A System Theoretic Approach to Behavioral Finance

by

Zalán Mátyás

Supervisor: László Gerencsér

Department of Probability Theory and Statistics,
Eötvös Loránd University,
Faculty of Sciences

June 2002

Contents

1 Introduction

Financial markets are playing an ever increasing role in modern life. The wide variety of stocks and the diversity of the participants make these markets complex systems. When modeling such systems in economics, the rationality of the participants is usually assumed. This, however, is typically not the case: a lot of people often behave less than rationally. The aim of this paper is to present a possible way of describing financial markets with irrational participants. The mathematical language chosen for this purpose is the field of systems and control theory.

The organization of this paper is as follows. In Chapter 2 we introduce some behavioral phenomena that may affect the participants of a financial market. In Chapter 3, based mainly on the work of Willems [14], we present a novel approach to constructing and analyzing mathematical models. In Chapter 4 we discuss the mathematical definition of dynamical systems and introduce some related concepts. Chapter 5 deals with a special subclass of dynamical systems, namely with discrete-time linear systems. Mainly on grounds of Kučera [11], we give a brief review of the main results of linear system theory. Armed with all these tools, in the next chapter we present a few simple models for financial markets affected by behavioral phenomena. Chapter 7 contains the main results of this paper: a criterion is presented for optimal linear predictors in the closed-loop system; furthermore, we investigate under what circumstances a given behavior can be replicated via a feedback interconnection. Finally some closing remarks are put forward in the last chapter.

Acknowledgements.
I would like to express my gratitude to László Gerencsér for the interesting and challenging topic, for supplying me with papers and references, and for carefully and patiently answering all my questions. Also thanks to Gábor Molnár-Sáska for his useful hints in typesetting this paper. And last, I am grateful to the MTA SZTAKI (Computer and Automation Research Institute, Hungarian Academy of Sciences) for an environment where I could undisturbed concentrate on my research field.

2 Behavioral Foundations

The traditional approach in economic modeling is based on two main concepts: rational behavior and efficient markets. The first assumption expresses the belief that people are rational, i.e. they interpret and process correctly public and private information. While the *rationality principle* is central to any economic explanation, its exact formulation depends on the economic paradigm. In the particular economic paradigm that finance has embraced, economic man is modeled as an independent maximizer of an expected utility function. The other main principle, the so called *efficient market hypothesis* assumes that people are fully informed, have complete knowledge of the fundamental structure of the economy (for example agents are assumed to have knowledge about supply and demand functions, about the stochastic laws governing the motion of the economy over time etc.), there is independence across individuals (no imitation occurs among them) and their decisions lead to a price equilibrium, the so called efficient price, which is rather stable in the absence of new information. It is also often assumed that markets have no memory: past prices records cannot help to predict future prices.

For a given stock, traditional rational expectations asset pricing models suppose everyone knows all the facts, uses them independently and makes a time discounting for projected earnings to arrive basically at the same price or a bracket of individual estimated prices due to each one's utility coefficient (a term measuring the given person's risk sensitivity). This range of buying/selling proposals will create a stable supply/demand balance (at least in the absence of new relevant data) in the market, with a stable price. The market's pricing of an item is thus the exact and full reflection of all relevant information and therefore it is the best estimate of its price; it also changes immediately and correctly as soon as new information arrive.

Based on the notions of efficient markets and economic rationality standard financial theory has difficulty explaining available empirical evidence. To explain why investors sometimes "get it wrong", financial economists have relaxed the traditional assumptions in many ways. Probably the best known are behavioral explanations that relax the assumption of completely rational information processing and allow the possibility that some of the agents in the economy behave less than fully rationally at least some of the time. Proponents of **behavioral finance** (as this set of theories is usually referred to) argue that psychological forces interfere with the main components of the traditional paradigm. They maintain that psychological phenomena prevent decision makers from acting in a rational manner, which creates market inefficiencies (in the shape of mispricings).

To see some examples that motivate the behavioral approach, consider the following two decision problems that have been used by psychologists in controlled experiments to study various aspects of human behavior and how they are manifested in decision-making.

Example 1. Relative to all the people you work with, how would you rate yourself as a driver? (1) Above average? (2) Average? (3) Below average? Here average is defined as the median.

When this question is asked from groups of people who actually work together, between 65 and 80 percent tend to rate themselves as being above the average. Since no more than 50 percent of any group can be above the median, this example shows us that people are generally *overconfident* about their own driving abilities. This turns out to be a general phenomenon. When it comes to difficult or challenging tasks, most people are overconfident about their own abilities and their own knowledge.

A more sophisticated example was posed in the article of A. Tversky and D. Kahneman [7]:

Example 2. Imagine that you face with the following pair of *concurrent* decisions. First examine both decisions and then indicate the option you prefer.

1. *First decision:* Choose between
   - A. a sure gain of $2,400
   - B. a 25 percent chance of gaining $10,000 and a 75 percent chance of gaining nothing.
2. *Second decision:* Choose between
   - C. a sure loss of $7,500
   - D. a 75 percent chance of losing $10,000 and a 25 percent chance of losing nothing.

Most people choose A in the first decision and D in the second. Notice that A is the risk-averse choice: most people find a sure gain of $2,400 more attractive than a gain of a considerably greater amount of money but which involves risk. In contrast, D is not the risk-averse choice. Yet most people take a chance instead of a guaranteed loss. Why? Because they hate to lose! And the uncertain choice holds out the hope that they won't have to lose. The psychologists call this phenomenon *loss aversion*. Empirical evidence shows that a loss has about two and a half times the impact of a gain of the same magnitude.

As noted in the description of the example, the first and second decision problems constitute a concurrent "package". However, most people do not see the package. They separate the choices into mental accounts, one account for the first decision and one for the second. And that brings us to what the behaviorists call *frame dependence*. People who choose A and D end up facing a 25 percent chance of winning $\$2,400$ and a 75 percent chance of losing $\$7,600$. However, they could do better. They could opt for B and C, which would lead them to face a 25 percent chance of winning $\$2,500$ and a 75 percent chance of losing $\$7,500$. But the manner in which these cash flows are packaged or framed is not straightforward. Therefore, even though people are fully informed, they act as if they preferred to give up $100. Thus framing matters.

These examples show that behavioral finance deals mostly with investor irrationality, bounded rationality, cognitive and decision biases. Four main types of behavioral phenomena are sources of mispricings:

- Individual cognitive biases
- Collective cognitive biases
- Individual emotions and passions
- Social behavior

2.1 Individual Understanding Anomalies

Overconfidence: individuals overestimate their information (knowledge illusion) and abilities.

Anchoring/conservatism: people have in memory some reference points (anchors), for example a previous stock price or price trend. They cling excessively to prior beliefs when exposed to new evidence, they reject new facts that are contrary to their preconceived ideas. Thus, when they get new information, they adjust their references insufficiently (this is termed underreaction by psychologists). Later, when those information get progressively confirmed, they adjust too much (overreaction).

Representativeness heuristic: people expect random sequences to reflect all the essential characteristics of an underlying distribution, even over short recent intervals. They seem to expect key parameters to be represented in any recent sequence of generated data, leading them to be excessively sensitive to recent data.

Rationalizing: getting stuck on immediate apparent explanations leads to explain by a rational story whatever action, even irrational (finding a good reason for an urge to buy or sell); whatever event, even of unclear origin (finding a reason for a market rise or fall); and whatever possible source of responsibility (finding an immediate culprit for a bad event).

Framing: most people are frame dependent and limit their approach to one angle. They select only one immediately apparent way of defining the question they ask themselves in such a way that it conditions their answer to it. The wording used to present an issue to them can also lead people to select inferior options.

Cognitive overload: human beings cannot process an arbitrarily big amount of information during a given time interval, they get overloaded. This may explain the tendency to follow the most apparent trends and discard weak signals.

Habits/conventions: these are other sources of oversimplification. In the absence of reference points for estimating the long-term yield of a capital asset, investors fall back on conventions. They assume that the present state of affairs will continue indefinitely into the future and that the current valuation is somewhat correct.

Tunnel vision: decisions made on shallow thinking, based on a too small number of keywords, anchors or decision models.

Hindsight bias: people tend to forget their original estimates. When seeing the outcome, they are likely to use it as a reference point and assume their estimates must have been close to it.

Confirmation bias: people suffer from the tendency to ascribe too much weight to evidence that confirms their views and too little weight to evidence that invalidates their views.

2.2 Collective/Group Cognitive Biases

These biases affect all the players in the economy or at least some dominant types of investors.

Imitation: there is often no independence across individuals, either concerning the whole population or some sections of it (see for example peer influence, neighborhood effect). People often find it rewarding to follow the most apparent trend.

Simplified common denominator: social behavior is usually based on common conventions and norms. This can lead individuals, as they get focused on these aspects, to overlook most weak signals.

Dominant investors: depositaries of economic power and some analysts, journalists acquire a guru status. The advice they give is taken for granted and their words get exaggerated interpretations.

Manipulations: already possible from person to person, manipulations get a new dimension in markets as groups and crowds are even more influencable. Also, people easily accept obedience even to apparent authority.

Error magnification: empirical data show that groups tend to amplify individual errors.

Investors profiling: investors, just as stocks, can be classified into different categories, each one having its own type of behavior.

2.3 Individual Emotions and Passions

Greed, fear, love, hate: these and other emotions all have effects on the market behavior of individuals.

Loss aversion: as seen above, even simple risk aversion can be biased. This may explain why in their portfolio people prefer to sell winners than losers: their aversion to an already real loss is stronger than their aversion to risk (the one they take by keeping downgraded stocks). Loss aversion is also responsible for the fact that downward price adjustments are more severe than their upward counterparts.

Commitment pride: once people start to accept to do a little thing (or just to listen to somebody), they feel committed towards themselves (or to others). Thus they accept more easily to go one step (and then several steps) further.

2.4 Social Psychology, Group and Crowd Behaviors

Collective hysterias: the collective behavior of the whole is qualitatively different from that of the sum of its individual parts. In a group or a crowd, individuals tend to lose their own reference and inhibitions. They share crowd emotions and behave like the crowd, sometimes going to extreme actions that they would never have done by themselves. This herd instinct can lead in the stockmarket to fashions and/or crashes.

Panic: panic contagions can spread faster than euphoria outbreaks. This makes for fast and massive price falls and higher volatility (price agitation) on the descending path than during the climb.

3 A Behavioral Approach to Modeling

In order to be able to describe a phenomenon occurring in a system, we first have to set up a model. Modeling is the process whereby the physical properties of a system are expressed in a mathematical form suited for further analysis. The model itself is *not* the system: it is just a mathematical idealization that represents selected aspects of the system's behavior relevant to the problem at hand; it is at best an approximation to reality. If the model is a good one, we can predict from it to a reasonable degree of accuracy how the system will perform under expected conditions. However, no model can exactly represent a real system under all possible circumstances since attempting to include all the phenomena present in the system would result in a mathematically untractable, hopelessly complex mass of detail. Golomb [4] expresses these ideas as follows:

> Basically, the model is like a *map*. It may obscure the complexities of the terrain, but it provides a simple enough picture to be grasped at a glance and is helpful in plotting a route from one point to another. But you will never strike oil by drilling through the map.

The general model structures that we use in this paper are referred to as the *behavioral approach*. The main underlying idea is that we view a mathematical model as an exclusion law: a model for a phenomenon permits the outcome of certain things, while other outcomes are declared impossible. Hence a model recognizes a certain subset from a university of possibilities. An elegant mathematical formalization of these ideas has been developed by Willems in [14].

Definition 3.0.1 A *mathematical model* is a pair $(\mathbb{U}, \mathfrak{B})$ with $\mathbb{U}$ a nonempty set, called the *universum* - its elements are called the *outcomes* - and $\mathfrak{B}$ a subset of $\mathbb{U}$, called the *behavior*.

In applications, models are often described by equations. In this case the behavior of the model is simply defined as those elements of the universum that satisfy the equations. This motivates the following definition:

Definition 3.0.2 Let $\mathbb{U}$ be a universum, $\mathbb{E}$ a set and $f_1, f_2 : \mathbb{U} \to \mathbb{E}$. The mathematical model $(\mathbb{U}, \mathfrak{B})$ with $\mathfrak{B} = \{u \in \mathbb{U} \mid f_1(u) = f_2(u)\}$ is said to be described by *behavioral equations* and is denoted by $(\mathbb{U}, \mathbb{E}, f_1, f_2)$. The set $\mathbb{E}$ is called the *equating space*. We also call $(\mathbb{U}, \mathbb{E}, f_1, f_2)$ a *behavioral equation representation* of $(\mathbb{U}, \mathfrak{B})$.

Example. Economists believe that there exists a relation between the amount P produced of a particular economic good, the capital K invested in the necessary infrastructure and the labor L expended towards its production. A typical model looks like $\mathbb{U} = \mathbb{R}_+^3$ with the behavior $\mathfrak{B} = \{(P, K, L) \in \mathbb{R}_+^3 \mid P = F(K, L)\}$, where $F : \mathbb{R}_+^2 \to \mathbb{R}_+$ is the production function. A well-known class of production functions are the so called Cobb-Douglas production functions: $F(K, L) = \alpha K^\beta L^\gamma$, where $\alpha, \beta, \gamma \in \mathbb{R}_+$, $0 \leq \beta \leq 1$, $0 \leq \gamma \leq 1$ are constant parameters depending on the production process.

Remark 3.0.3 Definition 3.0.2 can be simply modified to include models described by inequalities: just take $\mathbb{E}$ to be an ordered space and consider the behavioral inequality $f_1(u) \leq f_2(u)$. Many models in economics are of this nature.

Remark 3.0.4 Note that whereas behavioral equations specify the behavior uniquely, the converse is obviously not true, since given any bijection $f : \mathbb{E} \to \mathbb{E}'$, the pair of functions $(f \circ f_1, f \circ f_2)$ and (f_1, f_2) represent the same model. Since we have a tendency to think of mathematical models in terms of behavioral equations, most models being presented in this form, it is important to emphasize their ancillary role: *it is the behavior*, the solution set of the behavioral equations, not the behavioral equations themselves, *that is the essential result of a modeling procedure.*

In most modeling exercises, the behavior is most conveniently specified in terms of equations that in addition to the variables modeled (i.e the attributes in $\mathbb{U}$) also contain some additional variables. The variables whose behavior the model aims at describing are called *manifest variables*. We may think of them as external variables in contrast to the auxiliary variables (called *latent variables*) which are viewed as internal. These notions lead to the following definition.

Definition 3.0.5 A *mathematical model with latent variables* is defined as a triple $(\mathbb{U}, \mathbb{U}_l, \mathfrak{B}_f)$ with $\mathbb{U}$ the universum of *manifest variables*, $\mathbb{U}_l$ the universum of *latent variables*, and $\mathfrak{B}_f \subseteq \mathbb{U} \times \mathbb{U}_l$ the full behavior. It defines the *manifest mathematical model* $(\mathbb{U}, \mathfrak{B})$ with $\mathfrak{B} := \{u \in \mathbb{U} \mid \exists l \in \mathbb{U}_l$ such that $(u, l) \in \mathfrak{B}_f\}$; $\mathfrak{B}$ is called the manifest behavior or simply the behavior. We call $(\mathbb{U}, \mathbb{U}_l, \mathfrak{B}_f)$ a *latent variable representation* of $(\mathbb{U}, \mathfrak{B})$.

Some examples in economics: sales can be viewed as manifest, while consumer demand could be considered as a latent variable; prices are also implicit variables used in order to explain the production and exchange of economic goods; supply and demand functions are used when trying to figure out production volumes.

Latent variables appear whenever the system under consideration is viewed as an interconnection of several subsystems, and the modeling process is carried out by zooming in on the individual subsystems. The overall model is then obtained by combining the models of the subsystems with the interconnection constraints. This model invariably contains latent variables: the auxiliary variables introduced in order to express the interconnections play this role.

4 Dynamical Systems

A system in which time evolution is a crucial feature is usually called a dynamical system. We view a dynamical system as a mathematical model in which the objects of interest are functions: the universum is a function space and the system constrains the time signals that can be produced.

Definition 4.0.6 A *dynamical system* Σ is defined as a triple $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$ with $\mathbb{T}$ a subset of $\mathbb{R}$, called the *time axis*, $\mathbb{W}$ a non-empty set called the *signal space*, and $\mathfrak{B}$ a subset of $\mathbb{W}^{\mathbb{T}} := \{w \mid w : \mathbb{T} \to \mathbb{W}\}$ called the *behavior*.

The set $\mathbb{T}$ specifies the set of time instances relevant to our problem. In *discrete-time systems* it usually equals $\mathbb{Z}$ or $\mathbb{Z}_+$ while in *continuous-time systems* it is $\mathbb{R}$ or $\mathbb{R}_+$. The set $\mathbb{W}$ contains the outcomes of the dynamical system: the variables whose evolution in time we are describing. In economic models $\mathbb{W}$ is usually a finite dimensional vector space. When $\mathbb{W}$ is a finite set, the term *discrete-event systems* is often used. The behavior $\mathfrak{B}$ contains simply those time trajectories that are compatible with the laws that govern the system, i.e. all time signals $w : \mathbb{T} \to \mathbb{W}$ that according to the model can occur.

In dynamical systems the behavioral equations often take the form of difference equations (in discrete time) or differential equations (in continuous time). Since this paper deals mainly with discrete-time systems, we write down explicitly the formal definition only for these systems.

Definition 4.0.7 A *behavioral difference equation representation* of a discrete-time dynamical system with time axis $\mathbb{T} = \mathbb{Z}$ and signal space $\mathbb{W}$ is defined by a nonnegative integer L called the *lag*, a set $\mathbb{E}$ called the *equating space*, and two maps $f_1, f_2 : (\mathbb{W}^{\mathbb{T}})^{L+1} \to \mathbb{E}$, yielding the difference equations

$$f_1(w, qw, \ldots, q^{L-1}w, q^L w) = f_2(w, qw, \ldots, q^{L-1}w, q^L w)$$

where q^t denotes the backward t-shift, i.e. $(q^t w)(u) := w(u+t)$.

The definition of a latent variable model is easily generalized to dynamical systems.

Definition 4.0.8 A *dynamical system with latent variables* is defined as a quadruple $\Sigma_L = (\mathbb{T}, \mathbb{W}, \mathbb{L}, \mathfrak{B}_f)$ with $\mathbb{T} \subseteq \mathbb{R}$ the time axis, $\mathbb{W}$ the *manifest signal space*, $\mathbb{L}$ the *latent variable space*, and $\mathfrak{B}_f \subseteq (\mathbb{W} \times \mathbb{L})^{\mathbb{T}}$ the *full behavior*. It defines a *latent variable representation* of the dynamical system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$ with manifest behavior $\mathfrak{B} := \{w : \mathbb{T} \to \mathbb{W} \mid \exists l : \mathbb{T} \to \mathbb{L}$ such that $(w, l) \in \mathfrak{B}_f\}$.

Latent variable models are not just merely mathematical generalizations of dynamical systems since typical models will involve additional variables to those whose behavior we wish to model. Also, it is often impossible to obtain an explicit behavioral equation describing $\mathfrak{B}$ entirely in terms of the manifest variables.

4.1 Systems in Input/Output Forms

Until now we have viewed dynamical systems purely on a set-theoretic level. We did not pay attention to what the trajectories in the behavior actually look like. For our purposes, a system is viewed as a part of the world connected to its environment through the *input* and *output* terminals: it receives stimuli at the input and produces responses at the output. This leads us to the notion of input/output systems.

Definition 4.1.1 Let $\mathfrak{B}$ be a behavior with signal space $\mathbb{W} = \mathbb{R}^q$. Partition the signal space $\mathbb{R}^q = \mathbb{R}^m \times \mathbb{R}^l$ and $w \in {\mathbb{R}^q}^{\mathbb{T}}$ correspondingly as $w = \mathrm{col}(w_1, w_2)$ (i.e., $w_1 \in {\mathbb{R}^m}^{\mathbb{T}}$ and $w_2 \in {\mathbb{R}^l}^{\mathbb{T}}$). This partition is called an *input/output partition* if:

1. w_1 is free in the sense that for all $w_1 \in {\mathbb{R}^m}^{\mathbb{T}}$ there exists $w_2 \in {\mathbb{R}^l}^{\mathbb{T}}$ such that $w = \mathrm{col}(w_1, w_2) \in \mathfrak{B}$.

2. w_2 does not contain any further free components; i.e., given any w_1, none of the components of w_2 can be chosen arbitrarily.

If these conditions hold, then w_1 is called an *input variable* and w_2 is called an *output variable.*

Remark 4.1.2 The partition of w into input and output variables is in general not unique. A trivial example is the behavior defined by $w_1 = w_2$.

The system output may depend on the present input as well as on the past history of the system. To account for this memory, we introduce a special latent variable called the *state.*

Definition 4.1.3 Consider the latent variable dynamical system defined by definition 4.0.8. Let (w_1, l_1), $(w_2, l_2) \in \mathfrak{B}_f$ and $t_0 \in \mathbb{R}$ and suppose that l_1, l_2 are continuous. Define the concatenation of (w_1, l_1) and (w_2, l_2) at t_0 by (w, l) where

$$w(t) = \begin{cases} w_1(t), & t < t_0 \\ w_2(t), & t \geq t_0 \end{cases} \quad \text{and} \quad l(t) = \begin{cases} l_1(t), & t < t_0 \\ l_2(t), & t \geq t_0 \end{cases} .$$

Then $\mathfrak{B}_f$ is said to be a *state space model* with *state variable* l if $l_1(t_0) = l_2(t_0)$ implies $(w, l) \in \mathfrak{B}_f$.

The state property expresses that l splits the past and future of w. In other words, given (w^-, l^-), an observed past trajectory in the behavior, all we need to know to decide whether a given future trajectory can occur is the state $l^-(0)$: any trajectory (w^+, l^+) is a possible future continuation of (w^-, l^-) provided that $l^-(0) = l^+(0)$. As such, $l^-(0)$ contains all the information about the past required to be able to understand what the future may look like: $l^-(0)$ is indeed the *memory* of the system.

4.2 Some Basic Properties of Dynamical Systems

In order to get a mathematically well-tractable class of models, we need to impose some structure on our models. Below we define two algebraic properties of the behavior that play a central role in systems theory.

The first one is time-invariance: the laws of the system do not explicitly depend on time. This property is of utmost importance in applications.

Definition 4.2.1 A dynamical system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$ with $\mathbb{T} = \mathbb{R}$ or $\mathbb{Z}$ is said to be *time-invariant* if $q^t \mathfrak{B} = \mathfrak{B}$ for all $t \in \mathbb{T}$.

The other frequently used property is linearity:

Definition 4.2.2 A dynamical system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$ is said to be *linear* if $\mathbb{W}$ is a vector space over a field $\mathbb{F}$ and $\mathfrak{B}$ is a linear subspace of $\mathbb{W}^{\mathbb{T}}$.

Example. For a simple example of a behavior that is both linear and time-invariant consider the behavior defined by the behavioral equation $\frac{\mathrm{d}}{\mathrm{d}t} w = w$. It is given by $\{w \mid w(t) = ce^t,\ c \in \mathbb{R}\}$, which is easily seen to satisfy both properties.

Finally, a non-algebraic property is defined below.

Definition 4.2.3 An input/output partition $w = \mathrm{col}(w_1, w_2)$ is called *causal* if for any $t \in \mathbb{T}$ the output value $w_2(t)$ does not depend on the values of the input at times greater than t. If $w_2(t)$ depends only on the values of the input at times strictly less than t, then this partition is called *strictly causal.*

5 Linear Systems

In this chapter, based mainly on the work of Kučera [11], we discuss a very common class of dynamical systems, namely linear systems. The importance of such systems stems from at least three aspects. Firstly, many models used in applications are by their very nature linear. Secondly, under appropriate assumptions nonlinear models can be linearized in the neighborhood of a nominal trajectory, and the approximating linear system gives a local description of the nonlinear behavior. Thirdly, linear systems proved to be a mathematically well-tractable class of systems that lend themselves much better to analysis and synthesis techniques than nonlinear systems do. Much more is known about them. As such, the theory of linear systems not only plays an exemplary role for the nonlinear case, but has also reached a much higher degree of perfection.

5.1 Discrete-time Linear Systems

The study of discrete-time systems arises because frequently in practical situations system observations are made and control strategies are implemented at discrete-time instants only (sampled-data systems). An example of such a situation in the field of economics arises when certain statistics are compiled quarterly and budget controls are applied yearly. In these cases the discrete-time framework is a very natural one in which to give a model description.

As we have already pointed out, models are often described by equations. It can be shown (see for example [14]) that a behavior defined by linear constant-coefficient difference (or differential) equations always admits an input/output representation in the sense of definition 4.1.1. This insightful way of viewing a system can be combined with the notion of state. Thus we get a very common way of describing linear systems, namely the input/state/output representation. We now give a formal definition.

Definition 5.1.1 A *discrete-time linear input/state/output system* takes the form

$$\begin{aligned} x_{t+1} &= F_t x_t + G_t u_t \\ y_t &= H_t x_t + J_t u_t \end{aligned} \tag{1}$$

where the time subscript $t \in \{0, 1, \dots\}$, the input sequence $u_0, u_1, \dots$ takes values in $\mathbb{R}^m$, the output sequence $y_0, y_1, \dots$ in $\mathbb{R}^l$. The sequence of states $x_0, x_1, \dots$ takes values in $\mathbb{R}^n$, and n is called the *order* of the system. Finally, F_t, G_t, H_t and J_t are matrices of appropriate size with elements in $\mathbb{R}$. In the special case when F_t, G_t, H_t and J_t do not depend on t, we get the class of *discrete-time linear time-invariant systems.*

In most applications, the above defined discrete-time linear systems are somewhat unrealistic since the measurements may contain random errors (either due to the implementation of the given system or due to the disturbances imposed by nature). To account for these random phenomena, a broader class of linear systems is introduced, namely *linear stochastic systems.*

Definition 5.1.2 A *discrete-time linear stochastic system* is described by the recursive equations

$$\begin{aligned} x_{t+1} &= F_t x_t + G_t u_t + V_t v_t \\ y_t &= H_t x_t + J_t u_t + W_t w_t \end{aligned} \tag{2}$$

where $t \in \{0, 1, \ldots\}$, $u_t \in \mathbb{R}^m$ is the input variable (also called control variable), $y_t \in \mathbb{R}^l$ is the output variable, $x_t \in \mathbb{R}^n$ is the state variable and $\{v_t\} \in \mathbb{R}^n$ and $\{w_t\} \in \mathbb{R}^l$ are random sequences with mean zero

$$E \begin{bmatrix} v_t \\ w_t \end{bmatrix} = 0$$

and with covariance

$$E \begin{bmatrix} v_s \\ w_s \end{bmatrix} \begin{bmatrix} v_t^T & w_t^T \end{bmatrix} = \begin{bmatrix} Q_1 & S_1^T \\ S_1 & R_1 \end{bmatrix} \delta_{st}.$$

Finally, F_t, G_t, V_t, H_t, J_t and W_t are matrices of appropriate size with elements in $\mathbb{R}$. In the special case when these matrices do not depend on t, we get the class of *discrete-time linear time-invariant stochastic systems.*

In the remaining part of this chapter we define some concepts that play a major role in linear system theory and state the principle results needed later. For simplicity, we consider discrete-time linear time-invariant systems of the form

$$\begin{aligned} x_{t+1} &= F x_t + G u_t \\ y_t &= H x_t + J u_t \end{aligned} \tag{3}$$

with input variables u_t, output variables y_t and state variables x_t as defined above. Solving the system equations (3) we get

$$\begin{aligned} x_t &= F^t x_0 + F^{t-1} G u_0 + \cdots + G u_{t-1} \\ y_t &= H x_t + J u_t \end{aligned} \tag{4}$$

for $t \in \{0, 1, \ldots\}$. Thus the output sequence is conditioned by the initial state x_0 and the input sequence $u_0, \ldots, u_t$.

5.2 Reachability and Observability

Let us now examine separately the effect of the input on the state and the effect of the state on the output. First we concentrate on the input/state equation

$$x_{t+1} = Fx_t + Gu_t$$

and refer to the underlying system by the pair (F, G).
For $t = 0, 1, \ldots$ define

$$\mathcal{R}_t(F, G) := \text{Im}\ [G\ FG\ \ldots\ F^{t-1}G]\ .$$

In view of (4), $x_0 = 0$ gives

$$x_t = F^{t-1}Gu_0 + \cdots + Gu_{t-1}\ .$$

Hence $\mathcal{R}_t(F, G)$ is a subspace of $\mathbb{R}^n$ which consists of the states x_t that can be reached from the origin in t steps by applying an appropriate input sequence $u_0, u_1, \ldots, u_{t-1}$. That is why $\mathcal{R}_t(F, G)$, $t = 1, 2, \ldots$ are called the *reachability subspaces* of system (3). According to the Cayley-Hamilton theorem, for $t \geq n$ F^t can be expressed in terms of I, F, F^2, $\ldots$, F^{n-1}. Therefore, $\text{Im}\ [G\ FG\ \ldots\ F^{t-1}G] \subseteq \text{Im}\ [G\ FG\ \ldots\ F^{n-1}G]$ for all $t = 1, 2, \ldots$. This latter subspace plays a major part in linear system theory.

Definition 5.2.1 The pair (F, G) (and hence the underlying system) is said to be *reachable* if $\mathcal{R}_n(F, G) = \mathbb{R}^n$.

Remark 5.2.2 By definition, reachability holds if and only if the condition $\text{rank}[G\ FG\ \ldots\ F^{n-1}G] = n$ holds.

Now consider the effect of the state on the output. When $u_t = 0$ for $t = 0, 1, \ldots$ then (4) implies

$$\begin{aligned} y_0 &= Hx_0 \\ y_1 &= HFx_0 \\ &\vdots \\ y_{t-1} &= HF^{t-1}x_0\ . \end{aligned}$$

If the output sequence $y_0, \ldots, y_{t-1}$ uniquely determines the initial state x_0 (in other words, the above equations are not satisfied for any $x' \neq x_0$), then state x_t is said to be observable in t steps.

For $t = 0, 1, \ldots$ define

$$\mathcal{O}_t(F, H) := \{x \in \mathbb{R}^n \mid x \text{ is observable in } t \text{ steps } \} .$$

These subspaces of $\mathbb{R}^n$ are called the *observability subspaces* of system (3). Using again the Cayley-Hamilton theorem, it can be shown that $\mathcal{O}_t(F, H) \subseteq \mathcal{O}_n(F, H)$ for all $t = 1, 2, \ldots$. This motivates the following definition.

Definition 5.2.3 The pair (F, H) (and hence the underlying system) is said to be *observable* if $\mathcal{O}_n(F, H) = \mathbb{R}^n$.

Remark 5.2.4 It can be shown that observability of (F, H) is equivalent to reachability of (F^T, H^T). Therefore, system (3) is observable if and only if the following condition holds:

$$\text{rank} \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} = n.$$

Remark 5.2.5 For non-reachable and/or non-observable systems the question arises whether there exists a subsystem of the original system that has the desired properties. The so-called *Kalman decomposition* puts the reachability/observability structure of a system into evidence with the help of a non-singular state space transformation (see for example Rugh [16]).

5.3 Stability

Stability is a very common issue in many areas of mathematics. Intuitively, stability implies that small causes produce small effects. In linear systems stability captures the behavior of the unforced system $x_{t+1} = Fx_t$ (i.e when $u_t = 0$ for $t = 0, 1, \ldots$) for large values of t. There are several types of stability. In *structural stability* one wants small parameter changes to have a similar small influence on the behavior of the system. In *dynamic stability*, which is the topic of this section, one wants small disturbances in the initial conditions to have small effects on the solution of the dynamical equations.

Definition 5.3.1 The linear system (3) is said to be *asymptotically stable* (or just simply *stable*) if for arbitrary x_0, the solution x_t of the unforced system tends towards zero as $t \to \infty$.

Theorem 5.3.2 The linear system (3) is stable if and only if all eigenvalues of F have moduli less than unity.

Proof. The claim follows immediately from inspecting the canonical Jordan form of the matrix F. ■

Remark 5.3.3 The matrix F of a stable system is termed a *stability matrix*.

Let us now consider the controlled system

$$x_{t+1} = Fx_t + Gu_t \ .$$

If it is not stable, we might be able to stabilize it by finding a constant matrix L such that using the control law

$$u_t = -Lx_t$$

the resulting closed-loop system

$$x_{t+1} = (F - GL)x_t$$

will be stable.

Definition 5.3.4 The pair (F, G) is said to be *stabilizable* if there exists a constant matrix L such that $F - GL$ is a stability matrix.

Theorem 5.3.5 The pair (F, G) is stabilizable if and only if the system matrix of the non-reachable part of the system is a stability matrix.

Proof. By the Kalman decomposition there exists a basis in $\mathbb{R}^n$ in which F, G have the form

$$F = \begin{bmatrix} F_{11} & F_{12} \\ 0 & F_{22} \end{bmatrix}, \quad G = \begin{bmatrix} G_1 \\ 0 \end{bmatrix}$$

with (F_{11}, G_1) reachable. If (F, G) is stabilizable, there exists a matrix

$$L = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$$

such that

$$F - GL = \begin{bmatrix} F_{11} - G_1 L_1 & F_{12} - G_1 L_2 \\ 0 & F_{22} \end{bmatrix}$$

is a stability matrix. Hence F_{22}, the system matrix of the unreachable part of the system, is a stability matrix. On the other hand, reachability of (F_{11}, G_1) implies that L_1 can be found such that $F_{11} - G_1 L_1$ has any desired set of eigenvalues, see theorem 5.5.1. In particular, $F_{11} - G_1 L_1$ can be made into a stability matrix. Then stability of F_{22} implies stabilizability of (F, G). ■

5.4 Transfer Function

The input/state/output equations describe the full behavior of the system. Often we are interested only in the relation between certain variables, e.g. we would like to characterize the input/output behavior of the system. For this purpose, the concept of transfer functions is introduced.

Given a sequence a_t, $t = 0, 1, \dots$ define the *z-transform* $A(z)$ as

$$A(z) := \sum_{t=0}^{\infty} a_t z^{-t}$$

with z regarded as a complex variable. For a specific value of z this function is either finite or infinite. The set of complex numbers for which $A(z)$ is finite is called the *region of convergence* of $A(z)$.

Example. In everyday applications, the input is often bounded, i.e. there exists a constant $K > 0$ such that $|u_t| \leq K$ for all $t \in \{0, 1, \dots\}$. Since $\sum_{t=0}^{\infty} z^{-t} < \infty$ holds if and only if $|z| > 1$, it follows that the z-transform of the input sequence is defined for all $z \in \mathbb{C}$ outside the unit circle.

The z-transform provides a powerful method for solving difference equations by transforming them into algebraic equations. These algebraic equations then lend themselves to well-developed mathematical procedures. To illustrate this, take the z-transforms of both sides of equations (3) to get

$$\begin{aligned} zX(z) - zx_o &= FX(z) + GU(z) \\ Y(z) &= HX(z) + JU(z) \ . \end{aligned}$$

Here z is assumed to belong to the region of convergence of both $U(z)$ and $Y(z)$. Substituting $x_0 = 0$ in the above equation (which we may attain by a suitable state space transformation), it follows that

$$Y(z) = T(z)U(z)$$

with

$$T(z) := H(zI_n - F)^{-1}G + J \ .$$

The matrix $T(z) \in \mathbb{R}^{l \times m}$ relates the z-transform of the input sequence to that of the output sequence. For this reason it is called the *transfer function* of system (3). Using the standard decomposition, it can be shown that the transfer function depends only on the reachable and observable parts of the system. Thus transfer functions reflect just the input/output or external behavior of the system, while the internal behavior remains hidden.

Remark 5.4.1 The poles and zeros of the transfer function play an important role in system analysis. This topic is not pursued in this paper. A full treatment of poles and zeros can be found in Kailath [8].

Remark 5.4.2 Note that by formally replacing the complex variable of the transfer function with the shift operator q, the resulting operator gives explicitly the relation between the input process u and the output process y. By an abuse of terminology, this operator is also called the transfer function of system (3).

5.5 State Feedback

In this chapter we discuss an important control design question: that of choosing a control law such that the dynamics of the resulting closed-loop system fulfills certain requirements. If we think of the system (3) as describing an economic system, then we should think of the input u as being chosen by an actuator, by someone who is trying to achieve a desired behavior of the state trajectory x through a judicious choice of the input trajectory u. In this context u is thought to be a variable that can be manipulated, and therefore we call it the *control*. In *feedback control* the value of the control input is chosen not as an explicit function of time but on the basis of an observed output.

In *state feedback* we assume that the state x is observed and thus the control is based on the observed state trajectory. To be more formal, the control law is a map F that associates with the observed state trajectory $x : \mathbb{Z}_+ \to \mathbb{R}^n$ the chosen control input $u : \mathbb{Z}_+ \to \mathbb{R}^m$. For obvious reasons, this map has to be non-anticipating, meaning that $(Fx)(t)$ may depend only on the values taken by x_s for $s \leq t$. The map F may be required to have other properties as well: linear, time-invariant, with or without memory etc.

In this chapter we consider the situation where linear, static state feedback is applied,

$$u_t = -Lx_t + v_t \ ,$$

i.e. when the control is linear and memoryless: the control at time t depends on the state at the present time t only. The term $v_t \in \mathbb{R}^m$ accounts for other (possibly external) input factors on which the control may depend (for example v_t could be a disturbance imposed by nature). The matrix L is usually referred to as the *feedback gain matrix*. Substituting (5.5) into equation (3) produces the closed-loop system

$$x_{t+1} = (F - GL)x_t + Gv_t.$$

In order to ensure desirable properties for the closed-loop system, we need to consider the resulting closed-loop system matrix $F - GL$. As seen in the previous chapters, some basic properties of system (3) are strongly related to the eigenvalues of the system matrix, which in turn are the roots of the characteristic polynomial $\chi_F(z) = \det(zI_n - F)$ of the matrix F ($\chi_F(z)$ is often termed as the characteristic polynomial of the system). This leads us to the following question:

Given a system in the form (3), what closed-loop characteristic polynomials are achievable by choosing an appropriate feedback gain matrix L?

The following celebrated result, known as the *eigenvalue assignment theorem* (or pole placement theorem) provides an interesting answer to the question.

Theorem 5.5.1 The pair (F, G) is reachable if and only if for any monic polynomial[1] $c(z) \in \mathbb{R}[z]$ of degree n there exists a constant matrix L such that $F - GL$ has characteristic polynomial $c(z)$.

Proof. See for example Kučera [11]. ■

Remark 5.5.2 Using the Kalman decomposition (the canonical form that puts the reachability structure into evidence) the above theorem can be refined so that it gives the complete answer to the eigenvalue assignment problem for systems that are not necessarily reachable. Readers are referred to Kučera [11].

5.6 State Prediction

In everyday applications the states of the system are not available: either they are not directly measurable or the system is disturbed by some noise. Thus the results of the previous chapter cannot be used. To alter the dynamics of the system, however, it is not always necessary to measure all the state variables: by appropriate signal processing, we are often able to obtain good estimates of all the state variables from the measured outputs. In this chapter we investigate the problem of estimating the state of a system disturbed by random noise.

Consider a linear stochastic system governed by the equations

$$\begin{aligned} x_{t+1} &= Fx_t + Gu_t + v_t \\ y_t &= Hx_t + Ju_t + w_t \end{aligned} \tag{5}$$

[1] A polynomial $c(z) \in \mathbb{R}[z]$ is said to be monic if its leading coefficient is 1.

where $t = 0, 1, \ldots$, $u_t \in \mathbb{R}^m$, $y_t \in \mathbb{R}^l$, $x_t \in \mathbb{R}^n$ and $v_t \in \mathbb{R}^n$, $w_t \in \mathbb{R}^l$ are uncorrelated random sequences with mean zero. These random sequences can be interpreted as additive state and output noises that corrupt the behavior of the system (3). We furthermore suppose that u_t is a linear function of the measured outputs $y_{t-1}, y_{t-2}, \ldots , y_0$.

Clearly, exact reconstruction of the state is impossible in this case. Therefore, we propose the so-called *linear-quadratic predictor*: our aim is to give an estimate $\hat{x}_{t|t-1}$ of the state x_t generated by a linear system from the measurements $y_{t-1}, y_{t-2}, \ldots , y_0$ such that the loss $\vartheta_{t|t-1} = E(x_t - \hat{x}_{t|t-1})^T V(x_t - \hat{x}_{t|t-1})$ is minimized for every positive semi-definite matrix V. In other words, we are looking for a vector in

$$\mathcal{H}^y_{t-1} := \{x \mid x = A_{t-1}y_{t-1} + \ldots + A_0 y_0,\ A_j \in \mathbb{R}^{n\times l} \text{ for } j \leq t-1\} \qquad (6)$$

which minimizes the loss. The quadratic measure of the covariance of the estimation error is used for theoretical convenience: it is the quadratic structure which guarantees that the linear-quadratic predictor is fairly easily obtained. A major property of linear-quadratic estimates is the *principle of orthogonality.*

Theorem 5.6.1 The vector in $\mathcal{H}^y_{t-1}$ that minimizes the quadratic loss $\vartheta_{t|t-1}$ for every $V \geq 0$ is the orthogonal projection of x_t on $\mathcal{H}^y_{t-1}$.

Proof. Let $\hat{x}_t$ denote the orthogonal projection of x_t on $\mathcal{H}^y_{t-1}$, i.e.

$$x_t = \hat{x}_t + \tilde{x}_t$$

where $\hat{x}_t \in \mathcal{H}^y_{t-1}$ and $\tilde{x}_t$ is orthogonal to $\mathcal{H}^y_{t-1}$,

$$E\tilde{x}_t y^T = 0 \quad \text{for every } y \in \mathcal{H}^y_{t-1} \ .$$

Now for $y \in \mathcal{H}^y_{t-1}$

$$\begin{aligned} E(x_t - y)^T V(x_t - y) &= E(\tilde{x}_t + \hat{x}_t - y)^T V(\tilde{x}_t + \hat{x}_t - y) = \\ &= E\tilde{x}_t^T V \tilde{x}_t + E(\hat{x}_t - y)^T V(\hat{x}_t - y) + \\ &+ E\tilde{x}_t^T V(\hat{x}_t - y) + E(\hat{x}_t - y)^T V \tilde{x}_t \ . \end{aligned}$$

Since $\tilde{x}_t$ is orthogonal to any vector of $\mathcal{H}^y_{t-1}$, hence also to $\hat{x}_t - y$, the last two terms vanish. Both remaining terms are non-negative: the first one is independent of y, the second can be made zero by taking $y = \hat{x}_t$. Thus the orthogonal projection $\hat{x}_t$ is the vector of $\mathcal{H}^y_{t-1}$ that minimizes the quadratic loss $\vartheta_{t|t-1}$ for every $V \geq 0$. ■

We now give a procedure by means of which the predicted state of system (5) can be easily calculated from the transfer function R relating the variables v and x.

Lemma 5.6.2 Suppose that the transfer function R connecting the variables v and x is causal with a non-singular constant term and R^{-1} exists and is also causal. Furthermore, assume that no output noise is present in the system. Then the predicted state process $\hat{x}$ is given by the equation

$$\hat{x} = (I - R^{-1})x \ . \tag{7}$$

Proof. Let the expansion of R take the form $R(z) = \sum_{i=0}^{\infty} R_i z^{-i}$, where $R_i \in \mathbb{R}^{n \times n}$ (in other words, $x_t = \sum_{i=0}^{\infty} R_i v_{t-i}$). Since R_0 is non-singular, we might as well assume $R_0 = I$: consider simply the representation $x = R'v'$ with $R' = RR_0^{-1}$ and $v' = R_0 v$ (note that v' is again an uncorrelated random sequence with mean zero). Similarly to (6) we define $\mathcal{H}_t^v$ and $\mathcal{H}_t^x$ as the linear subspaces of $\mathbb{R}^n$ spanned by the random vectors v_i, $i \leq t$ and x_i, $i \leq t$ respectively. From the causality of R^{-1} it follows immediately that $\mathcal{H}_t^x = \mathcal{H}_t^v$ for any $t \in \mathbb{Z}_+$. Using our assumptions that $w_t = 0$ for all $t \in \{0, 1, \dots\}$, the second equation of (5) implies that $\mathcal{H}_t^y = \mathcal{H}_t^x$ for any t. Therefore, if the orthogonal projection of x on a linear subspace $\mathcal{A}$ is denoted by $E^*(x|\mathcal{A})$, then

$$\begin{aligned}
\hat{x}_t &= E^*(x_t|\mathcal{H}_{t-1}^y) = E^*(x_t|\mathcal{H}_{t-1}^v) = E^*\Big(\sum_{i=0}^{\infty} R_i v_{t-i}| < v_{t-1}, \dots, v_0 > \Big) = \\
&= \sum_{i=1}^{\infty} R_i v_{t-i} = \sum_{i=0}^{\infty} R_i v_{t-i} - R_0 v_t = (Rv)_t - v_t = ((R-I)v)_t = \\
&= ((R-I)R^{-1}x)_t = ((I - R^{-1})x)_t,
\end{aligned}$$

which was to be proved. ■

Remark 5.6.3 It can be shown that $\hat{x}_t$ can be implemented by a linear system similar to the original system (3) but one which is driven by a measure of the prediction error. Such systems are often called *state observers*. An elegant recursive algorithm for computing the state estimate is given by the celebrated *Kalman filter* which finds applications in signal processing, system identification, stochastic control and many other areas. The interested reader is referred to the original work of Kalman [9].

5.7 Output Feedback

Now we consider the case when the *complete* state vector cannot be directly measured, only the output y_t. Then, instead of state feedback, we can apply only output feedback.

First consider the case when system (3) is uncontrolled, i.e. $G = 0,\ J = 0$ and apply a linear static output injection of the form $-Ky_t$ for some matrix K. The closed-loop system equations obtained this way are

$$\begin{aligned} x_{t+1} &= (F - KH)x_t \\ y_t &= Hx_t \ . \end{aligned}$$

The dual of theorem 5.5.1 provides a mathematical link between observability and output feedback.

Theorem 5.7.1 The pair (F, H) is observable if and only if for any monic polynomial $c(z) \in \mathbb{R}[z]$ of degree n there exists a constant matrix K such that $F - KH$ has characteristic polynomial $c(z)$.

Proof. Consider the pair (F^T, H^T) and apply theorem 5.5.1. ■

Now consider the case of output injection in the controlled system

$$\begin{aligned} x_{t+1} &= Fx_t + Gu_t \\ y_t &= Hx_t + Ju_t \ . \end{aligned}$$

Since the value of y_t is not available when u_t is to be calculated, the simplest output feedback that can be implemented takes the form

$$u_{t+1} = -Ly_t$$

with a constant matrix L. Using this type of feedback, however, it is not possible to assign the eigenvalues of the closed-loop system

$$\begin{bmatrix} x_{t+1} \\ u_{t+1} \end{bmatrix} = \begin{bmatrix} F & G \\ -LH & -LJ \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix}$$

at will even if reachability and observability hold. What's more, the following example shows that the closed-loop system may fail to be stable for any L.

Example. Consider system (3) with $F = 2,\ G = 1,\ H = 1$ and $J = 0$. The resulting eigenvalues of the closed-loop system are $\lambda_1 = 1 + \sqrt{1 - L}$ and $\lambda_2 = 1 - \sqrt{1 - L}$. In case of $L \leq 1$ we get $|\lambda_1| \geq 1$; if $L > 1$ then

$|\lambda_1| = |\lambda_2| = \sqrt{L} > 1$. Therefore, even though this system is obviously reachable and observable, it cannot be stabilized for any $L \in \mathbb{R}$.

To solve this puzzle, we need to use feedback laws with memory. One way is to reconstruct the state and then apply linear static feedback in the reconstructed variable. The output feedback obtained this way is itself a dynamical system, often called a *compensator*. The design of a feedback compensator is based on the combination of two principles: the separation principle and the certainty equivalence. The *separation principle* states that the design of the state observer and of the controller gains are carried out separately. The observer produces an estimate of the state and the feedback compensator uses this estimate to compute the control action. The *certainty principle* states that for the control action we proceed as if the estimate were equal to the exact value of the state: the controller acts equivalently as if it were certain of the value of the state.

Using these ideas for system (3), all we have to do is to apply a linear static feedback of the form

$$u_t = -L\hat{x}_t \ .$$

By remark 5.6.3 the estimate of the state can be provided by an observer of the form

$$\hat{x}_{t+1} = F\hat{x}_t + Gu_t + K(y_t - H\hat{x}_t - Ju_t) \ .$$

The observer consists of a copy of system (3) with an extra driving input $K(y_t - H\hat{x}_t - Ju_t) = KH(x_t - \hat{x}_t)$ proportional to the estimation error. The resulting closed-loop system is described by the system equations

$$\begin{aligned} x_{t+1} &= Fx_t - GL\hat{x}_t \\ \hat{x}_{t+1} &= KHx_t + (F - KH - GL)\hat{x}_t \ . \end{aligned}$$

In terms of the estimation error $e_t := x_t - \hat{x}_t$, these equations can be given a more convenient form,

$$\begin{bmatrix} x_{t+1} \\ e_{t+1} \end{bmatrix} = \begin{bmatrix} F - GL & GL \\ 0 & F - KH \end{bmatrix} \begin{bmatrix} x_t \\ e_t \end{bmatrix} .$$

Since this system matrix is upper block diagonal, the overall closed-loop characteristic polynomial equals the product of the characteristic polynomials of $F - GL$ and $F - KH$. Combining theorems 5.5.1 and 5.7.1, we get the following result.

Theorem 5.7.2 For any monic polynomials $c_1(z)$ and $c_2(z)$ of $\mathbb{R}[z]$, each of degree n, there exist matrices K and L such that the closed-loop characteristic polynomial is equal to $c_1(z)c_2(z)$ if and only if (F, G) is a reachable pair and (F, H) is an observable pair.

Remark 5.7.3 It is possible to avoid the factorizability restriction of the closed-loop characteristic polynomial, i.e. for any closed-loop characteristic polynomial of order $2n$ (not necessarily factorizable into two real factors of order n) a compensator exists if and only if reachability and observability hold.

6 Modeling Behavioral Phenomena

In this chapter using the modeling language and techniques described in the previous chapters, we set up a few simple models for some of the behavioral phenomena affecting financial markets.

Throughout the chapter the following terminology and notations are used. The price of a given stock at time $t \in \mathbb{Z}_+$ is denoted by p_t, the market demand for this stock is denoted by d_t. If we consider a portfolio with n stocks then the above defined terms are multi-dimensional vectors in $\mathbb{R}^n$, their i^{th} component expressing the price of or demand for the i^{th} stock in the stock market. In the case when more than one actuator is present in the system, say m, the demand process of the k^{th} person is denoted by d_t^k. The stochastic disturbances in the system are denoted by e_t.

Naturally, all components of the price process p_t should be non-negative. However, if we think of the price of a stock as a measure of the profitability of the company that issues it, then a negative price could mean that the company is unprofitable. Therefore, we assume that the price process p_t of a stock can take any values in $\mathbb{R}$. The components of the demand sequence d_t can also take any values in $\mathbb{R}$: this time a negative value means that the broker would like to get rid of the stock concerned. (We implicitly assumed that stocks are infinitely divisible: any amount can be purchased or sold.)

The prices are given by the market. In general, the stock market is modeled as a dynamical system Σ_P that generates prices. In this paper prices are taken to depend on the past and present values of the demand process and on the current stochastic disturbances, i.e.

$$p_t = P(d_s^k,\ s \leq t,\ 1 \leq k \leq m,\ e_t)$$

for some function P defined on the appropriate space.

The models belonging to different behavioral phenomena are denoted by Σ_{B_j}, where j stands for the given behavioral phenomenon. These behavior models produce demands that are assumed to depend on the past values of the price and demand processes. Thus a generated demand at time t is given by a function

$$B_j(d_s^k,\ s < t,\ 1 \leq k \leq m,\ p_r,\ r < t)$$

defined on the appropriate signal space.

Now we present a simple model for rational investors. Suppose a broker at time t is trying to figure out how much of a given stock he is willing to buy. For this purpose, taking into consideration all the relevant information available, he makes an estimate of the future price, denote it by $\hat{p}_{t+1}$. Now, a reasonable broker would buy more of the stock whenever his estimate $\hat{p}_{t+1}$ is greater than the current price p_t, and would buy less if it is the other way. Thus a "rational" behavior could be described for example by the equation

$$d_t - \alpha_t d_{t-1} = B_t \ \mathrm{sgn}(\hat{p}_{t+1} - p_t - \delta_t)$$

where $\delta_t \geq 0$ is a threshold value, $B_t \in \mathbb{Z}_+$ is the number of stocks the broker wants to purchase and $\alpha_t \in \mathbb{R}$, $|\alpha_t| \approx 1$ is a parameter expressing the faith of the broker in the past observations. To get a mathematically more tractable model, we assume that this behavior is linear and time-invariant:

$$d_t - \alpha d_{t-1} = B(\hat{p}_{t+1} - p_t - \delta) \tag{8}$$

Now we give some possible models for the behavioral phenomena encountered in Chapter 2.

Overconfidence: the broker believes that he possesses the true market model Σ_P. Thus disregarding the disturbances, he can calculate the exact future price of the stock.

Anchoring/conservatism: the broker does not take into consideration the last r stock prices since he clings excessively to his prior beliefs. Thus $\hat{p}_{t+1}$ is a function of p_{t-r}, $p_{t-r-1}, \ldots$ only. As a special case, when the broker is anchored to the price trend, this function might depend only on the differences of the consecutive stock prices (that is, on $p_{t-r} - p_{t-r-1}$, $p_{t-r-1} - p_{t-r-2}$ and so on).

Remark 6.0.4 In this case the predicting procedure set forward in chapter 5.6 can be modified as follows. Let H denote the causal transfer function relating the processes e and p. Then supposing again invertibility and causality of H^{-1}, we proceed in exactly the same way:

$$\begin{aligned} \hat{p}_t &= E^*(p_t|\mathcal{H}^p_{t-r}) = E^*(p_t|\mathcal{H}^e_{t-r}) = \\ &= E^*\Big(\sum_{i=0}^{\infty} H_i e_{t-i}| < e_{t-r}, \ldots, e_0 > \Big) = \sum_{i=r}^{\infty} H_i e_{t-i} \ . \end{aligned}$$

To express this in a more compact way, we introduce a notation: for $f(z) = \sum_{i=-\infty}^{\infty} f_i z^i$ let $[f(z)]_- := \sum_{i=-\infty}^{0} f_i z^i$. Using this notation it is easily seen that

$$\hat{p} = q^{-r}[q^r H(q)]_- H^{-1} p$$

where q denotes the unit delay (or shift) operator.

Representativeness heuristic: a person affected by this phenomenon calculates his estimated stock price $\hat{p}_{t+1}$ based only on the latest l stock price data (i.e. on $p_t, p_{t-1}, \ldots p_{t-l+1}$) and leaves the previous prices completely out of consideration. In the extreme case, he might only consider the difference between the last two prices thus arriving at the estimate $\hat{p}_{t+1} = p_t + (p_t - p_{t-1})$, which yields the model

$$d_t - \alpha d_{t-1} = B(p_t - p_{t-1} - \delta) \ .$$

Rationalizing: the decision maker explicitly formulates one model and thinks that his model is a good approximation in the sense that the stock market data are generated by a model Σ_P that belongs to a set of models $\mathcal{P}$ constituting of the perturbations of the approximating model. When encountering data that contradicts his model, he simply switches to another model in $\mathcal{P}$ within which the existing data are no longer contradictory.

Remark 6.0.5 The possibility of model misspecification could lead the decision maker to prefer decision rules that work well over that set of nearby models. This leads us to the concept of *robustness*, i.e. the integrity of the control actions against model parameter variations. Robust control, the branch of system theory concerned with this problem has been rapidly developing in the past few years. For applications of robust control in the field of economics, see for example Sargent [6].

Framing: in this case, for some reason the agent does not see the real price of the stock, just a function of it. However, this function may hide some of the important aspects of the price process.

Cognitive overload: the broker uses extremely simple models, e.g. he supposes that the price of the stock will not change, $\hat{p}_{t+1} = p_t$. Substituting this into equation (8) yields $d_t - \alpha d_{t-1} = -B\delta$. Since this amount is non-positive, it means that the agent sells his stocks. This, by the way, is understandable given his assumptions since no one would keep a stock that earns no interest (correction: except for those affected by loss aversion and/or commitment pride).

Habits/conventions: now time-invariance of the market model Σ_P is supposed. Since the dynamics of the stock price process may change over time, the above assumption may be unreal.

Loss aversion: this phenomenon can be simply modeled by putting $\hat{p}_{t+1} := \max(p_t + \delta, \hat{p}^r_{t+1})$ where $\hat{p}^r_{t+1}$ denotes the original (unbiased) estimate of the agent. This way the agent will keep his downgraded stocks even when the original model (8) would urge him to sell them.

Commitment pride: the model for this phenomenon can be given by the original model (8) with the constraint $\hat{p}_{t+1} \geq p_t + \delta$. This inequality ensures that once the demand for the particular stock becomes positive, from then on it remains non-negative for ever.

Investors profiling: each group of investors is represented by an appropriate behavioral model. The aggregate demand for the given stock is then given by $d_t := \sum_{i=1}^{m} \lambda_i d_t^i$ where $\lambda_i \geq 0$, $\sum_{i=1}^{m} \lambda_i = 1$ are weights representing the ratio of the given group compared to all the investors.

Dominant investors: this can be viewed as a special case of investors profiling when all the groups are governed by essentially the same equations, $d_t^i \approx d_t'$ for all i. The values d_t' are of course determined by the dominant investors.

7 Closing The Loop

The two dynamical systems of the previous chapter, the market and the behavior are in close relation with each other: they are interconnected. Closing the loop constrains the laws of the market and it is the behavior of this interconnected system that we are interested in. This can be formalized in the same spirit as in definition 4.0.6.

Definition 7.0.6 Let $\Sigma_1 = (\mathbb{T}, \mathbb{W}, \mathfrak{B}_1)$ and $\Sigma_2 = (\mathbb{T}, \mathbb{W}, \mathfrak{B}_2)$ be two dynamical systems with the same time axis and the same signal space. The *interconnection* of Σ_1 and Σ_2 is defined as the dynamical system $\Sigma_1 \wedge \Sigma_2 := (\mathbb{T}, \mathbb{W}, \mathfrak{B}_1 \cap \mathfrak{B}_2)$.

A prototype of the closed-loop system arising by interconnecting the market and the behavior models is shown schematically in the following figure[1].

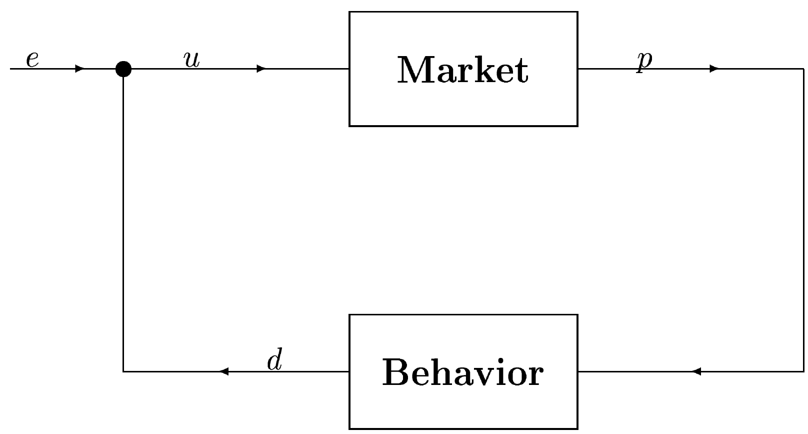

Figure 1: The Closed-Loop System.

7.1 The General Model

The market Σ_P is viewed as a black box, which is given by its transfer function P relating the input process u to the output process p (both take values in $\mathbb{R}^n$):

$$Pu = p \tag{9}$$

[1]The arrows in the figure show the direction of the signal flow, while the solid circle denotes summation.

Here P is assumed to be a causal linear operator of the form $P = \sum_{i=0}^{\infty} P_i q^{-i}$, where $P_i \in \mathbb{R}^{n \times n}$ and $q^{-i},\ i \in \mathbb{Z}$ denotes the backward i-shift: $(q^{-i}u)_t := u_{t-i}$.

The dynamical system Σ_B has inputs $p_t \in \mathbb{R}^n$ and outputs $d_t \in \mathbb{R}^n$. Its behavior is described by two equations. The first one gives the one-step price estimate of the stocks in the portfolio:

$$\hat{p}_t = (Mp)_t \tag{10}$$

where M is assumed to be a strictly causal linear predictor of the form $M = \sum_{i=1}^{\infty} M_i q^{-i},\ M_i \in \mathbb{R}^{n \times n}$. It is important to note that the summation in the above formula starts from $i = 1$, hence the name *predictor*.
The second equation takes the form

$$d_t - \alpha d_{t-1} = B(\hat{p}_t - p_{t-1}) \tag{11}$$

where $B \in \mathbb{R}^{n \times n}$ and $\alpha \in \mathbb{R},\ \alpha \neq 0,\ |\alpha| < 1$ is a known parameter that captures the memory of the agent (typically α is thought to be close to 1).

The interconnection of the two systems is then given by the equation

$$d_t + e_t = u_t \tag{12}$$

where the coordinates of e_t are uncorrelated random sequences with mean zero.

The interconnected system is visualized in Figure 2.

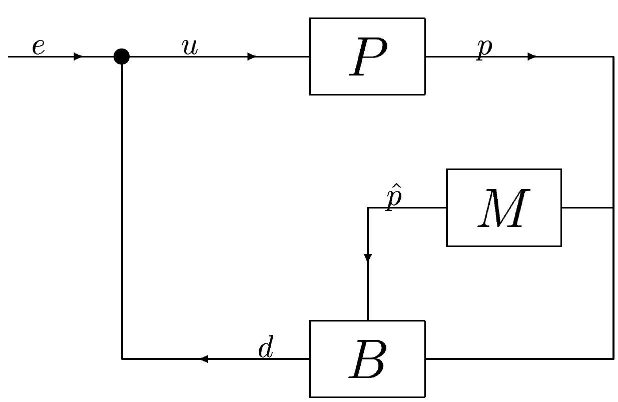

Figure 2: The General Model.

Remark 7.1.1 Note that while B and M are known by the agent, the transfer function P governing the stock market remains unknown.

Remark 7.1.2 A slight generalization of this model arises when $\alpha \in \mathbb{R}^{n\times n}$, $\alpha \neq 0$ is a matrix with $||\alpha|| < 1$. This expresses the fact that the demand of one stock may depend on the previous demands of *other* stocks in the portfolio (in other words, there is interdependence across the stocks).

Our first aim will be to determine an optimal linear predictor in the closed-loop system. In the sequel, we will need the following lemma.

Lemma 7.1.3 Let $A \in \mathbb{R}^{n\times n}$, $||A|| < 1$. Then the inverse of the operator $I - Aq^{-1}$ exists and is given by the following formula:

$$(I - Aq^{-1})^{-1} = \sum_{i=0}^{\infty} A^i q^{-i} \; . \tag{13}$$

Proof. Consider the above defined sum. Since the norm of the shift operator is 1 and $||A|| < 1$, it follows that $||\sum_{i=0}^{\infty} A^i q^{-i}|| \leq \sum_{i=0}^{\infty} ||A||^i < \infty$ and therefore the sum is convergent. Now

$$\begin{aligned}(I - Aq^{-1})\sum_{i=0}^{\infty} A^i q^{-i} &= \sum_{i=0}^{\infty} A^i q^{-i}(I - Aq^{-1}) = \\ &= \sum_{i=0}^{\infty} A^i q^{-i} - \sum_{i=0}^{\infty} A^{i+1} q^{-(i+1)} = A^0 q^0 = I\end{aligned}$$

proves our claim. ∎

As a first step, we calculate the transfer function H relating the variables e and p in the closed-loop system. Using our notations, (11) is equivalent to

$$(I - \alpha q^{-1})d = B(M - q^{-1})p \; .$$

Since $|\alpha| < 1$, in view of the previous lemma, we may multiply both sides of the equation by the inverse of $(I - \alpha q^{-1})$. Therefore

$$d = (I - \alpha q^{-1})^{-1} B(M - q^{-1})p \; .$$

Plugging this and (12) into the transfer function P, the following results:

$$\begin{aligned}p = Pu = P(d + e) &= \\ &= P(I - \alpha q^{-1})^{-1} B(M - q^{-1})p + Pe\end{aligned}$$

from which

$$[I - P(I - \alpha q^{-1})^{-1}B(M - q^{-1})]p = Pe$$
$$p = [I - P(I - \alpha q^{-1})^{-1}B(M - q^{-1})]^{-1}Pe = He$$

with $H = H(M) = [I - P(I - \alpha q^{-1})^{-1}B(M - q^{-1})]^{-1}P$. Here we assumed that the inverse of the operator $I - P(I - \alpha q^{-1})^{-1}B(M - q^{-1})$ exists; it is furthermore assumed that H is a causal operator, $H = \sum_{i=0}^{\infty} H_i q^{-i}$ with $H_0 := I$.

By theorem 5.6.1 the optimal linear predictor $\hat{p}_t$ of p_t is the orthogonal projection of p_t on $\mathcal{H}^p_{t-1}$. Supposing that H^{-1} exists and is causal and using lemma 5.6.2 we get that $\hat{p} = (I - H^{-1})p$. Comparing this with (10), we conclude that the optimal linear predictor (if it exists!) is a root of the equation

$$M = I - H^{-1}(M) \ . \tag{14}$$

Using the formula for H, this can be rewritten as

$$M = I - P^{-1}[I - P(I - \alpha q^{-1})^{-1}B(M - q^{-1})]$$
$$M = I - P^{-1} + (I - \alpha q^{-1})^{-1}B(M - q^{-1})$$

from which

$$[I - (I - \alpha q^{-1})^{-1}B]M = I - P^{-1} - q^{-1}(I - \alpha q^{-1})^{-1}B,$$

where use has been made of the fact that q^{-1} commutes with any linear operator. Thus we arrive at the following theorem:

Theorem 7.1.4 Assume that M is an optimal linear predictor in the system given by equations (9) - (12) and $H := [I - P(I - \alpha q^{-1})^{-1}B(M - q^{-1})]^{-1}P$ is an invertible causal operator with $H_0 = I$ and H^{-1} causal. Then M satisfies the equation

$$[I - (I - \alpha q^{-1})^{-1}B]M = I - P^{-1} - q^{-1}(I - \alpha q^{-1})^{-1}B \ . \tag{15}$$

On the other hand, if M is a linear predictor satisfying equation (15), then M is optimal.

Remark 7.1.5 It is worthwhile to notice that the interconnection of the market and behavior models is in fact nothing more than feeding back the price process p into the market Σ_P. Indeed, using the notation $K := K(B, M) := (I - \alpha q^{-1})^{-1}B(M - q^{-1})$, we get $d = Kp$. This value is then fed back into the market: $u = d + e = Kp + e$. However, this is not a static feedback in the sense that is uses past values of the price process as well (in other words, the feedback gain operator K contains in its expansion at least one power q^{-i} for some $i \geq 1$ with non-zero coefficient). This is obvious since K is the product of a causal operator and an operator that has only negative powers of q.

If K represented a linear static feedback in the *estimate* $\hat{p}_t$, i.e. $d_t = (Kp)_t = L\hat{p}_t$ for a fixed matrix $L \in \mathbb{R}^{n \times n}$, then the results of chapter 5.7 could be readily used. This property holds if and only if the coefficients of the powers of the shift operator are the same on both sides of the equation. The interested reader may check that this is the case if for example $M = \sum_{i=1}^{\infty} M_i q^{-i}$ with $M_i = c\alpha^{i-1}(1-c)^{i-1} I$ for some constant $0 < c < 1$. (This predictor, by the way, expresses a very rational behavior: the older the stock price is, the less influence it should have on the current price estimate. In this particular case, the weights of the previous stock prices are exponentially decaying.)

Remark 7.1.6 In view of theorem 7.1.4, for a given P the optimal linear predictor can be explicitly obtained by solving equation (15) for M. However, we have already noted that the agent does not know the true market model and therefore he cannot use the above mentioned method. Instead, he might proceed as follows: first he makes an estimate $\hat{P}_1$ of the transfer function P from the available stock market data. Then using equation (15) he calculates the optimal linear predictor $\hat{M}_1$ with respect to $\hat{P}_1$. In the following step, using the newly generated data he updates his estimate of the market model, thus getting $\hat{P}_2$ which in turn determines $\hat{M}_2$ and so on. This method is known by control theorists as *indirect adaptive control.*

Another approach would be to consider the function

$$f(M) = I - P^{-1} + (I - \alpha q^{-1})^{-1} B(M - q^{-1})$$

for some initial value $\hat{M}_0$ and begin iterating: $\hat{M}_k := f(\hat{M}_{k-1})$ for $k \in \mathbb{Z}_+$. We assume that the value of the function f at $\hat{M}_k$ can somehow be calculated even when the transfer function P is unknown (for example think of somehow simulating the stock market). We could then investigate whether the so-obtained sequence is convergent. In case it is, the limit is a fixed point of function f and therefore it must be an optimal linear predictor. (Note that this logic functions not unlike what happens in daily life: based on past stock market data and past experience, the broker has an initial price predictor. If the observed new prices do not agree with his estimated prices, then this causes him to update his predictor.)

Let us now consider the special case when the price process obeys equation

$$p_t - Ap_{t-1} = u_t$$

where $A \in \mathbb{R}^{n \times n}$, $||A|| < 1$, $A \neq 0$ and $B = cI$ for some $c > 0$, $c \in \mathbb{R}$. The matrix A captures the relationship among the different stock prices, while the parameter c can be interpreted as a measure of the agent's risk sensitivity.

First let us calculate the transfer function P relating the processes u and p explicitly. From $(I - Aq^{-1})p = u$ it follows immediately that $P = (I - Aq^{-1})^{-1}$, where lemma 7.1.3 is used to verify that this operator exists. Note that P is a causal operator in the sense of definition 4.2.3. Thus equation (15) takes the form

$$[I - c(I - \alpha q^{-1})^{-1}]M = Aq^{-1} - cq^{-1}(I - \alpha q^{-1})^{-1} .$$

We would like to take the inverse of the operator appearing next to M on the left-hand side of the equation. Assuming that $c \neq 1$ and using lemma 7.1.3

$$\begin{aligned}
&\{I - c(I - \alpha q^{-1})^{-1}\}^{-1} = \{(I - \alpha q^{-1})^{-1}[(I - \alpha q^{-1}) - cI]\}^{-1} = \\
&= [(1-c)I - \alpha q^{-1}]^{-1}(I - \alpha q^{-1}) = \\
&= \left[(1-c)\left(I - \frac{\alpha}{1-c}q^{-1}\right)\right]^{-1}\left(I - \alpha q^{-1}\right) = \\
&= \left[\frac{1}{1-c}\sum_{i=0}^{\infty}\left(\frac{\alpha}{1-c}\right)^i q^{-i}\right]\left(I - \alpha q^{-1}\right) = \\
&= \frac{1}{1-c}\sum_{i=0}^{\infty}\left(\frac{\alpha}{1-c}\right)^i q^{-i} - \sum_{i=0}^{\infty}\left(\frac{\alpha}{1-c}\right)^{i+1} q^{-(i+1)} = \\
&= \frac{1}{1-c} + \sum_{j=1}^{\infty}\beta_j q^{-j}
\end{aligned}$$

for some coefficients β_j. Thus

$$M = \left(\frac{1}{1-c} + \sum_{j=1}^{\infty}\beta_j q^{-j}\right)\left(Aq^{-1} - cq^{-1}(I - \alpha q^{-1})^{-1}\right)$$

$$M = \left(\frac{1}{1-c} + \sum_{j=1}^{\infty}\beta_j q^{-j}\right)\left(Aq^{-1} - c\sum_{i=1}^{\infty}\alpha^{i-1}q^{-i}\right) .$$

Hence M has indeed the desired form: the product of a causal operator and an operator that has only negative powers of q is indeed strictly causal. Note that when using lemma 7.1.3, we assumed that $|\frac{\alpha}{1-c}| < 1$. Since $c > 0$, this implies that the manipulations carried out above are legal if $0 < c < 1 - |\alpha|$ or $c > 1 + |\alpha|$. Thus we have proved the following theorem:

Theorem 7.1.7 Consider a system given by equations (9) - (12) with market transfer function $P = (I - Aq^{-1})^{-1}$, $A \in \mathbb{R}^{n \times n}$, $A \neq 0$, $||A|| < 1$. Then the behavior given by $B = cI$ with $0 < c < 1 - |\alpha|$ or $c > 1 + |\alpha|$ is realizable, i.e. it admits an optimal linear predictor M in the form $M = \sum_{i=1}^{\infty} M_i q^{-i}$.

Remark 7.1.8 It is interesting to see that the behaviors $B = cI$ for values of c close to 1 (which represent, by the way, a very natural behavior) cannot be modeled using our framework.

Remark 7.1.9 Note the stunning fact that the simplest behavior $B = I$ cannot be replicated for any α! Indeed, when $c = 1$ our calculations take the form

$$\begin{aligned} &\{I - (I - \alpha q^{-1})^{-1}\}^{-1} = \{(I - \alpha q^{-1})^{-1}[(I - \alpha q^{-1}) - I]\}^{-1} = \\ &= (-\alpha q^{-1})^{-1}(I - \alpha q^{-1}) = \\ &= -\frac{1}{\alpha} q (I - \alpha q^{-1}) = I - \frac{1}{\alpha} q \end{aligned}$$

thus yielding

$$M = \left(I - \frac{1}{\alpha} q\right)\left(Aq^{-1} - \sum_{i=1}^{\infty} \alpha^{i-1} q^{-i}\right) = \frac{1}{\alpha}(I - A) + \sum_{j=1}^{\infty} \gamma_j q^{-j}$$

for some coefficients γ_j. Since $A \neq I$, the resulting operator M is *not* a predictor. To deal with this peculiarity, we introduce multi-agent models.

7.2 A Multi-Agent Model

In this section we consider a model with two agents. The market dynamics is assumed to be the same as in the previous section. The behavior of the agents are characterized by a triple (M_i, B_i, α_i) for $i \in \{1, 2\}$, where our notation and assumptions follow that of the previous section. Thus Σ_{B_i} $(i = 1, 2)$ is given by the following equation:

$$d_t^i - \alpha_i d_{t-1}^i = B_i((M_i p)_t - p_{t-1}) \ .$$

The aggregate demand is taken to be

$$d_t := \lambda_1 d_t^1 + \lambda_2 d_t^2$$

where $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$. This model is shown in the next figure.

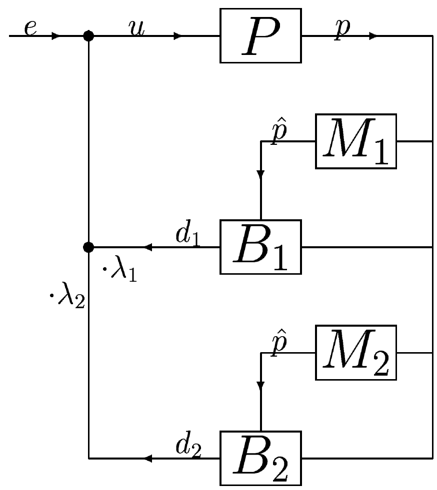

Figure 3: A Model With Two Agents.

Using the same argument as above, it can be shown that the optimal linear predictors M_1 and M_2 of the closed-loop system satisfy the equation

$$M_i = I - H^{-1}(M_1, M_2)$$

where

$$H^{-1}(M_1, M_2) := P^{-1} - \lambda_1(1 - \alpha_1 q^{-1})^{-1} B_1(M_1 - q^{-1}) - \\ - \lambda_2(1 - \alpha_2 q^{-1})^{-1} B_2(M_2 - q^{-1}) \; .$$

In the special case when $P = (I - Aq^{-1})^{-1}$, $\alpha_1 = \alpha_2$ and $M_1 = M_2$, $B_i = c_i I$ $(i = 1, 2)$, we get precisely the single-agent model with parameters α_1, M_1, and $B = (\lambda_1 c_1 + \lambda_2 c_2)I$. Thus by theorem 7.1.7 these behaviors can be replicated if $0 < \lambda_1 c_1 + \lambda_2 c_2 < 1 - |\alpha_1|$ or $\lambda_1 c_1 + \lambda_2 c_2 > 1 + |\alpha_1|$.

Remark 7.2.1 Note that in this setting we are already capable of replicating the behavior corresponding to $B = I$. Take for example $c_1 = 1$, $\lambda_1 = \lambda_2 = 0.5$, $|\alpha_1| = |\alpha_2| \approx 1$. Then choosing $c_2 > 1 + 2|\alpha_1|$ satisfies the required conditions. Thus, strange as it may seem, a very hectic participant may help to model the behavior of other actuators.

8 Concluding remarks

In this paper we presented a system theoretic approach to behavioral finance. Using the framework of discrete-time linear systems, we managed to establish a link between the fields of system theory and behavioral finance. While we gave some simple models for a number of behavioral phenomena, we investigated in detail just the general model. The general result of theorem 7.1.7 could be refined in many ways, for example we could examine under what circumstances a predictor of the form $M = \sum_{i=k}^{l} M_i q^{-i}$ with $k, l \in \mathbb{Z}_+$ could be optimal in the closed-loop system. For certain values of k and l this would lead to the characterization of some behavioral phenomena, e.g. that of representativeness heuristic and anchoring. Another point not addressed here is the iterating procedure mentioned in remark 7.1.6, which should prove useful in applications. It would also be worthwhile to examine the case when the market model is not stationary, i.e. the transfer function P itself changes during the time (we may think for example of different models describing the stock market during boom periods and during recession). Yet another possibility to extend our approach would be to consider the matrix A of the market model in 7.1.7 not as given, but as an element of a set $\mathcal{A}$ which consists of the perturbations of the decision maker's approximating model. Then our aim could be to choose a predictor that is optimal in some sense within the set of those nearby models. Finally, we should not forget that linear models were chosen mainly because of their relatively "simple" structure. The linear models considered in this paper should therefore serve as a model for the non-linear case.

References

[1] CAINES, PETER E.: *Linear Stochastic Systems.* Wiley Series in Probability and Mathematical Statistics, New York, 1988.

[2] CARLSON, BRUCE A. and FREDERICK, DEAN K.: *Linear Systems in Communication and Control.* John Wiley & Sons, New York, 1971.

[3] DEISTLER, M. and HANNAN, E. J.: *The Statistical Theory of Linear Systems.* John Wiley & Sons, New York, 1988.

[4] GOLOMB, S. W.: *Mathematical Models - Uses and Limitations.* Astronautics and Aeronautics, 6, 57-59, January 1968.

[5] GREENFINCH, PETER: *Behavioral Finance Definitions.* http:// //perso.wanadoo.fr/pgreenfinch/bfdef.htm, 2001.

[6] HANSEN, LARS PETER and SARGENT, THOMAS J.: *Robust Control and Filtering for Macroeconomics.* To be published, ftp://zia.stanford.edu/pub/sargent/webdocs/research/rgamesb.pdf, 2002.

[7] KAHNEMANN, D. and TVERSKY, A.: *Rational Choice and the Framing of Decisions.* Journal of Business, Vol. 59, 1986.

[8] KAILATH, T.: *Linear Systems.* Prentice Hall, Engelwood Cliffs, New Jersey, 1980.

[9] KALMAN, R. E.: *A New Approach to Linear Filtering and Prediction Problems.* Transactions of the ASME, Journal of Basic Engineering, 82D, 35-45, 1960.

[10] KOSTOLANY, ANDRÉ: *Tőzsdepszihológia.* Perfekt Pénzügyi Szakoktató és Kiadó Rt., 2000.

[11] KUČERA, VLADIMÍR: *Analysis and Design of Discrete Linear Control Systems.* Prentice Hall International, University Press, Cambridge, 1991.

[12] LUENBERGER, DAVID G.: *Investment Science.* Oxford University Press, New York, Oxford, 1998.

[13] MICHALETZKY, GYÖRGY: *Rendszerelmélet.* University lecture notes, http://www.math.elte.hu/probability/michaletzky/system99.dvi, 1999.

[14] POLDERMANN, JAN WILLEM and WILLEMS, JAN C.: *Introduction to Mathematical Systems Theory: A Behavioral Approach.* Springer Verlag, New York, 1998.

[15] POULARIKAS, ALEXANDER D. and SEELY, SAMUEL: *Signals and Systems.* Pws-Kent Publishing Company, Boston, 1991.

[16] RUGH, WILSON J.: *Linear System Theory.* Prentice Hall, Engelwood Cliffs, New Jersey, 1993.

[17] SHEFRIN, HERSH: *Behavioral Corporate Finance.* Working paper, Leavey School of Business, http://business.scu.edu/faculty//research/working_papers/workingpapers02.htm#0211WP, 2001.

[18] WILLEMS, J. C.: *On Interconnections, Control and Feedback.* IEEE Transactions on Automatic Control, 42:326-339, 1997.