}$$

describes the approximate excess demand function for assets in period t. Second, let $(q_t^{(0)}, V_t^{(0)})$ denote the beliefs of mediator 0 as before. Replacing Φ in (7.48) with (7.49) and solving for the ex-dividend price, we would get an approximate temporary equilibrium map

$$\hat{p}_t = \widehat{G}_t\big(\hat{\xi}_t, (\hat{\eta}_t^{(i)})_{i=1}^I, q_{t-1}, \dots, q_{t-L}, q_t^{(0)}, V_t^{(0)}\big), \tag{7.50}$$

provided that a solution exists. Third, in view of (7.8), taking expectations the condition for $q_{t-1}^{(0)}$ being unbiased is

$$\mathbb{E}_{t-1}\Big[\widehat{G}_t\big(\hat{\xi}_t, (\hat{\eta}_t^{(i)})_{i=1}^I, q_{t-1}, \dots, q_{t-L}, q_t^{(0)}, V_t^{(0)}\big)\Big] + d_t \stackrel{!}{=} q_{t-1}^{(0)}. \tag{7.51}$$

Hence any solution $q_t^{(0)}$ to (7.51) defines a approximation of the unbiased forecasting rule (7.9) in period t.

Up to the two invertibility conditions in (7.50) and (7.51) which may be violated during the estimation process, the forecasting scheme described in this section can, in principle, be carried out at each point in time using the estimation techniques of Chapter 5. The next section demonstrates the success of such an approach in the case with linear mean-variance preferences and linear forecasting rules.

7.7.2 The Case With Two Mediators

Consider now the case of two mediators treated in Section 7.6 in which households switch between mediators according to (7.27). The proposed learning scheme now takes a particularly simple form. Assume to this end that the estimates of the market shares are sufficiently good such that we can abstract from prediction errors of the distribution of households. This assumption seems not to be too strong because the sum $\eta_t^{(h0)} + \eta_t^{(h1)} = \overline{\eta}^{(h)}$ is constant over time for all $1 \le h \le H$ and each $\eta_t^{(h0)}$ is known to mediator 0. On the other hand, for the unknown choice intensities in (7.27) there are maximum-likelihood estimators available which are consistent under standard regularity assumptions. In the present case of our discrete choice model of the Logit type, these estimators are found in Judge, Griffiths, Hill, Lütkepohl & Lee (1985, Chap. 18, p. 765). The estimators for the choice intensities yield estimates $\hat{\eta}_t^{(h1)}$, $h = 1, \dots, H$ for the fractions of households $\eta_t^{(h1)}$, $h = 1, \dots, H$ employing mediator 1. The informational constraints faced by mediator 0 are now summarized as follows.

Assumption 7.2. *The information of mediator* 0 *encompasses the following:*

(i) The realized returns $R_t^{(1)}$ *of mediator* 1 *as given by (7.25) are observable quantities; a discrete choice model for the switching behavior of households has been estimated, consistent estimates of the fraction of households are known, such that, for simplicity of exposition, prediction errors are neglected and* $\hat{\eta}_t^{(h1)} = \eta_t^{(h1)}$, $h = 1, \ldots, H$.

(ii) The total amount of retradeable assets per capita $\bar{x}$ *is known;*

(iii) The correct functional form *of mediator* 1*'s forecasting rules is known; these consist of a simple technical trading rule of the form (7.40) and a constant subjective covariance matrix.*

To specify the noise traders, we impose the following so-called *identifiability assumption* (see Chapter 3, Assumption 3.1) which governs the transactions of the noise traders.

Assumption 7.3. *The noise-trader behavior is driven by an* MA(N)*-process* $\{\xi_t\}_{t\in\mathbb{N}}$*, given by*

$$\xi_t = \epsilon_t + \sum_{n=1}^{N} C^{(n)}\, \epsilon_{t-n},$$

where $\{\epsilon_t\}_{t\in\mathbb{N}}$ *is an* unobservable *iid process on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *with zero means* $\mathbb{E}[\epsilon_t] = 0$ *which satisfies the following:*

(1) The process $\{\epsilon_t\}_{t\in\mathbb{N}}$ *is bounded and, in particular,* $\sup_{t\geq 1} \mathbb{E}\left[\|\epsilon_t\|^\gamma\right] < \infty$ *for some* $\gamma > 2$.

(2) $C^{(1)}, \ldots, C^{(N)}$, $N \geq 0$ *are* $K \times K$ *matrices and the matrix polynomial* $\Gamma(z) = I_K + C^{(1)}z + \cdots + C^{(N)}z^N$, $z \in \mathbb{C}$ *is* strictly positive real*, that is,*

(i) $\det \Gamma(z) \neq 0$ *for all* $z \in \mathbb{C}$ *with* $|z| \leq 1$ *and*

(ii) for each $z \in \mathbb{C}$ *with* $|z| \leq 1$*, the matrix* $\Gamma(z) + \Gamma^\top(\bar{z})$*, with* $\bar{z} \in \mathbb{C}$ *denoting the complex conjugate of* z*, is strictly positive definite such that*

$$\zeta^\top \big(\Gamma(z) + \Gamma^\top(\bar{z})\big)\zeta > 0 \quad \textit{for all}\ \ 0 \neq \zeta \in \mathbb{R}^K.$$

Adopting Assumption 7.3, we have

$$\mathbb{E}_{t-1}[\xi_t] = \sum_{n=1}^{N} C^{(n)} \epsilon_{t-n} \quad \text{and} \quad \mathbb{V}_{t-1}[\xi_t] = \mathbb{V}[\epsilon_t].$$

Notice, moreover, that Assumption 7.3 includes the case in which $\{\xi_t\}_{t\in\mathbb{N}}$ is iid (Set $N = 0$).

By Assumption 7.2, mediator 0 is left to estimate the excess demand (7.49) of mediator 1. A major prerequisite for the success of the estimation is Assumption 7.3 which presumes that the noise trader behavior is not too arbitrary. Clearly, there is no economic justification why noise trader should behave according to this presumption other than without it, our learning scheme may fail to converge.[7] By Assumption 7.2, mediator 0 has correct knowledge about the functional form of the forecasting rule (7.40) used by mediator 1 but not its concrete parameterization. Set $x_t = \bar{x} - x_t^{(0)}$ to adjust mediator 0's portfolio for the total amount of available assets. Then the market-clearing condition (7.48) for period t reads

$$x_t = \sum_{h=1}^{H} \frac{\eta_t^{(h1)}}{\alpha^{(h)}} V^{(1)-1} \Big[\sum_{l=1}^{L} D^{(l)} q_{t-l} - R p_t \Big] + \epsilon_t + \sum_{n=1}^{N} C^{(n)} \epsilon_{t-n}. \tag{7.52}$$

Setting

$$\theta = (A^{(11)}, \ldots, A^{(HL)}, B^{(1)}, \ldots, B^{(H)}, C^{(1)}, \ldots, C^{(N)}) \tag{7.53}$$

with

$$A^{(hl)} = \tfrac{1}{\alpha^{(h)}} V^{(1)-1} D^{(l)}, \quad B^{(h)} = \tfrac{-R}{\alpha^{(h)}} V^{(1)-1} \tag{7.54}$$

for all $h = 1, \ldots, H$, $l = 1, \ldots, L$ and

$$y_t = \left(\eta_t^{(11)} q_{t-1}^\top, \ldots, \eta_t^{(H1)} q_{t-L}^\top, \eta_t^{(11)} p_t^\top, \ldots, \eta_t^{(H1)} p_t^\top, \epsilon_{t-1}^\top, \ldots, \epsilon_{t-N}^\top \right)^\top, \tag{7.55}$$

the market-clearing condition (7.52) may be rewritten as

$$x_t = \theta\, y_t + \epsilon_t. \tag{7.56}$$

We use the representation (7.56) to estimate the unknown coefficient matrix θ by applying the *approximate-maximum-likelihood (AML)* algorithm introduced in Chapter 3, Section 3.3. This recursive scheme generates the successive estimates

$$\hat{\theta}_t = (\hat{A}_t^{(11)}, \ldots, \hat{A}_t^{(HL)}, \hat{B}_t^{(1)}, \ldots, \hat{B}_t^{(H)}, \hat{C}_t^{(1)}, \ldots, \hat{C}_t^{(N)}), \tag{7.57}$$

for θ based upon information available at date t. Since the noise process $\{\epsilon_t\}_{t \in \mathbb{N}}$ is assumed to be unobservable, for each $t \in \mathbb{N}$ set

$$\begin{aligned} \hat{\epsilon}_t &:= x_t - \hat{\theta}_t \hat{y}_t, \\ \hat{y}_t &:= \left(\eta_t^{(11)} q_{t-1}^\top, \ldots, \eta_t^{(H1)} q_{t-L}^\top, \eta_t^{(11)} p_t^\top, \ldots, \eta_t^{(H1)} p_t^\top, \hat{\epsilon}_{t-1}^\top, \ldots, \hat{\epsilon}_{t-N}^\top \right)^\top, \end{aligned}$$

where the $\hat{\epsilon}_t$ are also called *a posteriori* prediction errors. The AML algorithm [8] is recursively defined by

---

[7]As already pointed out in Chapter 3, identifiability assumptions as stated in Assumption 7.3 are fundamental in the systems and control literature. Ljung, Söderström & Gustavsson (1975) show that without such identifiability assumptions, identification methods such as ordinary least squares may fail.

[8]We adjusted the dating according to our needs and made use of the matrix inversion lemma (e.g., see Caines 1988).

$$\begin{cases} \hat{\theta}_{t+1} = \hat{\theta}_t + (x_t - \hat{\theta}_t \hat{y}_t)\hat{y}_t^\top P_t, \\ P_t \quad = P_{t-1} - P_{t-1}(\hat{y}_t \hat{y}_t^\top) P_{t-1} \, (1 + \hat{y}_t^\top P_{t-1} \hat{y}_t)^{-1}, \end{cases} \tag{7.58}$$

for arbitrary initial conditions P_{-1} and $\hat{\theta}_0$. The estimated coefficients (7.57) obtained from applying (7.58) can now be used to construct an approximation of the perfect forecasting rules defined in (7.32) and (7.37) based on information available at date t. An approximation of (7.32) is defined by

$$\begin{aligned} q_t^{(0)} &= \widehat{\psi}_t^{(0)}\big((q_{t-1}^{(0)} - d_t), V_t^{(0)}; (\eta_t^{(h1)})_{h=1}^H, \hat{\theta}_t\big) \\ &:= R(q_{t-1}^{(0)} - d_t) \; - \; a_t^{(0)-1} V_t^{(0)} \hat{x}_t, \end{aligned} \tag{7.59}$$

where

$$\hat{x}_t = \sum_{h=1}^{H} \sum_{l=1}^{L} \eta_t^{(h1)} \hat{A}_t^{(hl)} \, q_{t-l} + \sum_{h=1}^{H} \eta_t^{(h1)} \hat{B}_t^{(h)} \, (q_{t-1}^{(0)} - d_t) + \sum_{n=1}^{N} \hat{C}_t^{(n)} \, \hat{\epsilon}_{t-n} \; - \; \bar{x}$$

denotes the estimated excess demand of shares in period t corresponding to (7.49). Observe that $\hat{\xi}_t := \sum_{n=1}^{N} \hat{C}_t^{(n)} \, \hat{\epsilon}_{t-n}$ is an approximation for the expected transactions of the noise traders $\mathbb{E}_{t-1}[\xi_t]$.

In order to obtain an approximation of the perfect forecasting rule for second moments (7.37), assume that $\mathbb{V}_{t-1}[\xi_t] \equiv \sigma_\xi I_K$ with known σ_ξ.[9] It follows from (7.54) that $\frac{1}{R} \sum_{h=1}^{H} \eta_t^{(h1)} \hat{B}_t^{(h)}$ is an approximation for $a_t^{(1)} V^{(1)-1}$. Setting

$$\widetilde{V}_t^{(0)} := R a_t^{(0)} \left(\sigma_\xi \sqrt{V_{t-1}^{(0)-1}} - \sum_{h=1}^{H} \eta_t^{(h1)} \hat{B}_t^{(h)} \right)^{-1},$$

an approximation of (7.37) is then given by

$$\begin{aligned} V_t^{(0)} \; &= \; \widehat{\varphi}_t^{(0)}\big(V_{t-1}^{(0)}; (\eta_t^{(h1)})_{h=1}^H, \hat{\theta}_t\big) \\ &:= \begin{cases} \widetilde{V}_t^{(0)-1} & \text{if } \widetilde{V}_t^{(0)} \text{ is positive definite,} \\ V_{t-1}^{(0)} & \text{otherwise.} \end{cases} \end{aligned} \tag{7.60}$$

The series of forecasts (7.59) and (7.60) generated by (7.58) defines a AML-based adaptive learning scheme for mediator 0. Together with the chartist's forecasts (7.40), the temporary equilibrium map (7.23), and the logit function (7.27), then prices and forecasts under this learning scheme evolve time according to the following set of equations:

[9] This assumption is made for simplicity. Since noise traders' transactions are observable, estimations of $\mathbb{V}_{t-1}[\xi_t]$ could, in principle, be obtained by standard methods.

$$\begin{cases} q_t &= G(\xi_t, (a_t^{(i)}, q_t^{(i)}, V_t^{(i)})_{i=0}^1) + d_t, \\ V_t^{(1)} &= V_0^{(1)}, \\ q_t^{(1)} &= \sum_{l=1}^L D^{(l)}\, q_{t-l}, \\ V_t^{(0)} &= \widehat{\varphi}_t^{(0)}\big(V_{t-1}^{(0)}; (\eta_t^{(h1)})_{h=1}^H, \hat{\theta}_t\big), \\ q_t^{(0)} &= \widehat{\psi}_t^{(0)}\big((q_{t-1}^{(0)} - d_t), V_t^{(0)}; (\eta_t^{(h1)})_{h=1}^H, \hat{\theta}_t\big), \\ \hat{\theta}_{t+1} &= \hat{\theta}_t + (x_t - \hat{\theta}_t \hat{y}_t)\hat{y}_t^\top P_t, \\ P_t &= P_{t-1} - P_{t-1}(\hat{y}_t\hat{y}_t^\top)P_{t-1}\,(1 + \hat{y}_t^\top P_{t-1}\hat{y}_t)^{-1}. \end{cases} \tag{7.61}$$

The proof that the learning scheme converges to rational expectations for mediator 0 in the sense that first and second moments are correct along each orbit of the system depends on whether or not the AML algorithm (7.58) generates strongly consistent estimates, that is, estimates $\hat{\theta}_t$ which converge $\mathbb{P}$-a.s. to θ. Observe that the requirement $\mathbb{V}_{t-1}[\xi_t] \equiv \sigma_\xi I_K$ is needed only for the approximation of the forecasting rule for second moments (7.37). Moreover, the adaptive learning scheme (7.58) is meaningful only if the dynamics under rational expectations for mediator 0 as well as under the scheme itself is stable. Before we state conditions under which strongly consistent estimates obtain, recall that $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ denote the minimal and maximal eigenvalue of a symmetric matrix B, respectively.

Theorem 7.2. *Assume that the following hypotheses are satisfied.*

(i) The noise process $\{\xi_t\}_{t\in\mathbb{N}}$ satisfies Assumption 7.3 with known $\mathbb{V}_{t-1}[\xi_t] = \sigma_\xi I_K$ and the dividend process $\{d_t\}_{t\in\mathbb{N}}$ satisfies Assumption 7.1 (ii).
(ii) Mediator 1 *uses a linear rule (7.40) and constant covariance matrices as forecasting rules.*
(iii) The informational constraints of mediator 0 *satisfy Assumption 7.2.*
(iv) The price process (7.23) is stable under rational expectations for mediator 0 *as given in (7.38) and under the application of the AML-based learning scheme (7.58) for mediator* 0 *as given by (7.61), in the sense that for both cases the respective processes of of prices and forecast remain bounded, i.e.,*

$$\|(q_t, q_t^{(0)}, q_t^{(1)})\| < M \quad \mathbb{P} - a.s.\ for\ all\ \ t \in \mathbb{N}$$

for some constant M.

If the stochastic process $\{y_t\}_{t\in\mathbb{N}}$ defined in (7.55) satisfies the weak excitation condition

$$\frac{\log\left(e + \lambda_{\max}\left(\sum_{s=1}^t y_s y_s^\top\right)\right)}{\lambda_{\min}\left(\sum_{s=1}^t y_s y_s^\top\right)} \to 0 \quad \mathbb{P} - a.s.\ as\ t \to \infty, \tag{7.62}$$

then the sequence of estimates $\{\hat{\theta}_t\}_{t\in\mathbb{N}}$ *generated by the AML algorithm (7.58) is strongly consistent, that is*

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

The proof follows directly from Chapter 3, Proposition 3.3. As in Chapter 3, the result is independent of initial conditions as long as the system is stable under the learning scheme. Combining the results of Chapter 3 with Proposition 3.4 in Horst & Wenzelburger (2005), it is intuitively clear how the forecasts could be 'censored' in such a way that the system remains stable under learning. The general idea of such a censoring was discussed in Chapter 5. The major obstacle for convergence is the *weak excitation condition* (7.62). In feedback systems such as (7.38), it is well-known that a violation of condition (7.62) may result in a failure of consistent estimates and thus in a failure of the learning scheme. However, as discussed in Chapter 3, this problem can be resolved by occasionally perturbing the forecasts such that (7.62) is satisfied.

Moreover, the success of the learning scheme depends essentially on the success of estimating the excess demand function. This estimation by means of the AML algorithm can only be successful if, apart from the switching behavior of households, the excess demand function is linear. Non-linear excess demand functions as, for example, in Brock & Hommes (1997b) will generally require non-linear estimation techniques. For the present model, this problem arises naturally as soon as chartists use empirical covariance matrices. However, the concept of estimating the excess demand readily carries over to the non-linear case. Then non-parametric techniques as presented in Chapter 5 can be applied to estimate the excess demand function.

Concluding Remarks

The analysis of adaptive learning in a financial market with heterogeneous traders showed that good forecasts hinge essentially on good estimates of the demand behavior of market participants. It was demonstrated that expectations should be based on estimates of the excess demand function of traders. The proposed learning scheme made use of the simple observation that the excess demand function is linear when households are characterized by linear mean-variance preferences and mediators by a simple linear forecasting technology. The remaining nonlinearity rests on the switching behavior of households for which consistent maximum likelihood estimators exist. In a special case with only one chartist, it was shown that this methodology leads to a learning scheme which converges under econometric assumptions that are standard in the literature on estimation and optimal control.

While situations in which an agent successfully learns to have rational expectations are likely to remain special cases, one should proceed with caution when assessing the plausibility of a learning scheme. As shown by Jungeilges

(2003), the underlying linearity hypothesis for applying the ordinary-least-squares learning scheme in nonlinear models may be rejected within finite time by means of techniques from linear statistics. Apart from the convergence issue, the main conclusion of this chapter is that a plausible learning scheme should take into account the correct trading structure of the market.

The potential of this line of research may be outlined as follows. First, the learning scheme should be extended to an algorithm which uses on-line estimations of the distributions of households instead of their true values. Second, non-parametric estimation techniques should be applied to identify forecasting rules for models with nonlinear excess demand functions. However, the biggest challenge is to abandon the assumption of price-taking behavior. The structure of an unbiased forecasting rule points at a possibility to exploit the knowledge of excess demand functions strategically. This will be carried out in Wenzelburger (2005).

7.8 Mathematical Appendix

Proof of Proposition 7.1. We preface the proof of the proposition by a technical lemma.

Lemma 7.3. *Let B, C be arbitrary symmetric positive definite matrices. Then*

$$D := \sqrt{B^{-1}} \left(\sqrt{(\sqrt{B}\, C \sqrt{B})} \right) \sqrt{B^{-1}}$$

is a symmetric positive definite matrix with $C = DBD$.

Proof. Since B is symmetric and positive definite, its square root $\sqrt{B}$, where $B = \sqrt{B}\sqrt{B}$, is again symmetric and positive definite and hence invertible with $(\sqrt{B})^{-1} = \sqrt{B^{-1}}$. It is straightforward to verify that D as defined above satisfies $C = DBD$. Since C, $\sqrt{B}$, and $\sqrt{B^{-1}}$ are all symmetric and positive definite, it is verified directly that D is symmetric and positive definite. □

By Lemma 7.3, Π_t as given by (7.36) is a solution to the equation

$$\Pi_t^{-1} \Lambda_{t-1} \Pi_t^{-1} = R^2 V_{t-1}^{(0)}.$$

If $\Pi_t - \sum_{i=1}^{I} a_t^{(i)} V_t^{(i)-1}$ is positive definite, then (7.37) is well defined and the assertion follows from (7.35). □

Proof of Lemma 7.1. The first statement follows from

$$A_t^{-1} = \left(\sum_{h=1}^{H} \frac{\overline{\eta}^{(h)}}{\alpha^{(h)}} \right) V_t^{(0)-1}$$

and Lemma 7.3 applied to (7.35). The second one is a special case with $V_t^{(0)} \equiv V_{t-1}^{(0)}$ for all times t.

Proof of Lemma 7.2. We seek a symmetric matrix $V_t^{(0)}$ which satisfies (7.35) and which stays positive definite for all times t. Under the given hypotheses, this is equivalent to the existence of a symmetric and positive matrix $V_t^{(0)}$ with

$$a_t^{(0)} V_t^{(0)-1} = \frac{\sigma_\xi}{R} \sqrt{V_{t-1}^{(0)-1}} - \sum_{i=1}^{I} a_t^{(i)} V_t^{(i)-1}. \tag{7.63}$$

Choose $V_0^{(0)}$ as in (iii). Then (7.63), the definition of the square root, and induction imply that

$$V_t^{(0)-1} = O^\top \operatorname{diag}(\lambda_t^{(01)}, \ldots, \lambda_t^{(0K)}) O,$$

where

$$\lambda_t^{(0k)} = \frac{\sigma_\xi}{R a_t^{(0)}} \sqrt{\lambda_{t-1}^{(0k)}} - \sum_{i=1}^{I} \frac{a_t^{(i)}}{a_t^{(0)}} \lambda^{(ik)}, \quad k = 1, \ldots, K. \tag{7.64}$$

Equation (7.64) is a list of K strictly monotonically increasing and strictly concave scalar maps of the form $\lambda_t = b_1 \sqrt{\lambda_{t-1}} - b_2$ with random coefficients b_1 and b_2. These determine the eigenvalues of the covariance matrix $V_t^{(0)}$. Condition (ii) now guarantees that for fixed $a_t^{(i)}$, $i = 0, \ldots, I$, each of the maps (7.64) has two positive deterministic fixed points, the larger one of which is asymptotically stable. Hence, there exists a compact set $\mathbb{K} \subset \mathbb{R}_{++}^K$ which is forward invariant under the maps (7.64) such that $(\lambda_t^{(01)}, \ldots, \lambda_t^{(0K)}) \in \mathbb{K}$ for all times $t \geq 0$.

□

References

Adams, R. A. (1975): *Sobolev Spaces*. Academic Press, New York a.o.

Anderson, S. P., A. de Palma & J.-F. Thisse (1992): *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge (Mass.) a.o.

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.

Auerbach, A. J. & L. J. Kotlikoff (1987): *Dynamic Fiscal Policy*. Cambridge University Press, Cambridge.

Azariadis, C. (1993): *Intertemporal Macroeconomics*. Blackwell Publishers, Oxford a.o.

Barnsley, M. (1988): *Fractals Everywhere*. Academic Press, New York a.o.

Bauer, H. (1991): *Wahrscheinlichkeitstheorie*. Walter de Gruyter Verlag, Berlin a.o., 4th edn.

——— (1992): *Maß– und Integrationstheorie*. Walter de Gruyter Verlag, Berlin a.o., 2nd edn.

Berger, M. S. (1977): *Nonlinearity and Functional Analysis*. Academic Press, New York a.o.

Bhattacharya, R. & M. Majumdar (2001): "On a Class of Stable Random Dynamical Systems: Theory and Application", *Journal of Economic Theory*, 96(1/2), 208–229.

Blume, L. & D. Easley (1982): "Learning to Be Rational", *Journal of Economic Theory*, 26, 340–351.

——— (1993): "What Has the Rational Learning Literature Taught Us?", in *Essays in Learning and Rationality in Economics*, ed. by A. Kirman & M. Salmon. Blackwell, Oxford.

——— (1998): "Rational Expectations and Rational Learning", in *Organizations with Incomplete Information: Essays in Economic Analysis: A Tribute to Roy Radner*, ed. by M. Majumdar. Cambridge University Press, Cambridge.

Böhm, V. & C. Chiarella (2005): "Mean Variance Preferences, Expectations Formations, and the Dynamics of Random Asset Prices", *Mathematical Finance*, 15(1), 61–97.

Böhm, V., N. Deutscher & J. Wenzelburger (2000): "Endogenous Random Asset Prices in Overlapping Generations Economies", *Mathematical Finance*, 10(1), 23–38.

Böhm, V., T. Kikuchi & G. Vachadze (2006): "On the Role of Equity for the Dynamics of Capital Accumulation", Discussion Paper No. 551, Department of Economics, Bielefeld University.

Böhm, V. & J. Wenzelburger (1999): "Expectations, Forecasting and Perfect Foresight - A Dynamical Systems Approach", *Macroeconomic Dynamics*, 3(2), 167–186.

——— (2002): "Perfect Predictions in Economic Dynamical Systems with Random Perturbations", *Macroeconomic Dynamics*, 6(5), 687–712.

——— (2004): "Expectational Leads in Economic Dynamical Systems", in *Economic Complexity: Non-linear dynamics, multi-agents economies, and learning*, ed. by W. Barnett, C. Deissenberg & G. Feichtinger, vol. 14 of *International Symposia in Economic Theory and Econometrics*, chap. 13, pp. 333–361. Elsevier, Amsterdam.

——— (2005): "On the Performance of Efficient Portfolios", *Journal of Economic Dynamics and Control*, 29(4), 721–740.

Borovkov, A. A. (1998): *Ergodicity and Stability of Stochastic Processes.* John Wiley & Sons, New York a.o.

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.

Brock, W. A. & C. H. Hommes (1997a): "Models of Complexity in Economics and Finance", in *Systems Dynamics in Economic and Finance Models*, ed. by B. Hanzon, C. Heij, J. Schumacher & C. Praagman, chap. 1, pp. 3–44. Wiley, New York.

——— (1997b): "A Rational Route to Randomness", *Econometrica*, 65(5), 1059–1095.

——— (1998): "Heterogeneous Beliefs and Routes to Chaos in a Simple Asset Pricing Model.", *Journal of Economic Dynamics and Control*, 22, 1235–1274.

Bullard, J. (1994): "Learning Equilibria", *Journal of Economic Theory*, 64(2), 468–485.

Bullard, J. & J. Duffy (2001): "Learning Excess Volatility", *Macroeconomic Dynamics*, 5(2), 272–302.

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 (1996): "Laws of Large Numbers For Hilbert-Space Valued Mixingales With Applications", *Econometric Theory*, 12, 284–304.

——— (1998): "Nonparametric Adaptive Learning with Feedback", *Journal of Economic Theory*, 82(1), 199–222.

——— (2002): "Asymptotic Properties of Some Projection-Based Robbins-Monro Procedures in a Hilbert Space", *Studies in Nonlinear Dynamics and Econometrics*, 6(1), 1–53.

Chiarella, C. & X.-Z. He (2002): "Heterogenous Beliefs, Risk and Learning in a Simple Asset Pricing Model", *Computational Economics*, 19, 95–132.

——— (2003a): "The Dynamics of Beliefs and Learning under a_L–Processes; the Heterogeneous Case", *Journal of Economic Dynamics and Control*, 27(3), 503–531.

——— (2003b): "Heterogenous Beliefs, Risk and Learning in a Simple Asset Pricing Model with a Market Maker", *Macroeconomic Dynamics*, 7, 503 – 536.

De Long, J. B., A. Shleifer, L. H. Summers & R. J. Waldmann (1990): "Noise Traders Risk in Financial Markets", *Journal of Political Economy*, 98, 703–738.

Deimling, K. (1980): *Nonlinear Functional Analysis*. Springer, Berlin.

Duffie, D., J. Geanakoplos, A. Mas-Colell & A. McLennan (1994): "Stationary Markov Equilibria", *Econometrica*, 62(4), 745–781.

Egghe, L. (1984): *Stopping Time Techniques For Analysts and Probabilists*. Cambridge University Press, Cambridge.

Elaydi, S. N. (1996): *An Introduction to Difference Equations*. Springer-Verlag, Berlin a. o.

Evans, G. W. (1985): "Expectational Stability and the Multiple Equilibria Problem in Linear Rational Expectations Models", *Quarterly Journal of Economics*, 100, 1217–1233.

——— (1986): "Selection Criteria for Models with Non-Uniqueness", *Journal of Monetary Economics*, 18, 147–157.

——— (1989): "The Fragility of Sunspots and Bubbles", *Journal of Monetary Economics*, 23, 297–317.

Evans, G. W. & S. Honkapohja (1986): "A Complete Characterization of ARMA Solutions to Linear Rational Expectations Models", *Review of Economic Studies*, 53, 227–239.

——— (1992): "On the Robustness of Bubbles in Linear RE Models", *International Economic Review*, 33, 1–14.

——— (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press.

Föllmer, H., U. Horst & A. P. Kirman (2005): "Equilibria in financial markets with heterogeneous agents: A probabilistic perspective", *Journal of Mathematical Economics*, 41, 123–155.

Fourgeaud, C., C. Gourieroux & J. Pradel (1986): "Learning Procedures and Convergence to Rationality", *Econometrica*, 54, 845–869.

Froot, K. A. & M. Obstfeld (1991): "Intrinsic Bubbles: The Case of Stock Prices", *American Economic Review*, 81, 1189–1214.

Fudenberg, D. & D. K. Levine (1998): *The Theory of Learning in Games*. MIT Press.

Gale, D. & H. Nikaido (1965): "The Jacobian Matrix and Global Univalence of Mappings", *Mathematische Annalen*, 159, 68–80.

Gallant, A. R. & H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell.

Geanakopolos, J., M. Magill & M. Quinzii (2004): "Demography and the Long-Run Predictability of the Stock Market", *Brookings papers on economic activity*, 1, 241–326.

Gohberg, I., P. Lancaster & L. Rodman (1982): *Matrix Polynomials*. Academic Press, New York.

Grandmont, J.-M. (1982): "Temporary General Equilibrium Theory", in *Handbook of Mathematical Economics*, ed. by K. Arrow & M. Intrilligator. North-Holland Publishing Company, Amsterdam a.o.

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.

Hillebrand, M. (2003): "Sequentielle Portefeuille-Entscheidungen mit mehrperiodigem Planungszeitraum – Ein Finanzmarktmodell mit endogener Preisbildung", Master's thesis, Bielefeld University.

——— (2006): “”Pension Systems, Demographic Change, and the Stock Market - Welfare and Sustainability in a Stochastic OLG Model”, Ph.D. thesis, Bielefeld University, in preparation.

Hillebrand, M. & J. Wenzelburger (2006a): “The Impact of Multiperiod Planning Horizons on Portfolios and Asset Prices in a Dynamic CAPM”, forthcoming *Journal of Mathematical Economics.*

——— (2006b): “On the Dynamics of Asset Prices and Portfolios in a Multiperiod CAPM”, *Chaos, Solitons and Fractals*, 29(3), 578–594.

Hommes, C. H. & G. Sorger (1998): “Consistent Expectations Equilibria”, *Macroeconomic Dynamics*, 2, 287–321.

Hommes, C. H., G. Sorger & F. Wagner (2004): “Consistency of Linear Forecasts in a Nonlinear Stochastic Economy”, CeNDEF working paper, University of Amsterdam, in preparation.

Horst, U. (2005): “Financial price fluctuations in a stock market model with many interacting agents”, *Economic Theory*, 25(4), 917–932.

Horst, U. & J. Wenzelburger (2005): “Non-ergodic behavior in a financial market with interacting investors”, Discussion Paper No. 544, Department of Economics, Bielefeld University.

Ingersoll, J. E. (1987): *Theory of Financial Decision Making.* Rowman & Littlefield, Totowa (NJ).

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl & T.-C. Lee (1985): *The Theory and Practice of Econometrics.* John Wiley and Sons, 2nd edn.

Jungeilges, J. A. (2003): “On Chaotic Consistent Expectations Equilibria”, Discussion Paper No. 497, Department of Economics, Bielefeld University.

Kailath, T. (1980): *Linear Systems.* Prentice-Hall, Inc., Englewood Cliffs, N. J.

Kelly, D. L. & J. Shorish (2000): “Stability of Functional Rational Expectations Equilibria”, *Journal of Economic Theory*, 95(2), 215–250.

Kirman, A. & M. Salmon (eds.) (1995): *Learning and Rationality in Economics.* Blackwell Publishers, Oxford a.o.

Kirman, A. P. (1975): “Learning by Firms about Demand Conditions”, in *Adaptive Economic Models*, ed. by R. Day & T. Groves, pp. 137–156. Academic Press.

Kuan, C.-M. & H. White (1994): “Adaptive Learning with Nonlinear Dynamics Driven By Dependent Processes”, *Econometrica*, 62(5), 1087–1114.

Kübler, F. & H. Polemarchakis (2004): “Stationary Markov Equilibria for Overlapping Generations”, *Economic Theory*, 24(3), 623–643.

Kushner, Harold J, . & G. G. Yin (2003): *Stochastic Approximation and Recursive Algorithms and Applications*, Stochastic Modelling and Applied Probability. Springer-Verlag, 2nd. edn.

Kushner, H. J. & D. S. Clark (1978): *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, no. 26 in Applied Mathematical Sciences. Springer-Verlag, New York a.o.

Lai, T. L. & C. Z. Wei (1982): “Least Squares Estimates In Stochastic Regression Models With Applications To Identification And Control Of Dynamic Systems”, *Journal of Multivariate Analysis*, 10(1), 154–166.

——— (1983): “A Note on Martingale Difference Sequences Satisfying the Local Marcinkiewicz-Zygmund Condition”, *Bulletin of the Institute of Mathematics Academia Sinica*, 11, 1–13.

——— (1986a): “Extended Least Squares and Their Applications to Adaptive Control and Prediction Linear Systems”, *IEEE Transactions on Automatic Control*, AC-31, 898–905.

——— (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.

Lang, S. (1968): *Analysis I.* Addison-Wesley Publishing Company, Reading (Mass.) a.o.

Lasota, A. & M. Mackey (1994): *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, vol. 97 of *Applied Mathematical Science.* Springer-Verlag, New York a.o.

LeRoy, S. & J. Werner (2001): *Principles of Financial Economics.* Cambridge University Press, Cambridge, UK.

Li, X. & J. Wenzelburger (2005): "Price Formation and Asset Allocations of the Electronic Equity Trading System *Xetra*", Discussion Paper No. 540, Department of Economics, Bielefeld University.

Lipschutz, S. (1965): *General Topology.* Schaum Publishing Company.

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 (1983): *Theory and Practice of Recursive Identification.* MIT Press, Cambridge MA.

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.

Lukacs, E. (1975): *Stochastic Convergence.* Academic Press, New York a.o.

Marcet, A. & T. J. Sargent (1989a): "Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models", *Journal of Economic Theory*, 48, 337–368.

——— (1989b): "Least Square Learning and the Dynamics of Hyperinflation", in *Chaos, Complexity, and Nonlinearity, Proceedings of the Fourth International Symposium in Economic Theory and Econometrics*, ed. by W. Barnett, J. Geweke & K. Shell. Cambridge University Press, Cambridge (Mass.) a.o.

Markowitz, H. (1952): "Portfolio Selection", *Journal of Finance*, 7, 77–91.

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 the Minimal State Variable Criterion in Rational Expectations Models", *International Tax and Public Finance*, 6, 621–639.

McLeish, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws", *Annals of Probability*, 3, 829–839.

Metivier, M. & P. Priouret (1984): "Applications of a Kushner and Clark Lemma to General Classes of Stochastic Algorithms", *IEEE Trans. Information Theory*, 30, 140–151.

Pliska, S. (1997): *Introduction to Mathematical Finance.* Blackwell Publishers Inc, Massachusetts.

Prakasa Rao, B. L. S. (1983): *Nonparametric Functional Estimation.* Academic Press, San Diego.

Rheinboldt, W. C. (1969): "Local Mapping Relations and Global Implicit Function Theorems", *Trans. Amer. Math. Soc.*, 138, 183–198.

ROBBINS, H. & S. MONRO (1951): "A Stochastic Approximation Method", *Annals of Mathematical Statistics*, pp. 400–407.

SARGENT, T. J. (1993): *Bounded Rationality in Macroeconomics*. Clarendon Press, Oxford.

SCHENK-HOPPÉ, K. R. & B. SCHMALFUSS (1998): "Random Fixed Points in a Stochastic Solow Growth Model", Discussion Paper No. 387, Department of Economics, Bielfeld University.

SCHMALFUSS, B. (1996): "A Random Fixed Point Theorem Based on Lyapunov Exponents", *Random and Computational Dynamics*, 4(4), 257–268.

SCHMALFUSS, B. (1998): "A Random Fixed Point Theorem and the Random Graph Transformation", *Journal of Mathematical Analysis and Application*, 225(1), 91–113.

SCHÖNHOFER, M. (1999): "Chaotic Learning Equilibria", *Journal of Economic Theory*, 89(1), 1–20.

——— (2001): "Can Agents Learn Their Way Out of Chaos?", *Journal of Economic Behavior and Organization*, 44(1), 71–83.

SIMON, H. A. (1982): *Models of Bounded Rationality*. MIT Press, Cambridge Mass.

SIMON, J. D. & S. K. MITTER (1968): "A Theory of Modal Control", *Information and Control*, 13, 316–353.

SOLO, V. (1979): "The Convergence of AML", *IEEE Transactions on Automatic Control*, AC-24, 958–962.

SPEAR, S. (1988): "Existence and Local Uniqueness of Functional Rational Expectations Equilibria in Dynamic Economic System Models", *Journal of Economic Theory*, 44(1), 124–155.

SPEAR, S. & S. SRIVASTAVA (1986): "Markov Rational Expectations Equilibria in a Stochastic Overlapping Generations Model", *Journal of Economic Theory*, 38(1), 35–46.

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

TOBIN, J. (1958): "Liquidity Preference as Behavior Towards Risk", *Review of Economic Studies*, 25, 65–86.

VARGA, R. S. (1962): *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs.

WENZELBURGER, J. (2002): "Global Convergence of Adaptive Learning in Models of Pure Exchange", *Economic Theory*, 19(4), 649–672.

——— (2004a): "Learning to play best response in duopoly games", *International Game Theory Review*, 6(3), 443–459.

——— (2004b): "Learning to Predict Rationally when Beliefs are Heterogeneous", *Journal of Economic Dynamics and Control*, 28(10), 2075–2104.

——— (2005): "Duopolistic competition in the CAPM", Mimeo, Department of Economics, Bielefeld University.

——— (2006): "Learning in Linear Models with Expectational Leads", forthcoming *Journal of Mathematical Economics*.

WHITE, H., A. GALLANT, K. HORNIK, M. STINCHCOMBE & J. WOOLRIDGE (1992): *Artificial Neural Networks: Approximation and Learning Theory*. Blackwell, Cambridge (Mass.) a.o.

YIN, G. (1992): "On $\mathbb{H}$-Valued Stochastic Approximation: Finite Dimensional Projections", *Stochastic Analysis and Applications*, 10(3), 363–377.

YIN, G. & Y. M. ZHU (1990): "On H-Valued Robbins-Monro Processes", *Journal of Multivariate Analysis*, 34, 116–140.

YOSIDA, K. (1978): *Functional Analysis*. Springer-Verlag, fifth edn.

Lecture Notes in Economics and Mathematical Systems

For information about Vols. 1–488
please contact your bookseller or Springer-Verlag

Vol. 489: T. Steger, Transitional Dynamics and Economic Growth in Developing Countries. VIII, 151 pages. 2000.

Vol. 490: S. Minner, Strategic Safety Stocks in Supply Chains. XI, 214 pages. 2000.

Vol. 491: M. Ehrgott, Multicriteria Optimization. VIII, 242 pages. 2000.

Vol. 492: T. Phan Huy, Constraint Propagation in Flexible Manufacturing. IX, 258 pages. 2000.

Vol. 493: J. Zhu, Modular Pricing of Options. X, 170 pages. 2000.

Vol. 494: D. Franzen, Design of Master Agreements for OTC Derivatives. VIII, 175 pages. 2001.

Vol. 495: I. Konnov, Combined Relaxation Methods for Variational Inequalities. XI, 181 pages. 2001.

Vol. 496: P. Weiß, Unemployment in Open Economies. XII, 226 pages. 2001.

Vol. 497: J. Inkmann, Conditional Moment Estimation of Nonlinear Equation Systems. VIII, 214 pages. 2001.

Vol. 498: M. Reutter, A Macroeconomic Model of West German Unemployment. X, 125 pages. 2001.

Vol. 499: A. Casajus, Focal Points in Framed Games. XI, 131 pages. 2001.

Vol. 500: F. Nardini, Technical Progress and Economic Growth. XVII, 191 pages. 2001.

Vol. 501: M. Fleischmann, Quantitative Models for Reverse Logistics. XI, 181 pages. 2001.

Vol. 502: N. Hadjisavvas, J. E. Martínez-Legaz, J.-P. Penot (Eds.), Generalized Convexity and Generalized Monotonicity. IX, 410 pages. 2001.

Vol. 503: A. Kirman, J.-B. Zimmermann (Eds.), Economics with Heterogenous Interacting Agents. VII, 343 pages. 2001.

Vol. 504: P.-Y. Moix (Ed.), The Measurement of Market Risk. XI, 272 pages. 2001.

Vol. 505: S. Voß, J. R. Daduna (Eds.), Computer-Aided Scheduling of Public Transport. XI, 466 pages. 2001.

Vol. 506: B. P. Kellerhals, Financial Pricing Models in Con-tinuous Time and Kalman Filtering. XIV, 247 pages. 2001.

Vol. 507: M. Koksalan, S. Zionts, Multiple Criteria Decision Making in the New Millenium. XII, 481 pages. 2001.

Vol. 508: K. Neumann, C. Schwindt, J. Zimmermann, Project Scheduling with Time Windows and Scarce Resources. XI, 335 pages. 2002.

Vol. 509: D. Hornung, Investment, R&D, and Long-Run Growth. XVI, 194 pages. 2002.

Vol. 510: A. S. Tangian, Constructing and Applying Objective Functions. XII, 582 pages. 2002.

Vol. 511: M. Külpmann, Stock Market Overreaction and Fundamental Valuation. IX, 198 pages. 2002.

Vol. 512: W.-B. Zhang, An Economic Theory of Cities.XI, 220 pages. 2002.

Vol. 513: K. Marti, Stochastic Optimization Techniques. VIII, 364 pages. 2002.

Vol. 514: S. Wang, Y. Xia, Portfolio and Asset Pricing. XII, 200 pages. 2002.

Vol. 515: G. Heisig, Planning Stability in Material Requirements Planning System. XII, 264 pages. 2002.

Vol. 516: B. Schmid, Pricing Credit Linked Financial Instruments. X, 246 pages. 2002.

Vol. 517: H. I. Meinhardt, Cooperative Decision Making in Common Pool Situations. VIII, 205 pages. 2002.

Vol. 518: S. Napel, Bilateral Bargaining. VIII, 188 pages. 2002.

Vol. 519: A. Klose, G. Speranza, L. N. Van Wassenhove (Eds.), Quantitative Approaches to Distribution Logistics and Supply Chain Management. XIII, 421 pages. 2002.

Vol. 520: B. Glaser, Efficiency versus Sustainability in Dynamic Decision Making. IX, 252 pages. 2002.

Vol. 521: R. Cowan, N. Jonard (Eds.), Heterogenous Agents, Interactions and Economic Performance. XIV, 339 pages. 2003.

Vol. 522: C. Neff, Corporate Finance, Innovation, and Strategic Competition. IX, 218 pages. 2003.

Vol. 523: W.-B. Zhang, A Theory of Interregional Dynamics. XI, 231 pages. 2003.

Vol. 524: M. Frölich, Programme Evaluation and Treatment Choise. VIII, 191 pages. 2003.

Vol. 525: S. Spinler, Capacity Reservation for Capital- Intensive Technologies. XVI, 139 pages. 2003.

Vol. 526: C. F. Daganzo, A Theory of Supply Chains. VIII, 123 pages. 2003.

Vol. 527: C. E. Metz, Information Dissemination in Currency Crises. XI, 231 pages. 2003.

Vol. 528: R. Stolletz, Performance Analysis and Optimization of Inbound Call Centers. X, 219 pages. 2003.

Vol. 529: W. Krabs, S. W. Pickl, Analysis, Controllability and Optimization of Time-Discrete Systems and Dynamical Games. XII, 187 pages. 2003.

Vol. 530: R. Wapler, Unemployment, Market Structure and Growth. XXVII, 207 pages. 2003.

Vol. 531: M. Gallegati, A. Kirman, M. Marsili (Eds.), The Complex Dynamics of Economic Interaction. XV, 402 pages, 2004.

Vol. 532: K. Marti, Y. Ermoliev, G. Pflug (Eds.), Dynamic Stochastic Optimization. VIII, 336 pages. 2004.

Vol. 533: G. Dudek, Collaborative Planning in Supply Chains. X, 234 pages. 2004.

Vol. 534: M. Runkel, Environmental and Resource Policy for Consumer Durables. X, 197 pages. 2004.

Vol. 535: X. Gandibleux, M. Sevaux, K. Sörensen, V. Tkindt (Eds.), Metaheuristics for Multiobjective Optimisation. IX, 249 pages. 2004.

Vol. 536: R. Brüggemann, Model Reduction Methods for Vector Autoregressive Processes. X, 218 pages. 2004.

Vol. 537: A. Esser, Pricing in (In)Complete Markets. XI, 122 pages, 2004.

Vol. 538: S. Kokot, The Econometrics of Sequential Trade Models. XI, 193 pages. 2004.

Vol. 539: N. Hautsch, Modelling Irregularly Spaced Financial Data. XII, 291 pages. 2004.

Vol. 540: H. Kraft, Optimal Portfolios with Stochastic Interest Rates and Defaultable Assets. X, 173 pages. 2004.

Vol. 541: G.-y. Chen, X. Huang, X. Yang, Vector Optimization. X, 306 pages. 2005.

Vol. 542: J. Lingens, Union Wage Bargaining and Economic Growth. XIII, 199 pages. 2004.

Vol. 543: C. Benkert, Default Risk in Bond and Credit Derivatives Markets. IX, 135 pages. 2004.

Vol. 544: B. Fleischmann, A. Klose, Distribution Logistics. X, 284 pages. 2004.

Vol. 545: R. Hafner, Stochastic Implied Volatility. XI, 229 pages. 2004.

Vol. 546: D. Quadt, Lot-Sizing and Scheduling for Flexible Flow Lines. XVIII, 227 pages. 2004.

Vol. 547: M. Wildi, Signal Extraction. XI, 279 pages. 2005.

Vol. 548: D. Kuhn, Generalized Bounds for Convex Multistage Stochastic Programs. XI, 190 pages. 2005.

Vol. 549: G. N. Krieg, Kanban-Controlled Manufacturing Systems. IX, 236 pages. 2005.

Vol. 550: T. Lux, S. Reitz, E. Samanidou, Nonlinear Dynamics and Heterogeneous Interacting Agents. XIII, 327 pages. 2005.

Vol. 551: J. Leskow, M. Puchet Anyul, L. F. Punzo, New Tools of Economic Dynamics. XIX, 392 pages. 2005.

Vol. 552: C. Suerie, Time Continuity in Discrete Time Models. XVIII, 229 pages. 2005.

Vol. 553: B. Mönch, Strategic Trading in Illiquid Markets. XIII, 116 pages. 2005.

Vol. 554: R. Foellmi, Consumption Structure and Macroeconomics. IX, 152 pages. 2005.

Vol. 555: J. Wenzelburger, Learning in Economic Systems with Expectations Feedback (planned) 2005.

Vol. 556: R. Branzei, D. Dimitrov, S. Tijs, Models in Cooperative Game Theory. VIII, 135 pages. 2005.

Vol. 557: S. Barbaro, Equity and Efficiency Considerations of Public Higer Education. XII, 128 pages. 2005.

Vol. 558: M. Faliva, M. G. Zoia, Topics in Dynamic Model Analysis. X, 144 pages. 2005.

Vol. 559: M. Schulmerich, Real Options Valuation. XVI, 357 pages. 2005.

Vol. 560: A. von Schemde, Index and Stability in Bimatrix Games. X, 151 pages. 2005.

Vol. 561: H. Bobzin, Principles of Network Economics. XX, 390 pages. 2006.

Vol. 562: T. Langenberg, Standardization and Expectations. IX, 132 pages. 2006.

Vol. 563: A. Seeger (Ed.), Recent Advances in Optimization. XI, 455 pages. 2006.

Vol. 564: P. Mathieu, B. Beaufils, O. Brandouy (Eds.), Artificial Economics. XIII, 237 pages. 2005.

Vol. 565: W. Lemke, Term Structure Modeling and Estimation in a State Space Framework. IX, 224 pages. 2006.

Vol. 566: M. Genser, A Structural Framework for the Pricing of Corporate Securities. XIX, 176 pages. 2006.

Vol. 567: A. Namatame, T. Kaizouji, Y. Aruga (Eds.), The Complex Networks of Economic Interactions. XI, 343 pages. 2006.

Vol. 568: M. Caliendo, Microeconometric Evaluation of Labour Market Policies. XVII, 258 pages. 2006.

Vol. 569: L. Neubecker, Strategic Competition in Oligopolies with Fluctuating Demand. IX, 233 pages. 2006.

Vol. 570: J. Woo, The Political Economy of Fiscal Policy. X, 169 pages. 2006.

Vol. 571: T. Herwig, Market-Conform Valuation of Options. VIII, 104 pages. 2006.

Vol. 572: M. F. Jäkel, Pensionomics. XII, 316 pages. 2006

Vol. 573: J. Emami Namini, International Trade and Multinational Activity, X, 159 pages, 2006.

Vol. 574: R. Kleber, Dynamic Inventory Management in Reverse Logisnes, XII, 181 pages, 2006.

Vol. 575: R. Hellermann, Capacity Options for Revenue Management, XV, 199 pages, 2006.

Vol. 576: J. Zajac, Economics Dynamics, Information and Equilibnum, X, 284 pages, 2006.

Vol. 577: K. Rudolph, Bargaining Power Effects in Financial Contracting, XVIII, 330 pages, 2006.

Vol. 578: J. Kühn, Optimal Risk-Return Trade-Offs of Commercial Banks, IX, 149 pages, 2006.

Vol. 579: D. Sondermann, Introduction to Stochastic Calculus for Finance, X, 136 pages, 2006.

Vol. 580: S. Seifert, Posted Price Offers in Internet Auction Markets, IX, 186 pages, 2006.

Vol. 581: K. Marti; Y. Ermoliev; M. Makowsk; G. Pflug (Eds.), Coping with Uncertainty, XIII, 330 pages, 2006 (planned).

Vol.583: Konnov, Igor V.; Luc, Dinh The; Rubinov, Alexander M. (Eds.), Generalized Convexity and Related Topics, IX, 469 pages, 2006 (planned).