Chapter 1

Introduction

The teaching philosophy of this book is that the best way to learn is by example. In that spirit, consider the following benchmark modeling problem from economic dynamics: At time t an economic agent receives income y_t. This income is split into consumption c_t and savings k_t. Savings is used for production, with input k_t yielding output

$$y_{t+1} = f(k_t, W_{t+1}), \qquad t = 0, 1, 2, \ldots \tag{1.1}$$

where $(W_t)_{t \geq 1}$ is a sequence of independent and identically distributed shocks. The process now repeats, as shown in figure 1.1. The agent gains utility $U(c_t)$ from consumption $c_t = y_t - k_t$, and discounts future utility at rate $\rho \in (0, 1)$. Savings behavior is modeled as the solution to the expression

$$\max_{(k_t)_{t \geq 0}} \mathbb{E} \left[\sum_{t=0}^{\infty} \rho^t U(y_t - k_t) \right]$$

$$\text{subject to } \; y_{t+1} = f(k_t, W_{t+1}) \text{ for all } t \geq 0, \text{ with } y_0 \text{ given}$$

This problem statement raises many questions. For example, from what set of possible paths is $(k_t)_{t \geq 0}$ to be chosen? And how do we choose a path at the start of time such that the resource constraint $0 \leq k_t \leq y_t$ holds at each t, given that output is random? Surely the agent cannot choose k_t until he learns what y_t is. Finally, how does one go about computing the expectation implied by the symbol $\mathbb{E}$?

A good first step is to rephrase the problem by saying that the agent seeks a savings *policy*. In the present context this is a map σ that takes a value y and returns a number $\sigma(y)$ satisfying $0 \leq \sigma(y) \leq y$. The interpretation is that upon observing y_t, the agent's response is $k_t = \sigma(y_t)$. Next period output is then $y_{t+1} = f(\sigma(y_t), W_{t+1})$, next period

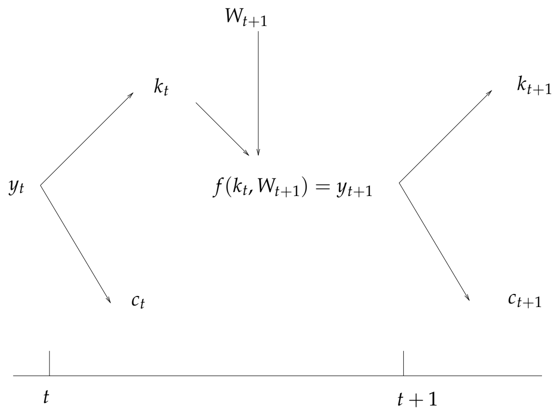

Figure 1.1 Timing

savings is $\sigma(y_{t+1})$, and so on. We can evaluate total reward as

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \rho^t U(y_t - \sigma(y_t))\right] \quad \text{where} \quad y_{t+1} = f(\sigma(y_t), W_{t+1}), \;\; \text{with } y_0 \text{ given} \tag{1.2}$$

The equation $y_{t+1} = f(\sigma(y_t), W_{t+1})$ is called a stochastic recursive sequence (SRS), or stochastic difference equation. As we will see, it completely defines $(y_t)_{t\geq 0}$ as a sequence of random variables for each policy σ. The expression on the left evaluates the utility of this income stream.

Regarding the expectation $\mathbb{E}$, we know that in general expectation is calculated by integration. But how are we to understand this integral? It seems to be an expectation over the set of nonnegative sequences (i.e., possible values of the path $(y_t)_{t\geq 0}$). High school calculus tells us how to take an integral over an interval, or perhaps over a subset of n-dimensional space $\mathbb{R}^n$. But how does one take an integral over an (infinite-dimensional) space of sequences?

Expectation and other related issues can be addressed in a very satisfactory way, but to do so, we will need to know at least the basics of a branch of mathematics called measure theory. Almost all of modern probability theory is defined and analyzed in terms of measure theory, so grasping its basics is a highly profitable investment. Only with the tools that measure theory provides can we pin down the meaning of (1.2).[1]

Once the meaning of the problem is clarified, the next step is considering how to solve it. As we have seen, the solution to the problem is a policy *function*. This is

[1] The golden rule of research is to carefully define your question before you start searching for answers.

rather different from undergraduate economics, where solutions are usually *numbers*, or perhaps *vectors* of numbers—often found by differentiating objective functions. In higher level applied mathematics, many problems have functions as their solutions.

The branch of mathematics that deals with problems having functions as solutions is called "functional analysis." Functional analysis is a powerful tool for solving real-world problems. Starting with basic functional analysis and a dash of measure theory, this book provides the tools necessary to optimize (1.2), including algorithms and numerical methods.

Once the problem is solved and an optimal policy is obtained, the income path $(y_t)_{t \geq 0}$ is determined as a sequence of random variables. The next objective is to study the dynamics of the economy. What statements can we make about what will "happen" in an economy with this kind of policy? Might it settle down into some sort of equilibrium? This is the ideal case because we can then make firm predictions. And predictions are the ultimate goal of modeling, partly because they are useful in their own right and partly because they allow us to test theory against data.

To illustrate analysis of dynamics, let's specialize our model so as to dig a little further. Suppose now that $U(c) = \ln c$ and $f(k, W) = k^\alpha W$. For this very special case, no computation is necessary: pencil and paper can be used to show (e.g., Stokey and Lucas 1989, §2.2) that the optimal policy is given by $\sigma(y) = \theta y$, where $\theta := \alpha\rho$. From (1.2) the law of motion for the "state" variable y_t is then

$$y_{t+1} = (\theta y_t)^\alpha W_{t+1}, \qquad t = 0, 1, 2, \ldots \tag{1.3}$$

To make life simple, let's assume that $\ln W_t \sim N(0, 1)$. Here $N(\mu, v)$ represents the normal distribution with mean μ and variance v, and the notation $X \sim F$ means that X has distribution F.

If we take the log of (1.3), it is transformed into the linear system

$$x_{t+1} = b + \alpha x_t + w_{t+1}, \text{ where } x_t := \ln y_t,\ w_{t+1} \sim N(0, 1), \text{ and } b := \alpha \ln \theta \tag{1.4}$$

This system is easy to analyze. In fact every x_t is normally distributed because x_1 is normally distributed (x_0 is constant and constant plus normal equals normal), and moreover x_{t+1} is normally distributed whenever x_t is normally distributed.[2]

One of the many nice things about normal distributions is that they are determined by only two parameters, the mean and the variance. If we can find these parameters, then we know the distribution. So suppose that $x_t \sim N(\mu_t, v_t)$, where the constants μ_t and v_t are given. If you are familiar with manipulating means and variances, you will be able to deduce from (1.4) that $x_{t+1} \sim N(\mu_{t+1}, v_{t+1})$, where

$$\mu_{t+1} = b + \alpha \mu_t \quad \text{and} \quad v_{t+1} = \alpha^2 v_t + 1 \tag{1.5}$$

[2] Recall that linear combinations of normal random variables are themselves normal.

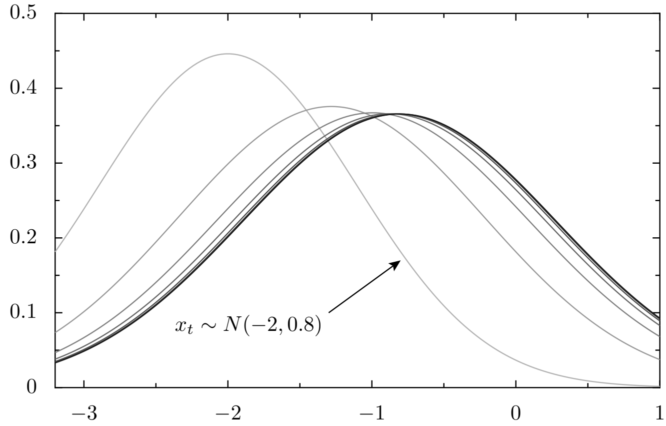

Figure 1.2 Sequence of marginal distributions

Paired with initial conditions μ_0 and v_0, these laws of motion pin down the sequences $(\mu_t)_{t \geq 0}$ and $(v_t)_{t \geq 0}$, and hence the distribution $N(\mu_t, v_t)$ of x_t at each point in time. A sequence of distributions starting from $x_t \sim N(-2, 0.8)$ is shown in figure 1.2. The parameters are $\alpha = 0.4$ and $\rho = 0.75$.

In the figure it appears that the distributions are converging to some kind of limiting distribution. This is due to the fact that $\alpha < 1$ (i.e., returns to capital are diminishing), which implies that the sequences in (1.5) are convergent (don't be concerned if you aren't sure how to prove this yet). The limits are

$$\mu^* := \lim_{t \to \infty} \mu_t = \frac{b}{1-\alpha} \quad \text{and} \quad v^* := \lim_{t \to \infty} v_t = \frac{1}{1-\alpha^2} \tag{1.6}$$

Hence the distribution $N(\mu_t, v_t)$ of x_t converges to $N(\mu^*, v^*)$.[3] Note that this "equilibrium" is a distribution rather than a single point.

All this analysis depends, of course, on the law of motion (1.4) being linear, and the shocks being normally distributed. How important are these two assumptions in facilitating the simple techniques we employed? The answer is that they are both critical, and without either one we must start again from scratch.

[3] What do we really mean by "convergence" here? We are talking about convergence of a sequence of *functions* to a given function. But how to define this? There are many possible ways, leading to different notions of equilibria, and we will need to develop some understanding of the definitions and the differences.

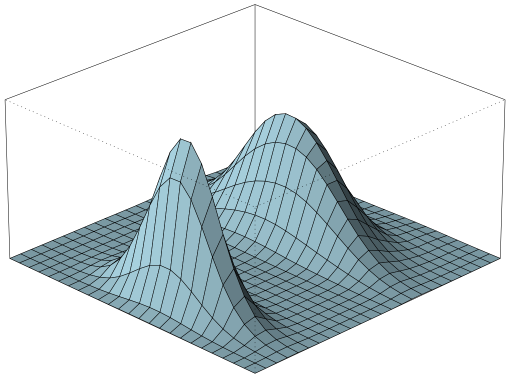

Figure 1.3 Stationary distribution

To illustrate this point, let's briefly consider the threshold autoregression model

$$X_{t+1} = \begin{cases} A_1 X_t + b_1 + W_{t+1} & \text{if } X_t \in B \subset \mathbb{R}^n \\ A_2 X_t + b_2 + W_{t+1} & \text{otherwise} \end{cases} \tag{1.7}$$

Here X_t is $n \times 1$, A_i is $n \times n$, b_i is $n \times 1$, and $(W_t)_{t \geq 1}$ is an IID sequence of normally distributed random $n \times 1$ vectors. Although for this system the departure from linearity is relatively small (in the sense that the law of motion is at least piecewise linear), analysis of dynamics is far more complex. Through the text we will build a set of tools that permit us to analyze nonlinear systems such as (1.7), including conditions used to test whether the distributions of $(X_t)_{t \geq 0}$ converge to some stationary (i.e., limiting) distribution. We also discuss how one should go about *computing* the stationary distributions of nonlinear stochastic models. Figure 1.3 shows the stationary distribution of (1.7) for a given set of parameters, based on such a computation.

Now let's return to the linear model (1.4) and investigate its sample paths. Figure 1.4 shows a simulated time series over 250 periods. The initial condition is $x_0 = 4$, and the parameters are as before. The horizontal line is the mean μ^* of the stationary distribution. The sequence is obviously correlated, and not surprisingly, shows no tendency to settle down to a constant value. On the other hand, the sample mean $\bar{x}_t := \frac{1}{t} \sum_{i=1}^{t} x_i$ seems to converge to μ^* (see figure 1.5).

That this convergence should occur is not obvious. Certainly it does not follow from the classical law of large numbers, since $(x_t)_{t \geq 0}$ is neither independent nor identically distributed. Nevertheless, the question of whether sample moments converge to the corresponding moments of the stationary distribution is an important one, with implications for both theory and econometrics.

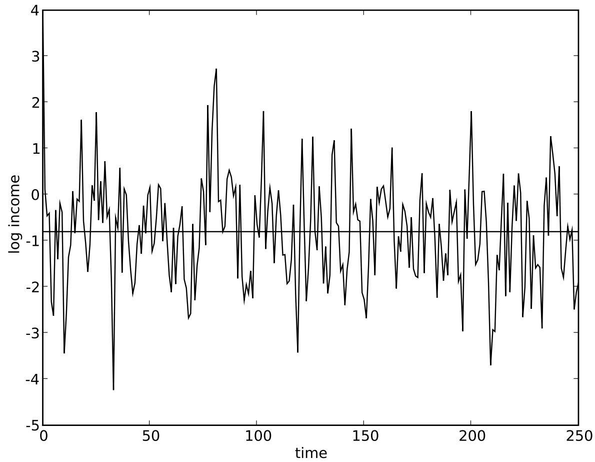

Figure 1.4 Time series

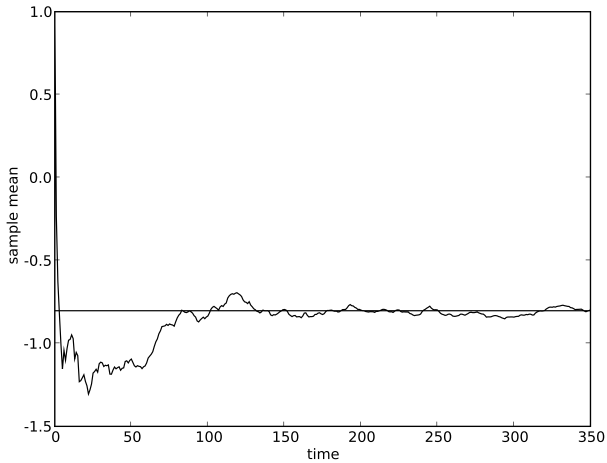

Figure 1.5 Sample mean of time series

For example, suppose that our simple model is being used to represent a given economy over a given period of time. Suppose further that the precise values of the underlying parameters α and ρ are unknown, and that we wish to estimate them from the data.[4] The method of moments technique proposes that we do this by identifying the first and second moments with their sample counterparts. That is, we set

$$\text{first moment } = \mu^*(\alpha, \rho) = \frac{1}{t}\sum_{i=1}^{t} x_t$$

$$\text{second moment } = v^*(\alpha, \rho) + \mu^*(\alpha, \rho)^2 = \frac{1}{t}\sum_{i=1}^{t} x_t^2$$

The right-hand side components $\frac{1}{t}\sum_{i=1}^{t} x_t$ and $\frac{1}{t}\sum_{i=1}^{t} x_t^2$ are collected from data, and the two equalities are solved simultaneously to calculate values for α and ρ.

The underlying assumption that underpins this whole technique is that sample means converge to their population counterparts. Figure 1.5 is not sufficient proof that such convergence does occur. We will need to think more about how to establish these results. More importantly, our linear normal-shock model is very special. Does convergence of sample moments occur for other related economies? The question is a deep one, and we will have to build up some knowledge of probability theory before we can tackle it.

To end our introductory comments, note that as well as studying theory, we will be developing computer code to tackle the problems outlined above. All the major code listings in the text can be downloaded from the text homepage, which can be found at `http://johnstachurski.net/book.html`. The homepage also collects other resources and links related to our topic.

[4]We have also used parameters b and θ, but $b = \alpha \ln \theta$, and $\theta = \alpha\rho$.