Long-run Uncertainty and Value

Lars Peter Hansen
University of Chicago and NBER

Thomas J. Sargent
New York University and NBER

Jose Scheinkman
Princeton University and NBER

July 2, 2009

Contents

List of Figures

Preface

This manuscript started off as the Toulouse Lectures given by Lars Peter Hansen. Our aim is to explore connections among topics that relate probability theory to the analysis of dynamic stochastic economic systems. Martingale methods have been a productive way to identify shocks with long-term consequences to economic growth and to characterize long-run dependence among macroeconomic time series. Typically they are applied by taking logarithms of time series such as output or consumption in order that growth can be modeled as accumulating linearly over time, albeit in a stochastic fashion. Martingale methods applied in this context have a long history in applied probability and applied time series analysis. We review these methods in the first part of this monograph. In the study of valuation, an alternative martingale approach provides a notion of long-term approximation. This approach borrows insights from large deviation theory, initiated in part to study the behavior of likelihood ratios of alternative time series models. We show how such methods provide characterizations of long-term model components and long-term contributions to valuation. Large deviation theory and the limiting behavior of likelihood ratios has also been central to some formulations of robust decision making. We develop this connection and build links to recursive utility theory in which investors care about the intertemporal composition of risk. Our interest in "robustness" and likelihood ratios is motivated by our conjecture that the modeling of the stochastic components to long-term growth is challenging for both econometricians and the investors inside the models that econometricians build.

More technical developments of some of these themes are given in Hansen and Scheinkman (1995), Anderson et al. (2003), Hansen and Scheinkman (2009) and Hansen (2008).

Acknowledgements

The author wishes to thank Jarda Borovicka, Adam Clemens, Serhiy Kokak and Junghoon Lee for their comments and assistance. These notes give some of the underlying support material for the Toulouse Lectures with the listed title.

Chapter 1

Stochastic Processes

In this chapter, we describe two ways of constructing stochastic processes. The first is one that is especially convenient for stating and proving limit theorems. The second is more superficial in the sense that it directly specifies objects that are outcomes in the first construction. However, the second construction is the one that is most widely used in modeling economic time series. We shall use these constructions to characterize limiting behavior both for stationary environments and for environments with stochastic growth.

1.1 Constructing a Stochastic Process: I

We begin with a method of constructing a stochastic process that is convenient for characterizing the limit of points of time series averages.[1] This construction works with a deterministic transformation $\mathbb{S}$ that maps a state of the world $\omega \in \Omega$ today into a state of the world $\mathbb{S}(\omega) \in \Omega$ tomorrow. The state of the world itself is not observed. Instead, a vector $X(\omega)$ that contains incomplete information about ω is observed. We assign probabilities over states of the world ω, then use the functions $\mathbb{S}$ and X to deduce a joint probability distribution for a sequence of X's. In more detail:

- The probability space is a triple $(\Omega, \mathfrak{F}, Pr)$, where Ω is a set of sample points, $\mathfrak{F}$ is an event collection (sigma algebra), and Pr assigns

[1] A good reference for the material in this section and the two that follow it is Breiman (1968).

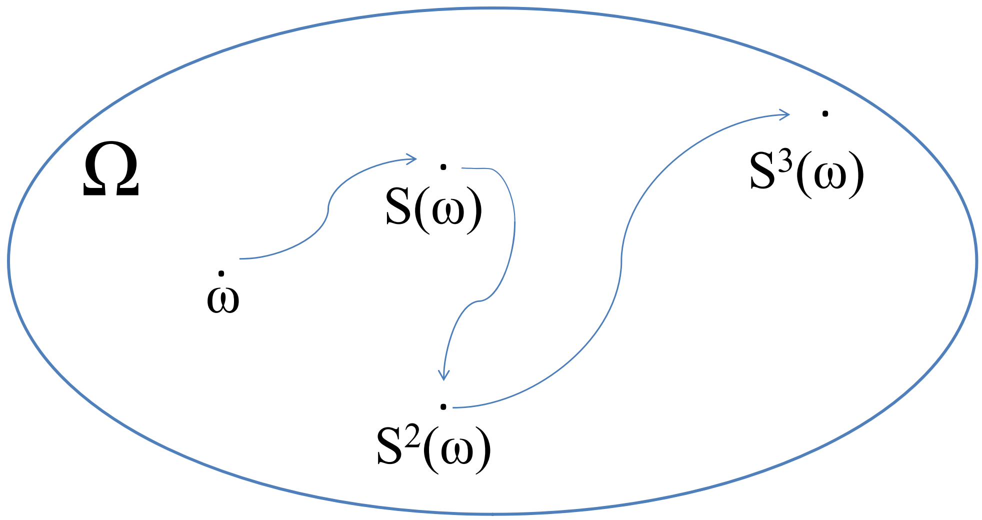

Figure 1.1: This figure depicts the state evolution as a function of sample point ω induced by the transformation $\mathbb{S}$. The oval shaped region is the collection Ω of all sample points.

probabilities to events.

- The (measurable) transformation $\mathbb{S} : \Omega \to \Omega$ used to model the evolution over time has the property that for any event $\Lambda \in \mathfrak{F}$,

$$\mathbb{S}^{-1}(\Lambda) = \{\omega \in \Omega : \mathbb{S}(\omega) \in \Lambda\}$$

  is an event. Notice that $\mathbb{S}^t(\omega)$ is a deterministic sequence of states of the world in Ω.

- The vector-valued function $X : \Omega \to \mathbb{R}^n$ used to model observations is Borel measurable. That is for any Borel set b in $\mathbb{R}^n$,

$$\Lambda = \{\omega \in \Omega : X(\omega) \in b\} \in \mathfrak{F}.$$

  In other words, X is a *random vector*.

- The stochastic process $\{X_t : t = 1, 2, ...\}$ used to model a sequence of observations is constructed via the formula:

$$X_t(\omega) = X[\mathbb{S}^t(\omega)]$$

or

$$X_t = X \circ \mathbb{S}^t.$$

The stochastic process $\{X_t : t = 1, 2, ...\}$ is a sequence of n-dimensional random vectors, and the probability measure Pr allows us to make probabilistic statements about the joint distribution of successive components of this sequence. It will sometimes be convenient to extend this construction to date zero by letting $X_0 = X$.

While this construction of a stochastic process may at first sight appear special, it is not, as the following example illustrates.

Example 1.1.1. *Let Ω be a collection of infinite sequences of elements of $\mathbb{R}^n$. Specifically, $\omega = (\mathbf{r}_0, \mathbf{r}_1, ...)$, $\mathbb{S}(\omega) = (\mathbf{r}_1, \mathbf{r}_2, ...)$ and $x(\omega) = \mathbf{r}_0$. Then $X_t(\omega) = \mathbf{r}_t$.*

1.2 Stationary Stochastic Processes

A *state* of a dynamic system is a complete description of its current position. The current state summarizes all information that can be gleaned from the past that is pertinent to forecasting the future. A *stationary* or *steady* state remains invariant as time passes. In a stochastic dynamic system, a stationary state is a probability distribution. In a stochastic steady state, for any finite τ the probability distribution of the composite random vector $[X_{t+1}{}', X_{t+2}{}', ..., X_{t+\ell}{}']'$ does not depend on t.

For a given $\mathbb{S}$, we can restrict the probability measure Pr to induce stationarity.

Definition 1.2.1. *The transformation $\mathbb{S}$ is **measure-preserving** if*

$$Pr(\Lambda) = Pr\{\mathbb{S}^{-1}(\Lambda)\}$$

for all $\Lambda \in \mathfrak{F}$.

Proposition 1.2.2. *When $\mathbb{S}$ is measure-preserving, the distribution function for X_t is identical for all $t \geq 0$.*

Given X, form a vector

$$X^{[\ell]}(\omega) \doteq \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \\ ... \\ X_\ell(\omega) \end{bmatrix},$$

Apply Proposition to $X^{[\ell]}$ and conclude that the joint distribution function for $(X_{t+1}, X_{t+2}, ..., X_{t+\ell})$ is independent of t for $t = 0, 1, \ldots$. The fact that this property holds for any choice of ℓ is equivalent to a statement that the process $\{X_t : t = 1, 2, ...\}$ is stationary.[2] Thus, the restriction that Pr be measure-preserving implies that the stochastic process $\{X_t : t = 1, 2, ...\}$ is stationary.

Example 1.2.3. *Suppose that Ω contains two states, $\Omega = \{\omega_1, \omega_2\}$. Consider a transformation $\mathbb{S}$ that maps ω_1 into ω_2 and ω_2 into ω_1: $\mathbb{S}(\omega_1) = \omega_2$ and $\mathbb{S}(\omega_2) = \omega_1$. Since $\mathbb{S}^{-1}(\omega_2) = \omega_1$ and $\mathbb{S}^{-1}(\omega_1) = \omega_2$, for $\mathbb{S}$ to be measure preserving we must have $Pr(\omega_1) = Pr(\omega_2) = .5$.*

Example 1.2.4. *Suppose that Ω contains two states, $\Omega = \{\omega_1, \omega_2\}$, and that $\mathbb{S}(\omega_1) = \{\omega_1\}$ and $\mathbb{S}(\omega_2) = \omega_2$. Since $\mathbb{S}^{-1}(\omega_2) = \omega_2$ and $\mathbb{S}^{-1}(\omega_1) = \omega_1$, $\mathbb{S}$ is measure preserving for $Pr(\omega_1) \geq 0$ and $Pr(\omega_1) + Pr(\omega_2) = 1$.*

1.3 Invariant Events and the Law of Large Numbers

In this subsection, we describe a Law of Large Numbers that tells us that time series averages converge when $\mathbb{S}$ is measure-preserving. We use the concept of an invariant event to understand possible limit points and how they are related to a conditional mathematical expectation.

Definition 1.3.1. *An event Λ is **invariant** if $\Lambda = \mathbb{S}^{-1}(\Lambda)$.*

Notice that if Λ is an invariant event and $\omega \in \Lambda$, then $\mathbb{S}^t(\omega) \in \Lambda$ for $t = 0, 1, ..., \infty$.

[2] Some people call this property 'strict stationarity' to differentiate it from notions that require only that some moments of the joint distribution be independent of time.

Let $\mathfrak{I}$ denote the collection of invariant events. Among the invariant events is the entire space, Ω, and the null set, $\varnothing$. Like $\mathfrak{F}$, this event collection is a sigma algebra. We are interested in constructed the conditional expectation $E(X|\mathfrak{I})$ as a random vector. Consider first the case in which the invariant events are unions of a finite partition Λ_j (along with the empty set). A finite partition consists of finite nonempty events Λ_j such that $\Lambda_j \cap \Lambda_k \neq \varnothing$ for $j \neq k$ and the union of all Λ_j is Ω. We assume that each member of the partition is itself and invariant event. The expectation condition on the event Λ_j is given by:

$$\frac{\int_{\Lambda_j} X dPr}{Pr(\Lambda_j)}.$$

This construction is applicable when $\omega \in \Lambda_j$. We extend this construction to the entire partition by

$$E(X|\mathfrak{I})(\omega) = \frac{\int_{\Lambda_j} X dPr}{Pr(\Lambda_j)} \text{ if } \omega \in \Lambda_j.$$

Thus the conditional expectation $E(X|\mathfrak{I})$ is constant within a partition and varies across partitions. This same construction extends directly to countable partitions.

There is an alternative way to think about a conditional expectation does not make reference to a partition but instead uses least squares when X has a finite second moment. Let Z be an n-dimensional measurement function such that

$$Z_t(\omega) = Z[\mathbb{S}^t(\omega)]$$

is time invariant (does not depend on calendar time). In the special case in which the invariant events are constructed from a finite partition, Z can vary across partitions but remains constant within a partition.[3] Let $\mathcal{Z}$ denote the collection of all such random vectors or measurement functions and solve the following least squares problem:

$$\min_{Z \in \mathcal{Z}} E[|X - Z|^2]$$

where we now assume that $E|X|^2 < \infty$. The solution to the least squares problem is $\tilde{Z} = E(X|\mathfrak{I})$. An implication of least squares is that

$$E\left[\left(X - \tilde{Z}\right) Z'\right] = 0$$

[3] More generally, Z is measurable with respect to $\mathfrak{I}$.

for Z in $\mathcal{Z}$ so that the vector $X - \tilde{Z}$ of regression errors must be orthogonal to any vector Z in $\mathcal{Z}$.

There is a more general measure-theoretic way to construct a conditional expectation. This construction extends the orthogonality property of least squares. Provided that $E|X| < \infty$, $E(X|\mathfrak{I})$ is essentially a unique random variable that for any invariant event Λ satisfies

$$E\left([X - E(X|\mathfrak{I})]\mathbf{1}_\Lambda\right) = 0$$

where $\mathbf{1}_\Lambda$ is the indicator function equal to one on the set Λ and zero otherwise.

The following states a key Law of Large Numbers.

Theorem 1.3.2. *(Birkhoff) Suppose that $\mathbb{S}$ is measure preserving.*

i) For any X such that $E|X| < \infty$

$$\frac{1}{N}\sum_{t=1}^{N} X_t \to E(X|\mathfrak{I})$$

with probability one;

ii) for any X such that $E|X|^2 < \infty$,

$$E\left[\left|\frac{1}{N}\sum_{t=1}^{N} X_t - E(X|\mathfrak{I})\right|^2\right] \to 0.$$

Definition 1.3.3. *The transformation $\mathbb{S}$ is **ergodic** if all invariant events have probability zero or one.*

From a probabilistic standpoint, when $\mathbb{S}$ is *ergodic* the invariant events are equivalent to the entire sample space Ω, which has probability one, or the empty set $\varnothing$, which has probability measure zero. This notion of ergodicity is a restriction on $\mathbb{S}$ and Pr that implies that conditioning on the invariant events is inconsequential.

Corollary 1.3.4. *Suppose that $\mathbb{S}$ is ergodic. Then $E(X|\mathfrak{I}) = E(X)$.*

Under ergodicity, the limit points of time series averages are the corresponding unconditional expectations. More generally, time series averages can only reveal expectations conditioned on the invariant events.

Consider again Example 1.2.3. Suppose that the measurement vector is

$$X(\omega) = \begin{cases} 1 & \omega = \omega_1 \\ 0 & \omega = \omega_2. \end{cases}$$

Then it follows directly from the specification of $\mathbf{S}$ that

$$\frac{1}{N} \sum_{t=1}^{N} X_t(\omega) \to \frac{1}{2}$$

for both values of ω. The limit point is the average across states. For Example 1.2.4, $X_t(\omega) = X(\omega)$ and hence the sequence is time invariant and equal to the time series average. The time series average equals the average across states only when one of the two states is assigned probability measure one. Theorem 1.3.2 covers the convergence in general and Corollary 1.3.4 covers the case in which the probability assignment is degenerate. These two examples are included merely for illustration, and we will explore much richer specifications of stochastic processes.

1.4 Constructing a Stochastic Process: II

Instead of specifying

$$X^{[\ell]} \doteq \begin{bmatrix} X_1 \\ X_2 \\ ... \\ X_\ell \end{bmatrix}$$

as in section 1.1, we could directly specify a collection of joint distributions $\hat{Pr}_\ell$ for $\ell \geq 1$. But we must make sure that $\hat{Pr}_{\ell+1}$ is consistent with $\hat{Pr}_\ell$ in the sense that both of these joint distributions assign the same probabilities for the same events, namely, $\{X_\ell^* \in b\}$ for (Borel) sets b. If this consistency condition is satisfied, then the famed Kolmogorov Extension Theorem guarantees that there exists a space $(\Omega, \mathfrak{F}, Pr)$ and a stochastic process $\{X_t : t = 1, 2, ...\}$ constructed as in section 1.1. Applied model builders typically use a direct specification approach.

A *Markov process* is an important tool for directly constructing a joint distribution $\hat{Pr}_\ell, \ell \geq 0$. Consider a state space $\mathcal{E}$ and a transition distribution $T(dx^*|x)$. The transition distribution T is a conditional probability measure for each choice $X_t = x$ in the state space, so it satisfies $\int T(dx^*|x) = 1$ for every x in the state space. There is an associated conditional expectation function. If in addition we specify a marginal Q_0 distribution for the initial state x_0 over $\mathcal{E}$, then we have completely specified all joint distributions for the stochastic process.

The notation $T(dx^*|x)$ denotes a conditional probability measure where the integration is over x^* and the conditioning is captured by x. Specifically, x^* is a possible realization of the next period state and x is a realization of the current period state. The conditional probability measure $T(dx^*|x)$ assigns conditional probabilities to the next period state given that the current period state is x. Often, but not always, the conditional distributions have densities against a common distribution $\lambda(dx^*)$ used to add up or integrate over states. In such cases we can use a *transition density* to represent the conditional probability measure. One example is that of first-order vector autoregression. In this case $T(dx^*|x)$ is a normal distribution with mean Ax and covariance matrix BB' for a square matrix A and a matrix B with full column rank.[4] In this we may write

$$X_{t+1} = AX_t + BW_{t+1}$$

where W_{t+1} is a multivariate standard normally distributed random vector that is independent of X_t. Another example is that of a discrete-state Markov chain in which $T(dx^*|x)$ can be represented as a matrix, one row for each realized value of the state x. The row entries give the vector of probabilities conditioned on this realized values. Both of these examples will be developed in more detail later.

An important object for us is a one-step conditional expectation operator that we apply to functions of a Markov state. Let $f : \mathcal{E} \to \mathbb{R}$. For bounded f, define:

$$\mathbb{T}f(x) = E\left[f(X_{t+1})|X_t = x\right] = \int f(x^*)T(dx^*|x)$$

[4]When BB' is singular, a density may not exist with respect to Lebesgue measure. Such singularity occurs when we convert a higher-order vector autoregression into a first-order process.

Iterating on $\mathbb{T}$ allows us to form expectations over longer time periods:

$$\mathbb{T}^j f(x) = E\left[f(X_{t+j})|X_t = x\right].$$

This is a statement of the Law of Iterated Expectations for our Markov setting.

Remark 1.4.1. *Instead of beginning with a conditional probability distribution, we could start with a conditional expectation operator $\mathbb{T}$ mapping a space of functions into itself. Provided that this operator is a) well defined on the space of bounded functions, b) preserves the bound, c) maps nonnegative functions into nonnegative functions, and d) maps the unit function into the unit function; we can construct a conditional probability measure $T(dx^*|x)$ from the operator $\mathbb{T}$.*

1.5 Stationarity reconsidered

We construct Markov processes that are stationary by appropriately choosing distributions of the initial state x_0.

Definition 1.5.1. *A* ***stationary distribution*** *for a Markov process is a probability measure Q over the state space $\mathcal{E}$ that satisfies*

$$\int T(dx^*|x)Q(dx) = Q(dx^*).$$

We will sometimes make reference to a stationary density q. A density is always relative to a some measure. With this in mind, let λ be a measure on the state space $\mathcal{E}$ used to add up or integrate over alternative Markov states. Then a density q is a nonnegative (Borel measurable) function of the state for which $\int q(x)\lambda(dx) = 1$.

Definition 1.5.2. *A* ***stationary density*** *for a Markov process is a probability density q with respect to a measure λ over the state space $\mathcal{E}$ that satisfies*

$$\int T(dx^*|x)q(x)\lambda(dx) = q(x^*)\lambda(dx^*).$$

The following example of a *reversible* Markov process occurs sometimes in simulations that implement Bayesian estimation.

Example 1.5.3. *Suppose that*

$$T(dx^*|x)q(x)\lambda(dx) = T(dx|x^*)q(x^*)\lambda(dx^*).$$

Because a transition density satisfies $\int T(dx|x^*) = 1$, *notice that*

$$\int T(dx^*|x)q(x)\lambda(dx) = \int T(dx|x^*)q(x^*)\lambda(dx^*) = q(x^*)d\lambda(dx^*).$$

Thus, q is a stationary density.

When the Markov process is initialized according to a stationary distribution, we can construct the process $\{X_t : t = 1, 2, ...\}$ with a measure-preserving transformation $\mathbb{S}$ of the type featured in the first method of constructing a stochastic process that we described in section 1.1.

Given a stationary distribution Q, form the space of functions $\mathcal{L}^2$ defined as

$$\mathcal{L}^2 = \{f : \mathcal{E} \to \mathbb{R} : \int f(x)^2 Q(dx) < \infty\}$$

It can be shown that $\mathbb{T} : \mathcal{L}^2 \to \mathcal{L}^2$. On this space there is a well defined norm give by:

$$\|f\| = \left[\int f(x)^2 Q(dx)\right]^{1/2}$$

1.6 Limiting Behavior

When the Markov process is not periodic, we are interested in situations when

$$\lim_{j\to\infty} \mathbb{T}^j f(x) = \mathbf{r} \tag{1.1}$$

for some $\mathbf{r} \in \mathbf{R}$ where the convergence is either pointwise in x or define using the $\mathcal{L}^2$ norm. This limit restricts the long-term forecasts eventually not to depend on the current Markov state. (See Meyn and Tweedie (1993) for a comprehensive treatment of this convergence.) Let Q be a stationary distribution. Then it is necessarily true that

$$\int \mathbb{T}^j f(x)Q(dx) = \int f(x)Q(dx)$$

for all j. Thus

$$\mathbf{r} = \int f(x)Q(dx).$$

Thus the limiting forecast is necessary the expectation under a stationary distribution. Notice that here we have not assumed that the stationary density is unique, although we did presume that the limit point is a number and not a random variable.

One reason we are interested in limit (1.1) is that when a stationary distribution exists, this limit implies the convergence of:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \left\| \mathbb{T}^j f(x) - \int f(x) dQ(x) \right\| = 0.$$

This suggests that limit point for a time series version of the Law of Large Numbers is $\int f(x)dQ(x)$. Also, if $\int f(x)Q(dx) = 0$ and the convergence is sufficiently fast, then

$$\lim_{N \to \infty} \sum_{j=1}^{N} \mathbb{T}^j f(x)$$

is a well-defined function of the Markov state. Under stationarity we can represent the limit in first case as the $\int f(x)Q(dx)$, and a necessary condition for the second limit is that $\int f(x)Q(dx) = 0$.

Ergodicity

Definition 1.6.1. *A function $\tilde{f} \in \mathcal{L}^2$ that solves the equation $\mathbb{T}f = f$ is called* an eigenfunction associated with a unit eigenvalue.

An eigenfunction of $\mathbb{T}$ is a generalization of an eigenvector of a matrix. Notice that if $\tilde{f}$ is used in calculation (1.1), the $\tilde{f}$ is necessarily constant.

Proposition 1.6.2. *Suppose that $\tilde{f}$ is an eigenfunction of $\mathbb{T}$ associated with a unit eigenvalue. Then $\{\tilde{f}(X_t) : t = 1, 2, ...\}$ is constant over time with probability one.*

Proof.

$$E\left[\tilde{f}(X_{t+1})\tilde{f}(X_t)\right] = \int (\mathbb{T}\tilde{f})(x)\tilde{f}(x)Q(dx) = \int \tilde{f}(x)^2 Q(dx) = E\left[\tilde{f}(X_t)^2\right].$$

Then because Q is a stationary distribution,

$$\begin{aligned} E\left([\tilde{f}(X_{t+1}) - \tilde{f}(X_t)]^2\right) &= E\left[\tilde{f}(X_{t+1})^2\right] + E\left[\tilde{f}(X_t)^2\right] \\ &\quad -2E\left[\tilde{f}(X_{t+1})\tilde{f}(X_t)\right] \\ &= 0. \end{aligned}$$

□

Obviously, time series averages of an such an eigenfunction $\mathbb{T}\tilde{f} = \tilde{f}$ do not move either, so

$$\frac{1}{N}\sum_{t=1}^{N} \tilde{f}(X_t) = \tilde{f}(X).$$

However, the time series average $\frac{1}{N}\sum_{t=1}^{N} \tilde{f}(X_t)$ differs from $\int \tilde{f}(x)Q(dx)$ when $\tilde{f}(x)$ when $\tilde{f}(x)$ is not constant across states x that occur with probability one under Q. This happens when the variation of $\tilde{f}(X_t)$ along a sample path for $\{X_t\}$ conveys an inaccurate impression of its variation across the stationary distribution $Q(dx)$. See example 1.7.2 below. This possibility leads us to use eigenfunctions to state a sufficient condition for ergodicity.

When f is an indicator function of a Borel set b in $\mathcal{E}$ and $\mathbb{T}f = f$, then

$$\Lambda = \{\omega \in \Omega : X \in b\}$$

is an invariant event in Ω under the corresponding probability measure Pr and transformation $\mathbb{S}$. For Markov processes, all invariant events can be represented like this, a result that is not obvious. A reference for it is Doob (1953), Theorem 1.1, page 460.

Proposition 1.6.3. *When the only solution to the eigenvalue equation*

$$\mathbb{T}f = f$$

is a constant function (with Q measure one), then it is possible to construct the process $\{X_t : t = 0, 1, ...\}$ using a transformation $\mathbb{S}$ that is measure preserving and ergodic.

Notice here that ergodicity is a property that obtains only relative to a stationary distribution for the Markov process. When there are multiple stationary distributions, a constant solution to the eigenvalue problem can be the only one that works for one stationary distribution, but non constant solutions can exist for other stationary distributions. For instance, consider example 1.2.4 . Although any assignment of probabilities constitutes a stationary distribution, we get ergodicity only when we assign probability one to one of the two states. Also see example 1.7.3.

Sufficient Conditions for Ergodicity

While finding nondegenerate eigenfunctions associated with a unit eigenvalue often gives a way to establish that a process is *not* ergodic, it can be difficult to establish ergodicity directly using Proposition 1.6.3. There are convenient sufficient conditions, including the drift conditions discussed in Meyn and Tweedie (1993). We explore one set of sufficient conditions in this subsection.

Let Q be a stationary distribution. Form the *resolvent* operator:

$$\mathbb{R}f(x) = (1-\delta)\sum_{j=1}^{\infty} \delta^j \mathbb{T}^j f$$

associated with some constant discount factor $0 < \delta < 1$. To study periodic components of processes, we introduce sampling at an interval p. If we sample a periodic process of period p, we want functions of the Markov state to be invariant, so

$$\mathbb{T}^p f = f$$

for some function f that is nondegenerate and not constant with probability one. This leads us to consider a sampled counterpart of the resolvent operator $\mathbb{R}$:

$$\mathbb{R}_p f(x) = (1-\delta)\sum_{j=1}^{\infty} \delta^j \mathbb{T}^{pj} f.$$

A set of sufficient conditions for

$$\lim_{j\to\infty} \mathbb{T}^j f(x^*) \to \int f(x) dQ(x) \tag{1.2}$$

for each $x \in \mathcal{E}$ and each $f \in \mathcal{L}^2$ that are bounded is:[5]

[5]Restriction 1.2 is stronger than ergodicity. it rules out periodic processes, although we know that periodic processes can be ergodic.

Condition 1.6.4. *Suppose that the stationary Markov process satisfies:*

(i) For any $f \geq 0$ such that $\int f(x)Q(dx) > 0$, $\mathbb{R}_p f(x) > 0$ for all $x \in \mathcal{E}$ and all $p \geq 0$. (The Markov process is (Q) irreducible and aperiodic.)

(ii) $\mathbb{T}$ maps bounded continuous functions into bounded continuous functions. (The Markov process satisfies the Feller property.)

(iii) The support of Q has a nonempty interior in $\mathcal{E}$.

(iv) $\mathbb{T}V(x) - V(x) \leq -1$ outside a compact subset of $\mathcal{E}$ for some nonnegative function V.

Sufficient condition (i) may be hard to check, but it suffices to show that there exists an m such that for any $f \geq 0$ such that $\int f(x)Q(dx) > 0$

$$\mathbb{T}^m f(x) > 0$$

for all $x \in \mathcal{E}$. Given this property holds for $\mathbb{T}^m$ it must also hold true for pm for any p. Condition (iv) is the *drift condition* for stability. It is constructive provided that we can establish the inequality for a conveniently chosen function V. Heuristically, this drift condition says that outside a compact subset of the state space, the conditional expectation must push inward. The choice of -1 as a comparison point is made for convenience since we can always multiply the function V by a number greater than one. Thus -1 could be replaced by any strictly negative number.

1.7 Finite State Markov Chain

Suppose that $\mathcal{E}$ consists of n states. We may label these state in any of a variety of ways but suppose that state x_j is the coordinate vector of all zeros except in position j where there is a one. Let $\mathbb{T}$ be an n by n transition matrix where entry i, j is the probability of moving from state i to state j in a single period. Thus the entries of $\mathbb{T}$ are all nonnegative and this matrix must satisfy:

$$\mathbb{T}\mathbf{1}_n = \mathbf{1}_n.$$

where $\mathbf{1}_n$ is an n-dimensional vector of ones. Let f be any n-dimensional vector.

Let q be an n-dimensional vector of probabilities. Then stationarity requires that

$$q'\mathbb{T} = q'$$

That is, q is a row eigenvector of $\mathbb{T}$ associated with a unit eigenvalue.

We use a vector f to represent a *function* from the state space to the real line, where each coordinate of f gives the value of the function at the corresponding coordinate vector. Consider column eigenvectors of $\mathbb{T}$ associated with a unit eigenvalue. Suppose that the only solutions to

$$\mathbb{T}f = f$$

are of the form $f \propto \mathbf{1}_n$. Then we can construct a process that is stationary and ergodic by initializing the process with density q.

We can weaken this condition to allow nonconstant right eigenvectors. A weaker condition is that the eigenvector and stationary distribution satisfy

$$\min_{\mathbf{r}} \sum_{i=1}^{n} (f_i - \mathbf{r})^2 q_i = 0$$

Notice that we are multiplying by probabilities, so that if q_i is zero, the contribution of f_i to the least squares objective can be neglected, which allows for non-constant f's, albeit in a limited way.

Three examples illustrate these concepts.

Example 1.7.1. *We now recast Example 1.2.3 as a Markov chain with transition matrix* $\mathbb{T} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. *This chain has a unique invariant distribution* $q = \begin{bmatrix} .5 & .5 \end{bmatrix}'$ *and the invariant functions are* $\begin{bmatrix} \alpha & \alpha \end{bmatrix}'$ *for any scalar* α. *Therefore, the process initiated from the stationary distribution is ergodic.*

Example 1.7.2. *Next we recast Example 1.2.4 as a Markov chain with transition matrix* $\mathbb{T} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. *This chain has a continuum of stationary distributions* $\pi \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (1-\pi) \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ *for any* $\pi \in [0,1]$ *and invariant functions* $\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$ *and for any scalars* α_1, α_2. *Therefore, the process is not ergodic when* $\pi \in (0,1)$, *for note that if* $\alpha_1 \neq \alpha_2$ *the resulting invariant function will* *fail to be constant across states receive positive probability according to a stationary distribution associated with $\pi \in (0,1)$. When $\pi \in (0,1)$, nature chooses state $i = 1$ or $i = 2$ with probabilities $\pi, 1-\pi$, respectively, at time 0. Thereafter, the chain remains stuck in the realized time 0 state. Its failure ever to visit the unrealized state prevents the sample average from converging to the population mean of an arbitrary function $\bar{y}$ of the state.*

Example 1.7.3. *A chain with transition matrix* $\mathbb{T} = \begin{bmatrix} .8 & .2 & 0 \\ .1 & .9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ *has a continuum of stationary distributions* $\pi \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix}' + (1-\pi) \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}'$ *for* $\pi \in [0,1]$ *and invariant functions* $\begin{bmatrix} \alpha_1 & \alpha_1 & \alpha_2 \end{bmatrix}'$ *and for any scalars* α_1, α_2. *With any stationary distribution associated with* $\pi \in (0,1)$, *the chain is not ergodic because some invariant functions are not constant with probability one under such a stationary distribution. But for stationary distributions associated with* $\pi = 1$ *or* $\pi = 0$, *the chain is ergodic.*

1.8 Vector Autoregression

When the eigenvalues of a square matrix A have absolute values that are strictly less than one we say that A is *stable*. For a stable A, suppose that

$$X_{t+1} = AX_t + BW_{t+1}$$

where $\{W_{t+1} : t = 1, 2, ...\}$ is an iid sequence of multivariate normally distributed random vectors and B has full column rank. Then

$$W_{t+1} = (B'B)^{-1}B'(X_{t+1} - AX_t),$$

so we can recover the shock vector W_{t+1} from X_{t+1} and X_t. To complete the specification of a Markov process, we specify an initial distribution $X_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$.

Let $\mu_t = EX_t$. Notice that

$$\mu_{t+1} = A\mu_t.$$

The mean μ of a stationary distribution satisfies

$$\mu = A\mu.$$

Because we have assumed that A is a stable matrix, the only μ that solves $(A - I)\mu = 0$ is $\mu = 0$. Thus, the mean of the stationary distribution is $\mu = 0$.

Let Σ_t be the covariance matrix for X_t. Then

$$\Sigma_{t+1} = A\Sigma_t A' + BB'.$$

For $\Sigma_t = \Sigma$ to be invariant over time, it must be true that

$$\Sigma = A\Sigma A' + BB'.$$

Because A is a stable matrix, this equation has a unique solution for a positive semidefinite matrix Σ. This is the covariance matrix of the stationary distribution.

Suppose that $\Sigma_0 = 0$ (a matrix of zeros). Then

$$\Sigma_t = \sum_{j=0}^{t-1} A^j BB'(A^j)'.$$

The limit of this sequence is:

$$\Sigma = \sum_{j=0}^{\infty} A^j BB'(A^j)'$$

which we have seen is the covariance matrix for the stationary distribution. Similarly,

$$\mu_t = A^t \mu_0,$$

converges to zero for all $\mu_0 = EX_0$. Recall that 0 is also the mean of the stationary distribution.

The linear structure of the model implies that the stationary distribution is Gaussian with mean μ and covariance matrix Σ.

To verify ergodicity, suppose that the covariance matrix Σ of the stationary distribution has full rank. Then restriction (iii) of Condition 1.6.4 is satisfied. Furthermore, Σ_t has full rank for some t, which guarantees that the process is irreducible and aperiodic (restriction (i). Let $V(x) = |x|^2$. Then

$$\mathbb{T}V(x) = x'A'Ax + \text{trace}(B'B).$$

Thus

$$\mathbb{T}V(x) - V(x) = x'(A'A - I)x + \text{trace}(B'B).$$

That A is a stable matrix implies that $A'A - I$ is negative definite, so that drift restriction (iv) of Condition 1.6.4 is satisfied for $|x|$ sufficiently large.[6]

We now show how to extend this example to obtain a nonzero mean for the stationary distribution. Partition the Markov state as:

$$x = \begin{bmatrix} x^{[1]} \\ x^{[2]} \end{bmatrix}$$

where $x^{[2]}$ is scalar. Similarly partition the matrix and the matrices A and B as:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & 1 \end{bmatrix}$$
$$B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}$$

where A_{11} is a stable matrix. Notice that

$$X_{t+1}^{[2]} = X_t^{[2]} = ... = X_0^{[2]}$$

and hence is invariant. Let $\mu^{[2]}$ denote the mean of $X_t^{[2]}$ for any t. In a stationary distribution we require that the mean $\mu^{[1]}$ of the first component of the state vector satisfy:

$$\mu^{[1]} = A_{11}\mu^{[1]} + A_{12}\mu^{[2]}.$$

Hence

$$\mu^{[1]} = (I - A_{11})^{-1} A_{12}\mu^{[2]}.$$

Imitating a previous argument, the covariance matrix, $\Sigma^{[11]}$ for this component satisfies:

$$\Sigma^{[11]} = \sum_{j=0}^{\infty} (A_{11})^j BB' (A_{11}')^j + (I - A_{11})^{-1} A_{12}\Sigma^{[22]}A_{12}' (I - A_{11}')^{-1}$$

where $\Sigma^{[22]}$ is the variance of $X_t^{[2]}$ for all t. Stationarity imposes no restriction on the mean $\mu^{[2]}$ and the variance $\Sigma^{[22]}$.

Since $\{X_t^{[2]} : t = 0, 1, ...\}$ is invariant, the process $\{X_t : t = 0, 1, ...\}$ is only ergodic when the variance $\Sigma^{[22]}$ is zero. Otherwise, the limit points for the Law of Large Numbers (Theorem 1.3.2) should be computed by conditioning on $X_0^{[2]}$.

[6]The Feller property can also be established.

Chapter 2

Additive Functionals

For economic applications, it is too limiting to consider only time series models that are stationary because we are interested in processes that display stochastic counterparts to geometric growth or, equivalently, arithmetic growth in logarithms. We suggest a convenient construction of such a process and produce a convenient decomposition into a time trend, a martingale and a stationary component.

2.1 Construction

Let $\{X_t\}$ be a stationary Markov process. We now build *functionals* of this process by accumulating the impact of the Markov process over time.

Definition 2.1.1. *A process* $\{Y_t : t = 0, 1, ...\}$ *is said to be and* ***additive functional*** *if it can be represented as*

$$Y_{t+1} - Y_t = \kappa(X_{t+1}, X_t), \tag{2.1}$$

or equivalently

$$Y_t = \sum_{j=1}^{t} \kappa(X_j, X_{j-1})$$

where we ininitialize $Y_0 = 0$.

The initialization, $Y_0 = 0$ is imposed for convenience, but it does allow us to construct Y_t as a function of *only* the underlying Markov process between date zero and t. Adding a nonzero initial condition will have obvious consequences for the results in this chapter.

A linear combination of two additive functionals $\{Y_t^{[1]}\}$ and $\{Y_t^{[2]}\}$ is an additive functional. If κ_1 is used to construct the first process and κ_2 the second process, then $\kappa = \kappa_1 + \kappa_2$ can be used to construct the sum of the two processes.

Example 2.1.2.

$$X_{t+1} = AX_t + BW_{t+1}$$

where $\{W_{t+1} : t = 1, 2, ...\}$ *is an iid sequence of multivariate normally distributed random vectors and* B *has full column rank. Premultiply by* B' *and obtain:*

$$B'X_{t+1} - B'AX_t = B'BW_{t+1}.$$

Then

$$W_{t+1} = (B'B)^{-1}(B'X_{t+1} - B'AX_t).$$

Form

$$\kappa(X_{t+1}, X_t) = \mu(X_t) + \sigma(X_t)W_{t+1}.$$

Then $\mu(X_t)$ *is the conditional mean of* $Y_{t+1} - Y_t$ *and* $|\sigma(X_t)|^2$ *is the conditional variance. When* σ *depends on the Markov state, this is referred to as a stochastic volatility model.*

Let $\mathfrak{F}_t$ be the information set (sigma algebra) generated by $X_0, X_1, ..., X_t$.

Definition 2.1.3. *A process* $\{Y_t : t = 0, 1, ...\}$ *is an* ***additive martingale*** *provided that* $E[\kappa(X_{t+1}, X_t)|X_t] = 0$.

Note that $E[\kappa(X_{t+1}, X_t)|X_t] = 0$ implies the usual martingale restriction

$$E[Y_{t+1}|\mathfrak{F}_t] = Y_t, \quad \text{for } t = 0, 1, ...$$

The process $\{Y_t : t = 0, 1, ...\}$ defined in example (2.1.2) is evidently a martingale if $\mu(X_t) = 0$.

2.2 Martingale Extraction

Additive processes have additive martingales embedded within them that capture all long-run variation. In this section, we show how to extract these martingales and suggest ways they can be used.

We use the following subspace of $\mathcal{L}^2$:

$$\mathcal{Z} = \left\{ f \in \mathcal{L}^2 : \int f(x) Q(dx) = 0 \right\}.$$

Thus functions in $\mathcal{Z}$ have mean zero under the stationary distribution. Define the norm $\|f\| = \left[\int f(x)^2 Q(dx)\right]^{1/2}$ on $\mathcal{L}^2$ and hence on $\mathcal{Z}$.

Definition 2.2.1. *The conditional expectation operator* $\mathbb{T}$ *is a* strong contraction *(on* $\mathcal{Z}$*) if there exists a* $0 < \rho < 1$ *such that*

$$\|\mathbb{T} f\| \leq \rho \|f\|$$

for all $f \in \mathcal{Z}$.[1]

Example 2.2.2. *Consider the Markov chain setting of subsection 1.7. The conditional expectation can be represented using a transition matrix* $\mathbb{T}$*. We have seen that to obtain a stationary density, we should solve*

$$q'\mathbb{T} = q'$$

for a nonnegative vector q *such that* $q \cdot \mathbf{1}_n = 1$*. If the only column eigenvector of* $\mathbb{T}$ *associated with a unit eigenvalue is a constant over states* i *for which* $q_i > 0$*, then the process is ergodic. If in addition, the only eigenvector of* $\mathbb{T}$ *with unit norm (this includes complex eigenvalues), is constant over states* i *for which* $q_i > 0$*, then* $\mathbb{T}^m$ *will be a strong contraction for some* m.[2] *In addition to imposing ergodicity, this rules out periodic components that can be forecast perfectly.*

For a Markov process $\{X_t : t = 0, 1, ...\}$, consider the following special algorithm that applies to a special type of an additive process for which $\kappa(x^*, x) = f(x)$ with $\int f(x)Q(dx) = 0$.

[1] When this property is satisfied the underlying process is said to be ρ-mixing.

[2] This result follows from Gelfand's Theorem. Let $\mathcal{Z}$ be the $n-1$ dimensional space of vectors that are orthogonal to q. Then $\mathbb{T}$ maps $\mathcal{Z}$ into itself. The spectral radius of this transformation is the maximum of the absolute values of the eigenvalues of the induced transformation. Gelfand's formula shows that the spectral radius governs the asymptotic decay of the transformation applied m times as m gets large implying the strong contraction property for any ρ larger than the spectral radius.

Algorithm 2.2.3. *Suppose that* $f \in \mathcal{Z}$ *and*

$$Y_{t+1} - Y_t = f(X_t).$$

Thus $\kappa(x^*, x) = f(x)$. *Solve the equation* $g(x) = f(x) + \mathbb{T}g(x)$ *for* g. *The solution is:*

$$g(x) = (\mathbb{I} - \mathbb{T})^{-1} f(x) = \sum_{j=0}^{\infty} \mathbb{T}^j f(x) = \sum_{j=0}^{\infty} E\left[f(X_{t+j}) | X_t = x\right], \qquad (2.2)$$

where $\mathbb{I}$ *is the identity operator, a legitimate calculation provided that the infinite sum on the right-hand side of* (2.2) *is finite. The function* g *gives the best forecast today of the long-term limit of the additive functional* $\{Y_t : t = 0, 1, ...\}$ *with the argument being the current Markov state. A sufficient condition for the sum on the right-hand side of* (2.2) *to be finite is that* $\mathbb{T}^m$ *be a strong contraction for some* m. *Evidently,* $(\mathbb{I} - \mathbb{T})g(x) = f(x)$.

Let

$$\breve{\kappa}(x^*, x) = g(x^*) - g(x) + f(x)$$

and note that $(\mathbb{I} - \mathbb{T})g(x) = f(x)$ *implies that*

$$\breve{\kappa}(x^*, x) = g(x^*) - \mathbb{T}g(x).$$

Thus, $\breve{\kappa}(X_{t+1}, X_t)$ *is the forecast error in forecasting* $g(X_{t+1})$ *given* X_t, *and in particular*

$$E\left[\breve{\kappa}(X_{t+1}, X_t) | X_t\right] = 0.$$

Therefore,

$$\begin{aligned} Y_{t+1} &= \sum_{j=1}^{t+1} f(X_{j-1}) \\ &= \sum_{j=1}^{t+1} \breve{\kappa}(X_j, X_{j-1}) - g(X_{t+1}) + g(X_0) \end{aligned}$$

is a martingale.

This algorithm is a martingale counterpart to a more general construction of Gordin (1969) for stationary processes.[3] We now use this algorithm as a component of a more general martingale extraction algorithm.

[3] See also Hall and Heyde (1980).

Algorithm 2.2.4. *Let $\{X_t : t = 0, 1, \ldots\}$ be a stationary, ergodic Markov process. Let $Y_t, t = 0, 1, \ldots$ be an additive process. Perform the following steps.*

(i) Compute the conditional expectation of the growth rate $E[\kappa(X_{t+1}, X_t)|X_t = x] = \bar{f}(x)$ and form the deviation from the conditional mean

$$\tilde{\kappa}(X_{t+1}, X_t) = \kappa(X_{t+1}, X_t) - \bar{f}(X_t).$$

Note that $E[\tilde{\kappa}(X_{t+1}, X_t)|X_t = x] = 0$.

(ii) Compute the deviation of the conditional mean from the unconditional mean of the growth rate $\nu = \int \bar{f}(x)q(x)d\lambda(x)$, namely, $f(x) = \bar{f}(x) - \nu$ and apply algorithm 2.2.3 to form g and $\check{\kappa}$ in the decomposition $f(x) = \check{\kappa}(x^*, x) - g(x^*) - g(x)$.

(iii) Note that

$$\begin{aligned} \kappa(x^*, x) &= \tilde{\kappa}(x^*, x) + \bar{f}(x) \\ &= \tilde{\kappa}(x^*, x) + f(x) + \nu \\ &= \tilde{\kappa}(x^*, x) + \check{\kappa}(x^*, x) - g(x^*) + g(x) + \nu. \end{aligned}$$

(iv) It follows that

$$Y_t = t\nu + \left[\sum_{j=1}^{t} \hat{\kappa}(X_j, X_{j-1}) \right] - g(X_t) + g(X_0)$$

where $\hat{\kappa}(X^*, X) = \check{\kappa}(X^*, X) + \tilde{\kappa}(X^*, X)$ and $E[\hat{\kappa}(X_{j+1}, X_j)|X_j] = 0$.

Thus, we have established that

Proposition 2.2.5. *Suppose that $\{Y_t : t = 0, 1, ...\}$ is an additive functional, $\mathbb{T}^n$ is a strong contraction on $\mathcal{Z}$ for some n and $E[\kappa(X_{t+1}, X_t)^2] < \infty$. Then*

$$Y_t = \underbrace{[t\nu]}_{\textbf{term one}} + \underbrace{\left[\textstyle\sum_{j=1}^{t} \hat{\kappa}(X_j, X_{j-1})\right]}_{\textbf{term two}} + \underbrace{[-g(X_t) + g(X_0)]}_{\textbf{term three}}.$$

The three terms in the decomposition are each additive processes initialized at zero. The first is a linear time trend, the second is an additive martingale, and the third is a stationary process.

The remainder of this section describes some applications.

2.3 Cointegration

Linear combinations of two additive processes are additive. Specifically, form

$$Y_t = \mathbf{r}_1 Y_t^{[1]} + \mathbf{r}_2 Y_t^{[2]}$$

where $Y_t^{[1]}$ is built with κ_1 and Y_t^2 is built with κ_2. Thus, we can build

$$Y_t = \mathbf{r}_1 Y_t^{[1]} + \mathbf{r}_2 Y_t^{[2]} = \sum_{j=1}^{t} \left[\mathbf{r}_1 \kappa_1(X_j, X_{j-1}) + \mathbf{r}_2 \kappa_2(X_j, X_{j-1}) \right]$$

The martingale of Proposition 2.2.5 for $\{Y_t : t = 0, 1, ...\}$ is the corresponding linear combination of the martingales for the two components.

Engle and Granger (1987) call two processes *cointegrated* if there exists a linear combination of these processes that is stationary. This occurs when

$$\begin{aligned} \mathbf{r}_1 \nu_1 + \mathbf{r}_2 \nu_2 &= 0 \\ \mathbf{r}_1 \hat{\kappa}_1 + \mathbf{r}_2 \hat{\kappa}_2 &= 0 \end{aligned}$$

where the ν's and $\hat{\kappa}$'s correspond to the first two components of the representation in Proposition 2.2.5. These two zero restrictions imply that the time trend and martingale component for the linear combination Y_t are both zero.[4] It is of particular interest when $\mathbf{r}_1 = 1$ and $\mathbf{r}_2 = -1$. In this case the two componenet additive processes $Y_t^{[1]}$ and $Y_t^{[2]}$ share a common growth component.

2.4 Identifying Shocks with Long-Run Consequences

Suppose that the Markov state $\{X_t : t = 0, 1, ...\}$ follows the first-order VAR

$$X_{t+1} = AX_t + BW_{t+1}$$

where A has stable eigenvalues. Let

$$Y_{t+1} - Y_t = \kappa(X_{t+1}, X_t) = D \cdot X_t + \nu + F \cdot W_{t+1}.$$

[4]The vector $(\mathbf{r}_1, \mathbf{r}_2)$ is the cointegration vector and is only determined up to scale.

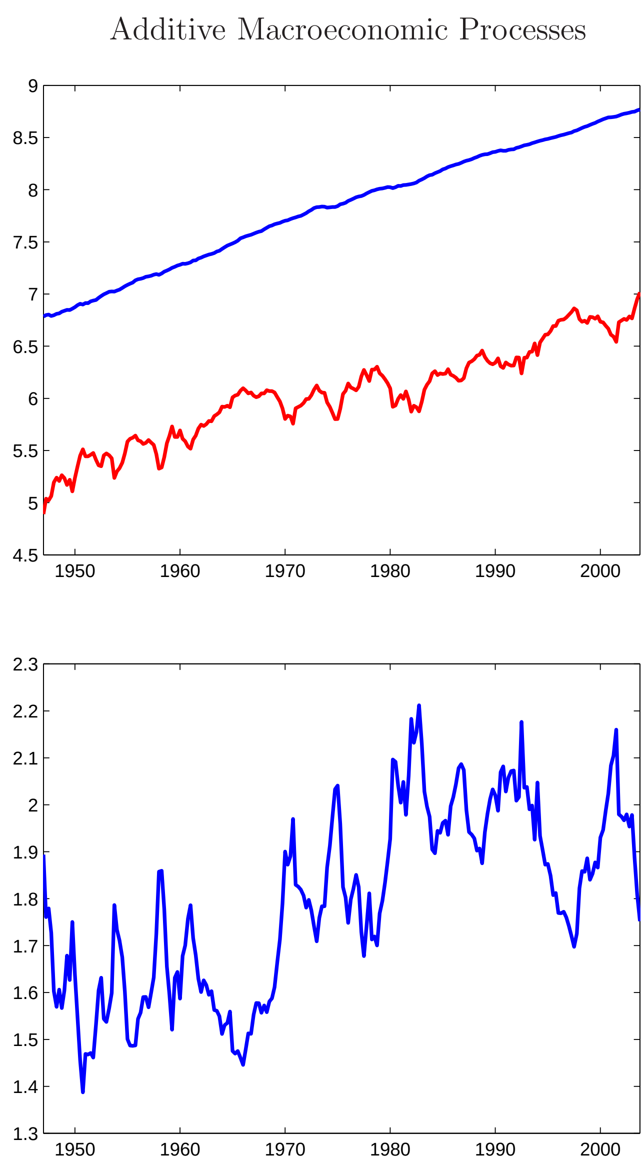

Figure 2.1: The top panel plots the logarithm of consumption (smooth blue series) and logarithm of corporate earnings (choppy red series). The bottom panel plots the difference in the logarithms of consumption and corporate earnings.

where H and F are vectors with the same dimensions as X_t and W_{t+1}, respectively.

We use this model to illustrate the four-step construction in algorithm 2.2.4.

(i) Form the conditional growth rate

$$\bar{f}(x) = D \cdot x + \nu$$

and the deviation

$$\tilde{\kappa}(X_{t+1}, X_t) = F \cdot W_{t+1}.$$

(ii) Remove the unconditional mean:

$$f(x) = D \cdot X_t + \nu - \nu = D \cdot X_t.$$

Here we are using that the unconditional mean of X is 0 because A is a stable matrix.

(iii) Where $g(x) = (\mathbb{I} - \mathbb{T})^{-1} f(x) = D'(I - A)^{-1}x$, form

$$\begin{aligned} \breve{\kappa}(x^*, x) &= f(x) + g(x^*) - g(x) \\ &= D \cdot x + D'(I - A)^{-1}(Ax + Bw^*) - D'(I - A)^{-1}x \\ &= [B'(I - A')^{-1}D] \cdot w^*. \end{aligned}$$

(iv) It follows that $\hat{\kappa} = \breve{\kappa} + \tilde{\kappa}$ is

$$\hat{\kappa}(X_{t+1}, X_t) = [F + B'(I - A')^{-1}D] \cdot W_{t+1}. \tag{2.3}$$

Blanchard and Quah (1989) use formula (2.3) in conjunction with a version of the decomposition in proposition 2.2.5 to identify a *supply shock* or a *technology shock*. In their application, the growth rate of output is one of the components of X_t, and it is assumed that only supply shocks or technology shocks have long-run consequences. Then $F + B'(I - A')^{-1}D$ identifies the linear combination of W_{t+1} that is the technology shock. This idea has been extended to include more than one shock with long-run consequences by Shapiro and Watson (1988) and Fisher (2006).

Similarly, for Beveridge and Nelson (1981), $[F + B(I - A')^{-1}D] \cdot W_{t+1}$ is the permanent shock in a permanent-transitory decomposition of a univariate time series model. When $\{W_{t+1} : t = 0, 1, ...\}$ is a univariate process, permanent and transitory shocks are necessarily (perfectly) correlated, but in a multivariate setting, transitory shocks can be restricted to be uncorrelated with permanent shocks.

2.5 Central Limit Theory

Consider an additive martingale process $\{Y_t : t = 0, 1, ...\}$ whose increments $Y_{t+1} - Y_t$ are stationary, ergodic,[5] martingale differences:

$$E\left[Y_{t+1} - Y_t | \mathfrak{F}_t\right] = 0.$$

Billingsley (1961) shows that this process obeys a central limit theorem:

$$\frac{1}{\sqrt{t}} Y_t \Longrightarrow N(0, E[(Y_{t+1} - Y_t)^2])$$

where $\Longrightarrow$ means convergence in distribution. This central limit theorem looks standard except that the terms in the sum (the increments in the additive process) are martingale differences rather than iid.

Gordin (1969) extends this result to allow for temporally dependent increments. We can regard Gordin's result as an application of Proposition 2.2.5. Under the assumptions of this proposition:

$$\frac{1}{\sqrt{t}} Y_t \Longrightarrow N(0, \sigma^2)$$

provided that $\eta = 0$. The variance used in the central limit approximation is

$$\sigma^2 = \lim_{t \to \infty} \frac{1}{t} \text{variance}(Y_t) = E\left(\left[\hat{\kappa}(X_j, X_{j-1})\right]^2\right).$$

a long-term concept that takes account of the temporal dependence of the increments.

Corollary 2.5.1. *(Gordin (1969)) Under the assumptions of Proposition 2.2.5,*

$$\frac{1}{\sqrt{t}} Y_t \Longrightarrow N(0, \sigma^2)$$

where $\sigma^2 = E\left(\left[\hat{\kappa}(X_j, X_{j-1})\right]^2\right)$.[6]

By way of illustration, we return to the first-order VAR example with $\nu = 0$:

$$X_{t+1} = AX_t + BW_{t+1}$$

[5]Ergodicity can be dispensed with if we replace the variance by $E[(Y_1 - Y_0)^2 | \mathfrak{I}]$.

[6]Hall and Heyde (1980) show how to extend this approach to functional counterparts to the Central Limit Theorem.

$$Y_{t+1} - Y_t \quad = \quad D \cdot X_t + F \cdot W_{t+1}$$

The variance that appears in this Central Limit Theorem is that of the martingale increment:

$$\sigma^2 = [F + B'(I - A')^{-1} D] \cdot [F + B'(I - A')^{-1} D]'.$$

This differs from both the conditional variance $|F|^2$ of Y_{t+1} and the unconditional variance, $D'\Sigma D + |F|^2$, of $Y_{t+1} - Y_t$ where Σ is the covariance matrix of X_t in the implied stationary distribution

$$\Sigma = \sum_{j=0}^{\infty} (A)^j BB'(A^j)'. \tag{2.4}$$

Since linear combinations of additive functionals are additive, Corollary 2.5.1 has a direct extension for multivariate counterparts to additive processes. The corollary can be applied to any linear combination of a vector of additive processes.

2.6 An Example with Discrete States

Suppose that $\{Z_t\}$ evolves according to an n-state Markov chain with transition matrix $\mathbb{T}$. In addition suppose that $\mathbb{T}$ has only one unit eigenvalue. The realized values of Z_t are the coordinate vectors in $\mathbb{R}^n$. Let q be corresponding row eigenvector:

$$q'\mathbb{T} = q'.$$

Consider an additive functional satisfying

$$Y_{t+1} - Y_t = D \cdot Z_t + {Z_t}' F W_{t+1},$$

where $\{W_t\}$ is an iid sequence of multivariate standard normally distributed random vectors. The date t composite state vector is:

$$X_t = \begin{bmatrix} Z_t \\ W_t \end{bmatrix}.$$

This is a model with discrete changes in the conditional mean and the conditional volatility of the process $\{Y_t\}$.

First compute

$$\tilde{\kappa}(X_{t+1}, X_t) = {Z_t}' F W_{t+1},$$

and

$$\nu = D \cdot q,$$

Let

$$f = D - \nu \mathbf{1}_n.$$

Then

$$g = (\mathbb{I} - \mathbb{T})^{-1} f$$

and $\check{\kappa}(x^*, x) = f \cdot z + g \cdot z^* - g \cdot z$. Then

$$Y_t = t\nu + \left[\sum_{j=1}^{t} \hat{\kappa}(X_j, X_{j-1}) \right] - g \cdot Z_t + g \cdot Z_0$$

where $\hat{\kappa} = \check{\kappa} + \tilde{\kappa}$. Notice that in this example the martingale increment has a continuous and a discrete component:

$$\hat{\kappa}(X_{t+1}, X_t) \quad = \quad \underbrace{{Z_t}' F W_{t+1}}_{\textbf{continuous}} \quad + \quad \underbrace{g \cdot Z_{t+1} - g \cdot Z_t + f \cdot Z_t}_{\textbf{discrete}}.$$

2.7 A Quadratic Model of Growth

Suppose that $\{X_t\}$ follows the first-order autoregression:

$$X_{t+1} = A X_t + B W_{t+1}$$

where A has stable eigenvalues, $B'B$ is nonsingular and $\{W_{t+1}\}$ is a sequence of independent and identically normally distributed random variables with mean zero and covariance matrix I. Consider an additive functional $\{Y_t\}$ given by

$$Y_{t+1} - Y_t = \epsilon + D \cdot X_t + \frac{1}{2} {X_t}' H X_t + F \cdot W_{t+1} + {X_t}' G W_{t+1}$$

where H is a symmetric matrix. First compute

$$\tilde{\kappa}(X_{t+1}, X_t) = F \cdot W_{t+1} + {X_t}' G W_{t+1}.$$

Next compute

$$\nu = \epsilon + E\left(\frac{1}{2}{X_t}'HX_t\right) = \epsilon + \frac{1}{2}\text{trace}(H\Sigma)$$

where Σ the covariance matrix in a stochastic steady state given by formula (2.4), and

$$f(x) = D \cdot x + \frac{1}{2}x'Hx - \frac{1}{2}\text{trace}(H\Sigma).$$

Recall that $g - \mathbb{T}g = f$ and guess that

$$g(x) = \hat{D} \cdot x + \frac{1}{2}x'\hat{H}x - \frac{1}{2}\text{trace}\left(\hat{H}\Sigma\right).$$

This gives rise to the following three relations:

$$\begin{aligned} \hat{D} - A'\hat{D} &= D, \\ \hat{H} - A'\hat{H}A &= H. \end{aligned} \tag{2.5}$$

It may be verified that

$$\begin{aligned} \hat{H} &= \sum_{j=0}^{\infty} \left(A^j\right)' H \left(A^j\right) \\ \hat{D} &= (I - A')^{-1} D. \end{aligned}$$

Since $\Sigma = BB' + A\Sigma A$,

$$\begin{aligned} \text{trace}\left(\hat{H}\Sigma\right) &= \text{trace}\left(\hat{H}BB'\right) + \text{trace}\left(\hat{H}A\Sigma A'\right) \\ &= \text{trace}\left(B'\hat{H}B\right) + \text{trace}\left(A'\hat{H}A\Sigma\right) \\ &= \text{trace}\left(B'\hat{H}B\right) + \text{trace}\left[\left(\hat{H} - H\right)\Sigma\right]. \end{aligned}$$

where the last equality follows from (2.5). Thus

$$\text{trace}\left(B'\hat{H}B\right) = \text{trace}\left(H\Sigma\right). \tag{2.6}$$

The increment in the martingale component to the additive functional is

$$\begin{aligned}
\hat{\kappa}(X_{t+1}, X_t) =& F \cdot W_{t+1} + {X_t}' G W_{t+1} + \left(B'\hat{D}\right) \cdot W_{t+1} \\
&+ \frac{1}{2} {X_{t+1}}' \hat{H} X_{t+1} + \frac{1}{2} {X_t}' \left(H - \hat{H}\right) X_t - \nu \\
=& \left(F + B'\hat{D}\right) \cdot W_{t+1} + {X_t}' \left(G + A'\hat{H}\right) W_{t+1} \\
&+ \frac{1}{2} {W_{t+1}}' B' \hat{H} B W_{t+1} - \frac{1}{2} \operatorname{trace}(H\Sigma) \\
=& \left(F + B'\hat{H}\right) \cdot W_{t+1} + {X_t}' \left(G + A'\hat{H}\right) W_{t+1} \\
&+ \frac{1}{2} {W_{t+1}}' B' \hat{H} B W_{t+1} - \frac{1}{2} \operatorname{trace}(B'\hat{H}B)
\end{aligned}$$

where the last equality follows from (2.6).

Bibliography

Alvarez, F. and U. Jermann. 2005. Using Asset Prices to Measure the the Persistence in the Marginal Utility of Wealth. *Econometrica* 73:1977–2016.

Anderson, Evan W., Lars Peter Hansen, and Thomas J. Sargent. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection. *Journal of the European Economic Association* 1:69–123.

Bansal, R. and B. N. Lehmann. 1997. Growth-optimal Portfolio Restrictions on Asset Pricing Models. *Macroeconomic Dynamics* 1:333–354.

Bansal, Ravi and Amir Yaron. 2004. Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance* 59 (4):1481–1509.

Beveridge, S. and C. R. Nelson. 1981. A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to the Measurement of the 'Business Cycle'. *Journal of Monetary Economics* 7:151–174.

Billingsley, P. 1961. The Lindeberg-Levy Theorem for Martingales. *American Mathematical Monthly* 12.

Blanchard, O. J and D. Quah. 1989. The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review* 79:655–673.

Breiman, Leo. 1968. *Probability Theory*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Cecchetti, Stephen G., Pok sang Lam, and Nelson C. Mark. 2000. Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True? *American Economic Review* 90 (4):787–805.

Chernoff, Herman. 1952. A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations. *Annals of Mathematical Statistics* 23:493–507.

Doob, J. L. 1953. *Stochastic Processes*. New York: John Wiley and Sons.

Engle, R. and C. W. J. Granger. 1987. Co-integration and Error Correction: Representation, Estimation and Testing. *Econometrica* 55:251–276.

Fisher, J. 2006. The Dynamic Effects of Neutral and Investment-Specific Technology Shocks. *Journal of Political Economy* 114:413–451.

Gordin, M. I. 1969. The Central Limit Theorem for Stationary Processes. *Soviet Mathematics Doklady* 10:1174–1176.

Hall, P. and C. C. Heyde. 1980. *Martingale Limit Theory and Its Application*. Boston: Academic Press.

Hamilton, J. D. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57:357–384.

Hansen, L. P. and J. Scheinkman. 1995. Back to the Future: Generating Moment Implications for Continuous Time Markov Processes. *Econometrica* 63:767–804.

Hansen, L. P., J. C. Heaton, and N. Li. 2008. Consumption Strikes Back?: Measuring Long Run Risk. *Journal of Political Economy* .

Hansen, Lars Peter. 2008. Modeling the Long Run: Valuation in Dynamic Stochastic Economies. Prepared for the Fisher-Schultz Lecture at the 2006 European Meetings of the Econometric Society.

Hansen, Lars Peter and Jose Alexandre Scheinkman. 2009. Long-term Risk: A Operator Approach. *Econometrica* 77 (1):177–234.

Kreps, David M. and Evan L. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice. *Econometrica* 46:185–200.

Meyn, S. and R. Tweedie. 1993. *Markov Chains and Stochastic Stability*. London: Springer-Verlag.

Sclove, S. L. 1983. Time-Series Segmentation: A Method and a Model. *Information Science* 29:7–25.

Shapiro, M. and M. Watson. 1988. Sources of Business Cycle Fluctuations. *NBER Macroeconomics Annual* 3:111–148.

Tallarini, Thomas D. 2000. Risk-sensitive real business cycles. *Journal of Monetary Economics* 45 (3):507–532.