Foundations and Trends® in
Finance
Vol. 6, No. 3 (2011) 165–219

DOI: 10.1561/0500000037

now
the essence of knowledge

Continuous-Time Linear Models

By John Cochrane

Contents

Foundations and Trends® in
Finance
Vol. 6, No. 3 (2011) 165–219

DOI: 10.1561/0500000037

now
the essence of knowledge

Continuous-Time Linear Models

John H. Cochrane

University of Chicago Booth School of Business and NBER.
5807 S. Woodlawn, Chicago, IL 60637, USA,
john.cochrane@chicagobooth.edu

Abstract

I translate familiar concepts of discrete-time time series to contnuous-time equivalent. I cover lag operators, ARMA models, the relation between levels and differences, integration and cointegration, and the Hansen–Sargent prediction formulas.

1

Introduction

Discrete-time linear ARMA processes and lag operator notation are convenient for lots of calculations. Continuous-time representations often simplify economic models, and can handle interesting nonlinearities as well. But standard treatments of continuous-time processes typically don't mention how to adapt the discrete-time linear model concepts and lag operator methods to continuous time. Here I attempt that translation.

The point of this monograph is to exposit the techniques, understand the intuition, and to make the translation from familiar discrete-time ideas. I do not pretend to offer anything new. I also don't discuss the technicalities. Hansen and Sargent (1991) is a good reference. Heaton (1993) describes many of these methods and provides a useful application. I assume basic knowledge of discrete-time time-series representation methods and continuous-time representations. Cochrane (2005a,b) cover the necessary background, but any standard reference covers the same material.

The concluding section collects the important formulas in one place.

2

Linear Models and Lag Operators

I start by defining lag operators and the inversion formulas.

2.1 Discrete Time Operators

As a reminder, discrete-time linear models can be written in the unique moving average or Wold representation

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \mathcal{Z}_b(L)\varepsilon_t, \tag{2.1}$$

where the operator L is defined by

$$L\varepsilon_t = \varepsilon_{t-1} \tag{2.2}$$

and

$$\mathcal{Z}_b(L) = \sum_{j=0}^{\infty} b_j L^j; \quad \mathcal{Z}_b(0) = b_0 = 1.$$

This last condition means that we define the variance of the shocks so that ε_t is the innovation in x_t.

The Wold representation and its error are defined from the autoregression

$$x_t = -\sum_{j=1}^{\infty} a_j x_{t-j} + \varepsilon_t. \tag{2.3}$$

We can write this autoregressive representation in lag-operator form,

$$\mathcal{Z}_a(L)x_t = \varepsilon_t; \quad \mathcal{Z}_a(L) = \sum_{j=0}^{\infty} a_j x_{t-j};\ \mathcal{Z}_a(0) = a_0 = 1. \tag{2.4}$$

We can connect the autoregressive and moving average representations by inversion,

$$\mathcal{Z}_a(L) = \mathcal{Z}_b(L)^{-1}; \quad \mathcal{Z}_b(L) = \mathcal{Z}_a(L)^{-1}.$$

To construct the inverse $\mathcal{Z}_a(L)^{-1}$, given the definition (2.2), we use a power-series interpretation. For example, suppose we want to invert the AR(1)

$$\mathcal{Z}_a(L)x_t = (1 - \rho L)x_t = \varepsilon_t.$$

To interpret $1/(1 - \rho L)$ – to find a $\mathcal{Z}_b(L)$ such that $\mathcal{Z}_b(L)\mathcal{Z}_a(L) = I$ – we use the expansion

$$\frac{1}{(1 - \rho L)} = \sum_{j=0}^{\infty} \rho^j L^j$$

for $\|\rho\| < 1$. With this interpretation, we can use the lag operator notation to represent the transformation from AR(1) to MA(∞) representations and back again,

$$(1 - \rho L)x_t = \varepsilon_t \Longleftrightarrow$$

$$x_t = \frac{1}{(1 - \rho L)}\varepsilon_t = \left(\sum_{j=0}^{\infty} \rho^j L^j\right)\varepsilon_t = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}. \tag{2.5}$$

2.2 A Note on Linear Processes

The fundamental autoregressive representations are linear; the conditional mean $E_t(x_{t+j})$ is a linear function of past x_t and the conditional

variance is constant. The process $\{x_t\}$ may *also* have a nonlinear representation, which allows greater predictability. For example, a random number generator is fully deterministic, $x_t = f(x_{t-1})$ with no error. The function f is just so complex that when you run linear regressions of x_t on its past, x_t looks unpredictable. A precise notation would use $E_{t-1}(x_t) = E(x_t|x_{t-1}, x_{t-2}, \ldots)$ to denote prediction using all linear and nonlinear functions, i.e. conditional expectation, which would give $E_{t-1}(x_t) = f(x_{t-1}) = x_t$ in this example. We would use a notation such as $P(x_t|x_{t-1}, x_{t-2}, \ldots)$ to denote linear prediction. I will not be so careful, so I will use E_{t-1} or $E(x_t|x_{t-1}, x_{t-2}, \ldots)$ and the word "expectation" to mean prediction given the linear models under consideration.

This clarification is especially important as we go to continuos time. One may object that a linear model is not "right" if there is an underlying "better" nonlinear model, say a square root process. That criticism is incorrect. Even if there is an underlying true, or better-predicting, nonlinear model, there is nothing wrong with *also* studying the processes' linear predictive representation. Analogously, just because there may be additional variables y_t, z_t that help to forecast x_{t+1}, there is nothing wrong with studying conditional (on past x_t alone) moments that ignore this extra information.

The conditioning-down assumption can cause trouble if you assume agents in a model only see the variables or information set that you the econometrician choose to model. But one does not have to make that assumption in order to study linear or otherwise conditioned-down representations.

2.3 Continuous-time Operators

We usually write continuous-time processes in differential or integral form. For example, the continuous-time AR(1) can be written in differential form,

$$dx_t = -\phi x_t dt + \sigma dB_t$$

or in integral form

$$x_t = \int_{\tau=0}^{\infty} e^{-\phi\tau} \sigma dB_{t-\tau},$$

where dB_t denotes increments to standard Brownian motion. I write the shock as σdB_t to preserve the discrete-time convention that a unit shock to the error is a unit shock to x_t, and the continuous-time convention that Brownian motion has a unit variance.

This integral form is the obvious analogue to the moving-average form of the discrete-time representation (2.5). Our job is to think about and manipulate these kinds of expressions using lag operators.

The lag operator can straightforwardly be extended to real numbers from integers, i.e.

$$L^{\tau} x_t = x_{t-\tau}.$$

Since we write differential expressions, dx_t in continuous time, it is convenient to define the differential operator D, i.e.

$$Dx_t = \frac{1}{dt} dx_t \tag{2.6}$$

where dx_t is the familiar continuous-time forward-difference operator,

$$dx_t = \lim_{\Delta \to 0} \left(x_{t+\Delta} - x_t \right). \tag{2.7}$$

(This is not a limit in the usual ε, δ sense, but I'll leave that to continuous time math books and continue to abuse notation.)

The D and L operators are related by

$$e^{-D} = L; \quad D = -\log(L). \tag{2.8}$$

We can see this relationship directly: From (2.6),

$$\begin{aligned} D &= \lim_{\Delta \to 0} \frac{L^{-\Delta} - 1}{\Delta} = \lim_{\Delta \to 0} \frac{e^{-\Delta \log(L)} - 1}{\Delta} \\ &= \lim_{\Delta \to 0} \frac{-\log(L) e^{-\Delta \log(L)}}{1} = -\log(L). \end{aligned}$$

Now we are ready to write the obvious general moving average processes:

$$x_t = \int_{\tau=0}^{\infty} b(\tau) \sigma dB_{t-\tau} = \mathcal{L}_b(D) \sigma D B_t \tag{2.9}$$

where we define

$$\mathcal{L}_b(D) = \int_{\tau=0}^{\infty} e^{-D\tau} b(\tau) d\tau; \; b(0) = 1.$$

Mirroring the convention that $b_0 = 1$ in discrete time, so that shocks ε_t translate one-to-one to shocks to x_t, I write the continuos time shock σDB_t with DB_t standard Brownian motion (variance $\sigma^2 dt$) and impose the normalization $b(0) = 1$.

It is useful to verify just how each step of this operation works:

$$\begin{aligned}\mathcal{L}_b(D)\sigma DB_t &= \int_{\tau=0}^{\infty} e^{-D\tau} b(\tau) d\tau \left(\frac{1}{dt}\sigma dB_t\right) \\ &= \int_{\tau=0}^{\infty} b(\tau) e^{-D\tau} \sigma dB_t = \int_{\tau=0}^{\infty} b(\tau)\sigma dB_{t-\tau}.\end{aligned}$$

Though it breaks the analogy with discrete time a bit, it is more convenient to describe continuous-time lag functions in terms of D rather than L. We could have written $\mathcal{Z}_b(L) = \int_{\tau=0}^{\infty} L^\tau b(\tau) d\tau$. However, we will have to use the D operator frequently, to describe Dx_t and DB_t, so it is simpler to use D everywhere. This change means that familiar quantities from discrete time such as the impact multiplier $\mathcal{Z}_b(L = 0)$ and the cumulative multiplier $\mathcal{Z}_b(L = 1)$ will have counterparts corresponding to $\mathcal{L}_b(D = -\infty)$ and $\mathcal{L}_b(D = 0)$.

For example, the continuous-time AR(1) process in differential form reads

$$\begin{aligned}dx_t + \phi x_t \, dt &= \sigma dB_t \\ (D + \phi)x_t &= \sigma DB_t.\end{aligned}$$

We can "invert" this formula by inverting the "lag operator polynomial" as we do in discrete time:

$$\begin{aligned}x_t = \left(\frac{1}{D+\phi}\right)\sigma DB_t &= \left(\int_{\tau=0}^{\infty} e^{-\phi\tau} e^{-D\tau} \, d\tau\right)\sigma DB_t \\ = \left(\int_{\tau=0}^{\infty} e^{-\phi\tau} DL^\tau \, d\tau\right)\sigma B_t &= \int_{\tau=0}^{\infty} e^{-\phi\tau}\sigma dB_{t-\tau}.\end{aligned}$$

The second equality uses the formula for the integral of an exponential $\int_{\tau=0}^{\infty} e^{-(D+\phi)\tau} d\tau$ to interpret $1/(D + \phi)$ given the definition of D, as we used the power series expansion $\sum_{j=0}^{\infty} \rho^j L^j$ to interpret $1/(1 - \rho L)$ given the definition of L.

2.4 Laplace Transforms

The justification for these techniques fundamentally comes from Laplace transforms. While it is not necessary to know a lot about Laplace transforms to use lag and differential operators, it helps to have some familiarity with the underlying idea.

If a process $\{y_t\}$ is generated from another $\{x_t\}$ by

$$y_t = \int_{\tau=0}^{\infty} b(\tau) x_{t-\tau} d\tau,$$

the Laplace transform of this operation is defined as

$$\mathcal{L}_b(D) = \int_{\tau=0}^{\infty} e^{-D\tau} b(\tau) d\tau$$

where D is a complex number.

Given this definition, the Laplace transform of the lag operation $y_t = L^j x_t = x_{t-j}$ is

$$\mathcal{L}_{L^j}(D) = e^{-jD}.$$

This definition directly establishes the relationship between lag and differential operators (2.8), avoiding my odd-looking limits.

One difference in notation between discrete and continuous-time notation is necessary. It is common to write the discrete-time lag polynomial as

$$b(L) = \sum_{j=0}^{\infty} b_j L^j.$$

It would be nice to write similarly

$$b(D) = \int_{\tau=0}^{\infty} e^{-D\tau} b(\tau) d\tau,$$

but we cannot do that, since $b(\tau)$ is already a function. If in discrete time we had written $b_j = b(j)$, then $b(L)$ would not have made any sense either. For this reason, we will have to use a different letter. In deference to the Laplace transform I use the notation

$$\mathcal{L}_b(D) \equiv \int_{\tau=0}^{\infty} e^{-D\tau} b(\tau) d\tau.$$

For clarity I also write discrete-time lag polynomial functions as

$$\mathcal{Z}_b(L) = \sum_{j=0}^{\infty} b_j L^j$$

rather than the more common $b(L)$. ($\mathcal{Z}$ stands for z-transform, the discrete counterpart to Laplace transforms.)

To use a lag polynomial expansion

$$\mathcal{Z}_b(L) = \frac{1}{1-\rho L} x_t = \sum_{j=0}^{\infty} \rho^j L^j \varepsilon_{t-j},$$

we must have $\|\rho\| < 1$. In general, the poles $L : \mathcal{Z}_b(L) = \infty$ and the roots $L : \mathcal{Z}_a(L) = \mathcal{Z}_b(L)^{-1} = 0$ must lie outside the unit circle. The domain of $\mathcal{Z}_b(L)$ is $\|L\| < \|\rho\|^{-1}$; for which $\|L\| < 1$ will suffice.

When $\rho > 1$, or if the poles of $\mathcal{Z}_b(L)$ are inside the unit circle, we solve in the opposite direction:

$$\begin{aligned} \|\rho\| > 1 \Longrightarrow \frac{1}{1-\rho L}\varepsilon_t &= -\frac{\rho^{-1}L^{-1}}{1-\rho^{-1}L^{-1}}\varepsilon_t = -\left(\sum_{j=1}^{\infty} \rho^{-j} L^{-j}\right)\varepsilon_t \\ &= -\sum_{j=1}^{\infty} \rho^{-j}\varepsilon_{t+j}. \end{aligned}$$

In the corresponding general case, the domain of $\mathcal{Z}_b(L)$ must be L *outside* the unit circle.

Similarly, to interpret

$$\begin{aligned} \mathcal{L}_b(D)DB_t &= \frac{1}{\phi + D}\sigma DB_t = \left(\int_{\tau=0}^{\infty} e^{-\phi\tau} e^{-D\tau}\, d\tau\right)\sigma DB_t \\ &= \int_{\tau=0}^{\infty} e^{-\phi\tau}\sigma dB_{t-\tau} \end{aligned}$$

we must have $\|\phi\| > 0$, and the domain $\mathrm{Re}(D) > 0$ so that $\|e^{-D}\| < 1$. More generally, the poles $\mathcal{L}_b(D)$ must lie where $\mathrm{Re}(D) < 0$, i.e. where $L = e^{-D}$ is outside the unit circle.

In the other circumstance, we expand forward, i.e.

$$\begin{aligned}\mathcal{L}_b(D)\sigma DB_t &= \frac{1}{\phi - D}\sigma dB_t = \left(\int_{\tau=0}^{\infty} e^{-\phi\tau}e^{D\tau}\ d\tau\right)\sigma DB_t \\ &= \int_{\tau=0}^{\infty} e^{-\phi\tau}\sigma dB_{t+\tau},\end{aligned}$$

and use the domain $\text{Re}(D) < 0$ so that $\|e^D\| < 1$. More generally, in this case the poles of $\mathcal{L}_b(D)$ must lie where $\text{Re}(D) > 0$, i.e. where $L = e^{-D}$ is inside the unit circle. (Here I found it clearer to keep $\phi > 0$ and introduce the negative sign directly.)

Sometimes operators $\mathcal{L}_b(D)$ will have poles at both positive and negative values of $\text{Re}(D)$. Then, as in discrete time, we solve "unstable" roots forward and stable roots backward, and obtain an integral that runs over both past and future dB_t.

Lag operators (Laplace transforms) commute, so we can simplify expressions by taking them in any order that is convenient,

$$\begin{aligned}\mathcal{L}_a(D)\mathcal{L}_b(D) &= \mathcal{L}_b(D)\mathcal{L}_a(D), \\ \mathcal{Z}_a(L)\mathcal{Z}_b(L) &= \mathcal{Z}_b(L)\mathcal{Z}_a(L).\end{aligned}$$

This is one of the great simplifications allowed by operator representations. More generally, lots of the hard integrals one runs into while manipulating lag operators are special cases of well-known Laplace transform tricks, and looking up the latter can save a lot of time.

3

Moving Average Representation and Moments

The moving average representation

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \mathcal{Z}_b(L)\varepsilon_t$$

is also a basis for all the second-moment statistical properties of the series. The variance is

$$\sigma^2(x_t) = \left(\sum_{j=0}^{\infty} b_j^2 \right) \sigma_\varepsilon^2,$$

the covariance is

$$\operatorname{cov}(x_t, x_{t-k}) = \left(\sum_{j=0}^{\infty} b_j b_{j+k} \right) \sigma_\varepsilon^2,$$

and the spectral density is

$$S_x(\omega) = \sum_{j=-\infty}^{\infty} e^{-i\omega j} \operatorname{cov}(x_t, x_{t-j}) = \mathcal{Z}_b(e^{i\omega}) \mathcal{Z}_b(e^{-i\omega}) \sigma_\varepsilon^2.$$

The inversion formula

$$\text{cov}(x_t, x_{t-k}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega k} S_x(\omega) d\omega = \frac{\sigma_\varepsilon^2}{2\pi}\int_{-\pi}^{\pi} e^{i\omega k} \mathcal{Z}_b(e^{i\omega})\mathcal{Z}_b(e^{-i\omega}) d\omega$$

gives us a direct connection between the function $\mathcal{Z}_b(e^{i\omega})$ and the second moments of the series. The variance formula quickly shows you why square-summable lag coefficients, $\sum_{j=0}^{\infty} b_j^2 < \infty$ are a standard technical condition on the moving-average representation.

The continuous-time moving-average representation

$$x_t = \int_{\tau=0}^{\infty} b(\tau)\sigma dB_{t-\tau} = \mathcal{L}_b(D)\sigma D B_t$$

is also the basis for standard moment calculations,

$$\begin{aligned}
\sigma^2(x_t) &= \left(\int_{\tau=0}^{\infty} b^2(\tau) d\tau\right)\sigma^2, \\
\text{cov}(x_t, x_{t-k}) &= \left(\int_{\tau=0}^{\infty} b(\tau) b(\tau + k) d\tau\right)\sigma^2 \\
S_x(\omega) &= \mathcal{L}_b(i\omega)\mathcal{L}_b(-i\omega)\sigma^2,
\end{aligned}$$

and the inversion formula

$$\text{cov}(x_t, x_{t-k}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega k} S_x(\omega) d\omega = \frac{\sigma^2}{2\pi}\int_{-\infty}^{\infty} e^{i\omega k}\mathcal{L}_b(i\omega)\mathcal{L}_b(-i\omega) d\omega.$$

The variance formula shows we we impose $\int_{\tau=0}^{\infty} b^2(\tau) d\tau < \infty$.

For example, the AR(1) gives

$$\begin{aligned}
x_t &= \int_{\tau=0}^{\infty} e^{-\phi\tau}\sigma dB_{t-\tau} = \frac{1}{\phi + D}\sigma D B_t \\
\sigma^2(x) &= \sigma^2\int_{\tau=0}^{\infty} e^{-2\phi\tau} d\tau = \frac{\sigma^2}{2\phi} \\
cov(x_t, x_{t-k}) &= \sigma^2\int_{\tau=0}^{\infty} e^{-\phi\tau} e^{-\phi(t+k)} d\tau = \frac{\sigma^2}{2\phi} e^{-\phi k} \\
S_x(\omega) &= \mathcal{L}_b(i\omega)\mathcal{L}_b(-i\omega) = \frac{\sigma}{\phi + i\omega}\frac{\sigma}{\phi - i\omega} = \frac{\sigma^2}{\phi^2 + \omega^2}.
\end{aligned}$$

4

ARMA Models

In discrete time, ARMA models provide a tractable class that generalizes the AR(1) and captures interesting dynamics. Here, I describe the counterpart to those models in continuous time.

4.1 Discrete Time

We can write ARMA models in lag-polynomial notation

$$(1-\lambda_1 L)(1-\lambda_2 L)\ldots x_t = (1+\theta_1 L)(1+\theta_2 L)\ldots \varepsilon_t. \tag{4.1}$$

We can express these processes in autoregressive form

$$\frac{(1-\lambda_1 L)(1-\lambda_2 L)\ldots}{(1+\theta_1 L)(1+\theta_2 L)\ldots} x_t = \varepsilon_t$$

or moving average form

$$x_t = \frac{(1+\theta_1 L)(1+\theta_2 L)\ldots}{(1-\lambda_1 L)(1-\lambda_2 L)\ldots}\varepsilon_t$$

To calculate and interpret the denominator polynomials, it is useful to use partial fraction decompositions,

$$\frac{1}{(1-\lambda_1 L)(1-\lambda_2 L)\cdots} = \frac{A}{1-\lambda_1 L} + \frac{B}{1-\lambda_2 L} + \cdots$$

For example, the AR(2) is equivalent in this way to the sum of two AR(1),

$$x_t = \frac{1}{(1-\lambda_1 L)(1-\lambda_2 L)}\varepsilon_t = \left(\frac{\frac{\lambda_1}{\lambda_1-\lambda_2}}{1-\lambda_1 L} + \frac{\frac{\lambda_2}{\lambda_2-\lambda_1}}{1-\lambda_2 L}\right)\varepsilon_t \quad (4.2)$$

$$= \frac{\lambda_1}{\lambda_1-\lambda_2}\sum_{j=0}^{\infty}\lambda_1^j\varepsilon_{t-j} + \frac{\lambda_2}{\lambda_2-\lambda_1}\sum_{j=0}^{\infty}\lambda_2^j\varepsilon_{t-j}. \quad (4.3)$$

4.2 Continuous Time

The continuous-time analogue to lag-operator polynomial models are differential-operator polynomial models, of the form

$$x_t = \frac{(D+\theta_1)(D+\theta_2)\cdots}{(D+\lambda_1)(D+\lambda_2)(D+\lambda_3)\cdots}\sigma DB_t \quad (4.4)$$

Unlike the discrete-time case, the order of the denominator must always be one greater than the order of the numerator, for reasons I discuss below.

The partial-fractions decomposition is useful to understand the moving-average form of (4.4). For example, the next-simplest model after the AR(1) is

$$x_t = \frac{(D+\theta_1)}{(D+\lambda_1)(D+\lambda_2)}\sigma DB_t$$

$$= \frac{1}{\lambda_1-\lambda_2}\left(\frac{\lambda_1-\theta_1}{D+\lambda_1} - \frac{\lambda_2-\theta_1}{D+\lambda_2}\right)\sigma DB_t$$

$$= \frac{\lambda_1-\theta_1}{\lambda_1-\lambda_2}\int_{\tau=0}^{\infty} e^{-\lambda_1\tau}\sigma dB_{t-\tau} + \frac{\lambda_2-\theta_1}{\lambda_2-\lambda_1}\int_{\tau=0}^{\infty} e^{-\lambda_2\tau}\sigma dB_{t-\tau}. \quad (4.5)$$

This formula is the analogue of the AR(2) expressed as the sum of two AR(1) in (4.2). More generally, we can express (4.4) as

$$x_t = \left[\frac{A}{(D+\lambda_1)} + \frac{B}{(D+\lambda_2)} + \frac{C}{(D+\lambda_3)} + \cdots\right]\sigma DB_t \quad (4.6)$$

and understand the general process (4.4) as a sum of many AR(1)s. The normalization $b(0) = 1$ implies

$$A + B + C + \cdots = 1,$$

which you can verify in (4.5).

To understand the autoregressive representation of this polynomial operator,

$$\frac{(D+\lambda_1)(D+\lambda_2)(D+\lambda_3)\cdots}{(D+\theta_1)(D+\theta_2)\cdots}x_t = \sigma DB_t \tag{4.7}$$

it is useful to reexpress the differential-operator polynomial in a different way. For example, we can write the second-order model

$$\frac{(D+\lambda_1)(D+\lambda_2)}{(D+\theta_1)}x_t = \sigma DB_t \tag{4.8}$$

in the form

$$\left[D + (\lambda_1+\lambda_2-\theta_1) + \frac{(\theta_1-\lambda_1)(\theta_1-\lambda_2)}{D+\theta_1}\right]x_t = \sigma DB_t$$

or, writing it out,

$$\begin{aligned} dx_t = &-[(\lambda_1+\lambda_2)-\theta_1]x_t\,dt \\ &-\left((\theta_1-\lambda_1)(\theta_1-\lambda_2)\int_{\tau=0}^{\infty} e^{-\theta_1\tau}x_{t-\tau}d\tau\right)dt + \sigma DB_t \end{aligned}$$

Here you see a natural generalization of the AR(1), and see the "autoregressive" nature of the process. We forecast dx_t as a linear function of the history of $\{x_t\}$. More generally, we can express (4.7) in the form

$$\left[D + A + \frac{B}{D+\theta_1} + \frac{C}{D+\theta_2} + \cdots\right]x_t = \sigma DB_t$$

In this form, we forecast dx_t by its level x_t and a sum of geometrically-weigthed integrals over the history of x_t.

4.3 How not to Define ARMA Models

The class of models I described in (4.4) displays some notable differences from the discrete-time ARMA class that I used to motivate them. Other natural attempts to take ARMA models to continuous time do not work.

First, I announced the rule that the order of the numerator in (4.4) must be one less than the denominator, while the order of polynomials in (4.1) is arbitrary. The underlying reason for this difference is that,

while the L^2 operator takes a double lag, the D^2 operator takes a second derivative. For example, consider

$$D^2 x_t = \sigma D B_t$$

writing it out, this means

$$\frac{1}{dt}\, d\left(\frac{1}{dt}\, dx_t\right) = \sigma \frac{1}{dt}\, dB_t$$

$$dx_t = \left(\frac{1}{dt}\, dx_0 + \sigma \int_{\tau=0}^{t} dB_\tau\right) dt.$$

This dx_t does not have a dB_t term. x_t is a differentiable function of time, perfectly forecastable dt ahead. Taking the D^2 operator takes us out of the kind of process we are looking for.

As a less trivial example, suppose we tried to write a "continuous time AR(2)" as

$$(D + \lambda_1)(D + \lambda_2) x_t = \sigma D B_t.$$

Then, we would have

$$(D + \lambda_1) x_t = \frac{1}{(D + \lambda_2)} \sigma D B_t$$

$$dx_t + \lambda_1 x_t\, dt = \left(\int_{\tau=0}^{\infty} e^{-\lambda_2 \tau} \sigma dB_{t-\tau}\right) dt$$

Again, we lose the σdB_t term and x_t is differentiable.

Second, the main feature of ARMA models, that only a finite past of $\{x_t\}$ or shocks $\{\varepsilon_t\}$ forms a state vector for forecasting, is not preserved in models of the form (4.4). One could create perfectly good models with that feature, but those models do not have the convenience or tractability that they posses in discrete time. For example, we can write finite-length processes such as

$$dx_t = \left(\gamma x_t + \int_{\tau=0}^{k} a(\tau) x_{t-\tau}\right) d\tau + \sigma dB_t$$

or a finite-length moving average

$$x_t = \int_{\tau=0}^{k} b(\tau) dB_{t-\tau}.$$

But the finiteness k of the AR or MA representation does not lead to easy inversion or manipulation as it does in discrete time.

Similarly, we could try to take the continuous-time limit of an AR(2) by keeping the second lag fixed, not letting it contract towards zero so that it create the troublesome second derivative. We would start with

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t$$
$$x_t - x_{t-1} = -(1-\rho_1)x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t$$

Then, take the limit by letting the first difference get smaller but keeping the second lag fixed. We get

$$dx_t = (-\phi x_t + \phi_2 x_{t-\kappa})dt + dB_t$$
$$(D + \phi + e^{-\kappa D})x_t = DB_t$$

with $\kappa = 2$. This is a legitimate process, but the tractability is clearly lost, as inverting this lag operator will not be fun.

5

Differences

In discrete time, you usually choose to work with levels x_t or differences Δx_t depending on which is stationary. In continuous time, we often work with differences even though the series is stationary in levels. For example, we write the continuous-time AR(1) as $dx_t = -\phi x_t\, dt + \sigma dB_t$, which corresponds to expressing the discrete-time AR(1) as $x_{t+1} - x_t = -(1-\rho)x_t + \varepsilon_{t+1}$. This fact accounts for the major difference between the look of continuous and discrete-time formulas, and means we must spend a little more time than usual describing the relation between level and differenced processes.

5.1 Levels to Differences in Discrete Time

First-differencing is simple in discrete time. Given a process in levels,

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}$$

we can write the same process in differences as

$$x_t - x_{t-1} = \varepsilon_t + \sum_{j=1}^{\infty} (b_j - b_{j-1})\varepsilon_{t-j}. \tag{5.1}$$

In operator notation, we transform from the moving average for levels

$$x_t = \mathcal{Z}_b(L)\varepsilon_t \tag{5.2}$$

to a moving average for differences

$$(1 - L)x_t = \mathcal{Z}_c(L)\varepsilon_t. \tag{5.3}$$

One way to construct $\mathcal{Z}_c(L)$ is straightforwardly shown by (5.1),

$$\mathcal{Z}_c(L) = (1 - L)\mathcal{Z}_b(L) = 1 + \mathcal{Z}_{\Delta b}(L). \tag{5.4}$$

Remember, we normalized the lag polynomial so that $b_0 = \mathcal{Z}_b(0) = 1$, and so that $(E_t - E_{t-1})x_t = 1 \times \varepsilon_t$ is the impact response to a shock. In discrete time $(E_t - E_{t-1})(x_t - x_{t-1}) = 1 \times \varepsilon_t$ as well so we have $\mathcal{Z}_c(0) = 1$ and $\mathcal{Z}_{\Delta b}(0) = 0$.

5.2 Levels to Differences in Continuous Time

In continuous time, we can similarly model levels or differences,

$$x_t = \mathcal{L}_b(D)\sigma DB_t \tag{5.5}$$

or

$$Dx_t = \mathcal{L}_c(D)\sigma DB_t. \tag{5.6}$$

Obviously, we can write

$$\mathcal{L}_c(D) = D\mathcal{L}_b(D),$$

but there are several other ways to construct, express, and interpret the differenced representation given the level representation.

Mirroring (5.1) and (5.4), we can find $\mathcal{L}_c(D)$ from

$$\mathcal{L}_c(D) = D\mathcal{L}_b(D) = 1 + \mathcal{L}_{b'}(D) \tag{5.7}$$

or, explicitly,

$$dx_t = \left(\int_{\tau=0}^{\infty} b'(\tau)\sigma dB_{t-\tau}\right) dt + \sigma dB_t. \tag{5.8}$$

This formula is the obvious analogue to (5.1). However, in continuous time, this expression gives familiar drift and diffusion terms.

Expression (5.7) and the resulting (5.8) is a standard property of Laplace transforms

$$D\mathcal{L}_b(D) = b(0) + \mathcal{L}_{b'}(D) \tag{5.9}$$

together with the normalization $b(0) = 1$. To derive it, integrate by parts:

$$\begin{aligned}\mathcal{L}_{b'}(D) &= \int_{\tau=0}^{\infty} e^{-D\tau}\frac{db(\tau)}{d\tau}\,d\tau = b(\tau)e^{-D\tau}\big|_0^{\infty} + \int_{\tau=0}^{\infty} De^{-D\tau}b(\tau)d\tau \\ &= -b(0) + D\int_{\tau=0}^{\infty} e^{-D\tau}b(\tau)d\tau = -b(0) + D\mathcal{L}_b(D).\end{aligned}$$

I assume here that $b(\tau)$ is differentiable except at $\tau = 0$. The formulas can be extended to include $b(\tau)$ with jumps, which give rise to additional lagged diffusion terms. Correspondingly, to represent something like (5.8) as a Laplace transform, I allow a δ function in $c(\tau)$ at $\tau = 0$, whose Laplace transform is the constant $c(0)$. A typical moving average representation for differences will have such a delta function, i.e. its integral expansion will be of the form

$$\mathcal{L}_c(D) = c(0) + \int_{\tau=0}^{\infty} e^{-D\tau}c(\tau)d\tau.$$

In the case of a differential-operator polynomial, this transformation from levels to differences is simply algebra. For the AR(1), we can write

$$Dx_t = \frac{D}{D+\phi}\sigma DB_t = \left(1 - \frac{\phi}{D+\phi}\right)\sigma DB_t \tag{5.10}$$

i.e.

$$dx_t = -\phi\left(\int_{\tau=0}^{\infty} e^{-\phi\tau}\sigma dB_{t-\tau}\right)dt + \sigma dB_t.$$

Recognizing the first term on the right as x_t itself, you recognize the AR(1), but see that it is now written in a moving average representation for dx_t, which is what we were looking for. Construction (5.8) gives the same answer which is a fun exercise.

For the more general polynomial operator, we can apply the same algebra to the partial-fractions expansion of the moving average polynomial,

$$\begin{aligned}\mathcal{L}_c(D) &= D\mathcal{L}_b(D) = \frac{DA}{D+\lambda_1} + \frac{DB}{D+\lambda_2} + \cdots \\ &= A - \frac{\lambda_1 A}{D+\lambda_1} + B - \frac{\lambda_2 B}{D+\lambda_2} + \cdots \\ &= 1 - \frac{\lambda_1 A}{D+\lambda_1} - \frac{\lambda_2 B}{D+\lambda_2} - \cdots \end{aligned} \tag{5.11}$$

In each case, notice that $\mathcal{L}_c(0) = 0$. That follows in (5.7) with the fact that $\mathcal{L}_b(0)$ is finite, and it is clear in (5.10) and (5.11). That ends up being the condition that x_t is stationary in levels. The Beveridge–Nelson decomposition and cointegration follow later from the case of a differenced representation $Dx_t = \mathcal{L}_c(D)DB_t$ in which $\mathcal{L}_c(0) \neq 0$ or is not full rank.

6

Impulse-response Function

6.1 Discrete Time

The discrete-time moving-average representation *is* the impulse-response function. In

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \mathcal{Z}_b(L)\varepsilon_t,$$

the terms of b_j measure the response of x_{t+j} to a shock ε_t,

$$(E_t - E_{t-1})x_{t+j} = b_j \varepsilon_t.$$

In particular, we can read the impact multiplier — the response $(E_t - E_{t-1})x_t$ off the lag polynomial evaluated at $L = 0$,

$$b_0 = \mathcal{Z}_b(0) = 1; \tag{6.1}$$

we can read the cumulative response — the response of $\sum_{j=0}^{\infty} x_{t+j}$ to a shock — off the lag polynomial evaluated at $L = 1$,

$$\mathcal{Z}_b(1) = \sum_{j=0}^{\infty} b_j;$$

and we can read the final response, which needs to be zero for a stationary process, from the lag polynomial at $L = \infty$,

$$b_\infty = \lim_{j\to\infty} b_j = \lim_{L\to\infty} \mathcal{Z}_b(L) = \mathcal{Z}_b(\infty).$$

6.2 Continuous Time

In continuous time, the moving average representation is

$$x_t = \int_{\tau=0}^{\infty} b(\tau)\sigma dB_{t-\tau}. \tag{6.2}$$

The quantity $b(\tau)$ again gives an "impulse–response" function, namely how expectations at t about $x_{t+\tau}$ are affected by the shock σdB_t.

The concept $\lim_{\Delta\to 0}(E_{t+\Delta} - E_t)x_t$ does not really make sense. It makes more sense in continuous time to understand the "impulse–response" as the loading of a difference dx_t on the Brownian motion σdB_t term. By transforming the moving-average representation of levels in (6.2) to differences as in (5.8),

$$dx_t = \left(\int_{\tau=0}^{\infty} b'(\tau)\sigma dB_{t-\tau}\right) dt + b(0)\sigma dB_t,$$

we get a better sense of $b(0) = 1$ as the "response of x_t to a shock" — that concept represents how dx_t responds to a Brownian increment σdB_t. In discrete time,

$$x_{t+1} - x_t = b_0\varepsilon_{t+1} + \sum_{j=0}^{\infty}(b_{j+1} - b_j)\varepsilon_{t-j}$$

the innovation in x_{t+1} and Δx_{t+1} (i.e. x_{t+1} and dx_{t+1}) are the same. The difference version makes more sense in continuous time.

Similarly, to see what an "impulse–response" past the first term really means in continuous time, define

$$y_t = E_t(x_{t+k}) = \int_{\tau=0}^{\infty} b(\tau + k)\sigma dB_{t-\tau}.$$

Then, following the same logic as in (5.8),

$$dy_t = b(k)\sigma dB_t + \left(\int_{\tau=0}^{\infty} b'(\tau + k)\sigma dB_{t-\tau}\right) dt.$$

Here you see directly what it means to say that $b(k)$ is the shock to today's expectations of x_{t+k}. (We get the same result whether we interpret dy_t as

$$y_{t+\Delta} - y_t = E_{t+\Delta}(x_{t+k}) - E_t(x_{t+k})$$

or if we interpret dy_t as

$$y_{t+\Delta} - y_t = E_{t+\Delta}(x_{t+k+\Delta}) - E_t(x_{t+k}).$$

These quantities are the same because $E_t(dx_{t+k})$ is of order dt.)

We can recover the impact multiplier from the level operator function (6.2) via

$$b(0) = \lim_{D\to\infty} D\mathcal{L}_b(D). \tag{6.3}$$

This expression is the analogue to (6.1). I am normalizing so that $b(0) = 1$ for moving average representations, and this expression allows us to check that fact for general differential-operator functions.

Statement (6.3) is the "initial value theorem" of Lapalce transforms. To derive this formula, take the limit on both sides of (5.7), which I repeat here,

$$D\mathcal{L}_b(D) = b(0) + \mathcal{L}_{b'}(D),$$

and note that

$$\lim_{D\to\infty} \mathcal{L}_{b'}(D) = \lim_{D\to\infty} \int_{\tau=0}^{\infty} e^{-D\tau} b'(\tau) d\tau = 0.$$

The form of the differential-operator polynomials (4.4) imposes this normalization

$$\lim_{D\to\infty} D\mathcal{L}_b(D) = \lim_{D\to\infty} D\frac{(D+\theta_1)(D+\theta_2)\cdots}{(D+\lambda_1)(D+\lambda_2)(D+\lambda_3)\cdots} = 1,$$

but only if there is one less D on top than on the bottom. This observation gives a little deeper insight for that requirement.

Applying $b(0) = \lim_{D\to\infty} D\mathcal{L}_b(D) = 1$ to the partial-fractions expansion of the differential operator polynomial, (4.6),

$$x_t = \left[\frac{A}{(D+\lambda_1)} + \frac{B}{(D+\lambda_2)} + \frac{C}{(D+\lambda_3)} + \cdots\right]\sigma DB_t \tag{6.4}$$

gives a swift demonstration and interpretation of the fact that $A + B + C + \cdots = 1$.

Since the differenced moving average $\mathcal{L}_c(D) = D\mathcal{L}_b(D)$, the corresponding requirement is

$$\lim_{D\to\infty} \mathcal{L}_c(D) = 1$$

Since the "impact multiplier" is most easily understood in continuous time as the response of dx_t to σdB_t, this requirement makes better sense of the expression (6.3)

The "final value theorem" of Laplace transforms states

$$b(\infty) = \lim_{D\to 0} D\mathcal{L}_b(D). \tag{6.5}$$

As in discrete time, to obtain a stationary (finite-variance) series, moving averages must tail off,

$$\lim_{\tau\to\infty} b(\tau) = 0$$

(Actually we need $\int_{\tau=0}^{\infty} b^2(\tau) < \infty$ which is stronger.) As in discrete time, (6.5) tells us how to measure this quantity directly from the differential operator function $\mathcal{L}_b(D)$.

To see the "final value theorem," simply take the limit of

$$\int_{\tau=0}^{\infty} De^{-D\tau} b(\tau) d\tau.$$

We also want the equivalent of the cumulative response function, which measures the response of $E_t \int_{\tau=0}^{\infty} x_{t+\tau}\, d\tau$ to a shock. Corresponding to $\mathcal{Z}_b(1)$ in discrete time, we have

$$\mathcal{L}_b(0) = \int_{\tau=0}^{\infty} b(\tau) d\tau.$$

We often model the differences

$$dx_t = \mathcal{L}_c(D)\sigma D B_t$$

and want to find the final response of the level x_t to the shock. Since $\lim_{T\to\infty} x_{t+T} = \int_{\tau=0}^{\infty} dx_{t+\tau}$, the final response of x_t is

$$\mathcal{L}_c(0) = 1 + \int_{\tau=0}^{\infty} c(\tau) d\tau.$$

(The right-hand expansion is for the standard case of a δ function at zero with $c(0) = 1$). If x_t is stationary, this number like b_∞ in (6.5) should be zero. If dx_t is stationary but x_t is not, this number is not zero, and is the key distinguishing level and difference stationary series. More later.

(Befitting the nontechnical nature of this article, I am not making an important distinction between $\mathcal{L}_c(0)$ and $\lim_{D\to 0}\mathcal{L}_c(D)$. With $\mathcal{L}_c(D) = D\mathcal{L}_b(D)$ you can see why the latter formulation might be preferred. But we can usually write $\mathcal{L}_c(D)$ in such a way that the limit and limit point are the same. For the AR(1) example, $D\mathcal{L}_b(D) = D/(D+\phi)$, and $\mathcal{L}_c(D) = 1 - \phi/(D+\phi)$. These are the same except at the limit point $D = 0$.)

7

Hansen–Sargent Formulas

Here is one great use of the operator notation — and the application that drove me to figure all this out and write it up. Given a process x_t, how do you calculate

$$E_t \int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau} \, d\tau?$$

This is an operation we run into again and again in modern intertemporal macroeconomics and in asset pricing.

7.1 Discrete Time

Hansen and Sargent (1980) gave an elegant answer to this question in discrete time. You want to calculate $E_t \sum_{j=0}^{\infty} \beta^j x_{t+j}$. You are given a moving average representation $x_t = \mathcal{Z}_b(L)\varepsilon_t$. (Here and below, ε_t can be a vector of shocks, which considerably generalizes the range of processes you can write down.) The answer: the moving-average representation of the expected discounted sum is

$$E_t \sum_{j=0}^{\infty} \beta^j x_{t+j} = \left(\frac{L\mathcal{Z}_b(L) - \beta \mathcal{Z}_b(\beta)}{L - \beta} \right) \varepsilon_t = \left(\frac{\mathcal{Z}_b(L) - \beta L^{-1} \mathcal{Z}_b(\beta)}{1 - \beta L^{-1}} \right) \varepsilon_t. \tag{7.1}$$

Hansen and Sargent give the first form. The second form is a bit less pretty but shows a bit more clearly what you are doing. $\mathcal{Z}_b(L)\varepsilon_t$ is just x_t. $(1-\beta L)^{-1} = \sum_{j=0}^{\infty}\beta^j L^{-j}$ takes the forward sum so $(1-\beta L^{-1})^{-1}\mathcal{Z}_b(L)\varepsilon_t$ is the actual, ex-post value whose expectation we seek. But that expression would leave you many terms in ε_{t+j}. The second term ends up subtracting off all the ε_{t+j} terms leaving only ε_{t-j} terms, which thus is the conditional expectation.

For example, consider an AR(1). We start with

$$x_t = \mathcal{Z}_b(L)\varepsilon_t = (1-\rho L)^{-1}\varepsilon_t.$$

Then the expected discounted sum follows

$$\begin{aligned} E_t\sum_{j=0}^{\infty}\beta^j x_{t+j} &= \left(\frac{\frac{L}{1-\rho L} - \frac{\beta}{1-\rho\beta}}{L-\beta}\right)\varepsilon_t = \frac{1}{(1-\rho\beta)}\frac{1}{(1-\rho L)}\varepsilon_t \\ &= \frac{1}{(1-\rho\beta)}\sum_{j=0}^{\infty}\rho^j\varepsilon_{t-j} = \frac{1}{(1-\rho\beta)}x_t. \end{aligned}$$

The formula is even prettier if we start one period ahead, as often happens in finance:

$$E_t\sum_{j=1}^{\infty}\beta^{j-1}x_{t+j} = \left(\frac{\mathcal{Z}_b(L) - \mathcal{Z}_b(\beta)}{L-\beta}\right)\varepsilon_t \tag{7.2}$$

Just subtract $x_t = \mathcal{Z}_b(L)\varepsilon_t$ from (7.1). This version turns out to look exactly like the continuous-time formula below.

We often want the impact multiplier – how much does a price react to a shock? The Hansen–Sargent formula (7.1) says the answer is $\mathcal{Z}_b(\beta)$, i.e.

$$(E_t - E_{t-1})\sum_{j=0}^{\infty}\beta^j x_{t+j} = \mathcal{Z}_b(\beta)\varepsilon_t. \tag{7.3}$$

This formula is particularly lovely because you do not have to construct, factor, or invert any lag polynomials. Suppose you start with

an autoregressive representation

$$\mathcal{Z}_a(L)x_t = \varepsilon_t.$$

Then, you can first evaluate $\mathcal{Z}_a(\beta)$ (a number) and then invert that number, rather than invert a lag-operator polynomial (hard) and then substitute in a number:

$$(E_t - E_{t-1})\sum_{j=0}^{\infty}\beta^j x_{t+j} = [\mathcal{Z}_a(\beta)]^{-1}\varepsilon_t.$$

7.2 Continuous Time

Hansen and Sargent (1991) show that if we express a process in moving-average form,

$$x_t = \int_{\tau=0}^{\infty} b(\tau)\sigma dB_{t-\tau} = \mathcal{L}_b(D)\sigma DB_t,$$

then we can find the moving average representation of the expected discounted value by

$$E_t\int_{\tau=0}^{\infty} e^{-r\tau}x_{t+\tau}\ d\tau = \left(\frac{\mathcal{L}_b(D) - \mathcal{L}_b(r)}{r - D}\right)\sigma DB_t. \tag{7.4}$$

The formula is almost exactly the same as (7.2).

The pieces work as in discrete time. The operator

$$\frac{1}{r - D} = \int_{\tau=0}^{\infty} e^{-r\tau}e^{D\tau}d\tau$$

takes the discounted forward integral, and creates the ex-post present value. Subtracting off $\mathcal{L}_b(r)/(r - D)$ removes all the terms by which the discounted sum depends on future realizations of $\sigma dB_{t+\tau}$, leaving an expression that only depends on the past and hence is the conditional expectation.

Here is the AR(1) example in continuous time. x_t follows

$$x_t = \frac{1}{D + \phi}\sigma DB_t.$$

Applying (7.4),

$$\begin{aligned} E_t \int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau}\, d\tau &= \frac{1}{(r-D)}\left(\frac{1}{D+\phi} - \frac{1}{r+\phi}\right)\sigma D B_t \\ &= \frac{1}{(r+\phi)(D+\phi)}\sigma D B_t \\ &= \frac{1}{r+\phi}\int_{\tau=0}^{\infty} e^{-\phi\tau}\sigma dB_{t-\tau} = \frac{1}{r+\phi}x_t. \end{aligned}$$

We recover the same result as in discrete time.

The innovation in the expected discounted value, the counterpart to (7.3), is found as we found impact multipliers in (6.3). From (7.4), the impact multiplier of the expected discounted value is

$$\lim_{D\to\infty}\left(D\frac{\mathcal{L}_b(D)-\mathcal{L}_b(r)}{r-D}\right) = \mathcal{L}_b(r). \tag{7.5}$$

($\lim_{D\to\infty} D\mathcal{L}_b(D) = b(0) = 1$ is the impact multiplier of x_t, so, dividing by $r - D$, the first numerator term is zero.) Thus, if we define

$$y_t = E_t \int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau}\, d\tau,$$

then

$$dy_t = ()dt + \mathcal{L}_b(r) dB_t.$$

This expression reminds us what an impact multiplier means in continuous time. As in discrete time, (7.5) is a lovely formula because you may be able to find $\mathcal{L}_b(r)$ without knowing the whole $\mathcal{L}_b(D)$ function. (As an example, I use this formula in Cochrane (2012) (below Equation (4.2), p. 178) to evaluate how much consumption must react to an endowment shock, in order to satisfy the present-value budget constraint in a permanent-income style model with complex habits and durability. In this case, the habits or durability add "autoregressive" terms, and it is convenient to invert them as scalar $\mathcal{L}(r)$ rather than functions $\mathcal{L}(D)$.)

7.3 Derivation

7.3.1 Operator Derivation

Hansen and Sargent give an elegant derivation that illustrates the power of thinking in terms of Laplace transforms. Start with the ex-post present value. It has a moving average representation, whose terms I will denote by $d(\tau)$. Then, we want to separate $d(\tau)$ into its positive (past) and negative (future) components. Write

$$\begin{aligned}\int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau}\, d\tau &= \int_{\tau=-\infty}^{\infty} d(\tau) dB_{t-\tau} \\ &= \int_{\tau=-\infty}^{0} d(\tau) dB_{t-\tau} + \int_{\tau=0}^{\infty} d(\tau) dB_{t-\tau} \frac{\mathcal{L}_b(D)}{r-D} \sigma D B_t \\ &= \mathcal{L}_d(D) \sigma D B_t \\ &= [\mathcal{L}_{d^-}(D) + \mathcal{L}_{d^+}(D)] \sigma D B_t\end{aligned}$$

The second integral runs from $-\infty$ to ∞, because the ex-post present value depends on future shocks. The differential-operator function $\mathcal{L}_d(D)$ has a pole at $D = r$, so must be in part solved forward.

In order to break $\mathcal{L}_d(D)$ into past and future components, Hansen and Sargent suggest that we simply add and subtract $\mathcal{L}_b(r)$

$$\frac{\mathcal{L}_b(D)}{r-D} \sigma D B_t = \left\{ \left[\frac{\mathcal{L}_b(D) - \mathcal{L}_b(r)}{r-D} \right] + \left[\frac{\mathcal{L}_b(r)}{r-D} \right] \right\} \sigma D B_t$$

The first term no longer has a pole at $D = r$, and removing that pole is a motivation for subtracting $\mathcal{L}_b(r)$. Thus, the first term corresponds to past $dB_{t-\tau}$ only. The numerator of the second term is a constant, so that term has only a pole at $D = r$, and no poles with negative values of D. Thus it is expressed in terms of future $dB_{t-\tau}$ only.

We have achieved what we are looking for! We broke the moving average of the ex-post present value into one term that depends only on past dB_t and one that depends only on future dB_t. The part loading only on the past, the first term after the equality, must be the conditional expectation.

Wait a minute, you say. We could have added and subtracted anything. But the answer is no, this separation is *unique*: if you find *any*

way of adding and subtracting something that breaks $\mathcal{L}_d(D)$ into past and future components, you have found the *only* way of doing so. Suppose we add and subtract an arbitrary $\mathcal{L}(D)$. It must have $\mathcal{L}(r) = \mathcal{L}_b(r)$ so the numerator of the first term removes the pole at $D = r$. Still, any backwards-solvable $\mathcal{L}(D)$ with $\mathcal{L}(r) = \mathcal{L}_b(r)$ would work in the first term. But any other backwards-solvable $\mathcal{L}(D)$ would induce backwards-solvable parts of the *second* term. A constant is the only thing we can add and subtract which removes the pole in the first term, making that term backwards-solvable, but does not introduce backwards-solvable parts in the second term. And that constant must be $\mathcal{L}_b(r)$ to remove the pole in the first term.

7.3.2 Brute force

It is easy to check the Hansen–Sargent formula by brute force. It is useful to confirm that the operator logic is correct. Write out the moving average representation for the ex-post present value, $\sum_{j=0}^{\infty} \beta^j x_{t+j}$, then verify that the $\mathcal{Z}_b(\beta)/(L-\beta)$ term subtracts off the forward-looking terms. The ex-post present value is

$$\left(\frac{\mathcal{Z}_b(L)}{1-\beta L^{-1}}\right)\varepsilon_t = \sum_{j=0}^{\infty} \beta^j x_{t+j} \tag{7.6}$$

$$\begin{array}{rccccc} & & +b_0\varepsilon_t & +b_1\varepsilon_{t-1} & +b_2\varepsilon_{t-2} & \cdots \\ & +\beta b_0\varepsilon_{t+1} & +\beta b_1\varepsilon_t & +\beta b_2\varepsilon_{t-1} & +\beta b_3\varepsilon_{t-2} & \cdots \\ = \quad +\beta^2 b_0\varepsilon_{t+2} & +\beta^2 b_1\varepsilon_{t+1} & +\beta^2 b_2\varepsilon_t & +\beta^2 b_3\varepsilon_{t-1} & +\beta^2 b_4\varepsilon_{t-3} & \cdots \\ +\beta^3 b_0\varepsilon_{t+3} & +\beta^3 b_1\varepsilon_{t+2} & +\beta^3 b_2\varepsilon_{t+1} & \cdots & & \\ & & \cdots & & & \end{array}$$

Summing the columns,

$$\begin{aligned} &= \cdots + \beta^3\mathcal{Z}_b(\beta)\varepsilon_{t+3} + \beta^2\mathcal{Z}_b(\beta)\varepsilon_{t+2} + \beta\mathcal{Z}_b(\beta)\varepsilon_{t+1} + \mathcal{Z}_b(\beta)\varepsilon_t \\ &\quad + (..)\varepsilon_{t-1} + (..)\varepsilon_{t-2} + \cdots \end{aligned} \tag{7.7}$$

The second part of the formula (7.1) gives

$$\begin{aligned} \frac{\beta L^{-1}}{1-\beta L^{-1}}\mathcal{Z}_b(\beta)\varepsilon_t &= (\beta L^{-1} + \beta^2 L^{-2} + \beta^3 L^{-3} + \cdots)\mathcal{Z}_b(\beta)\varepsilon_t \\ &= \cdots + \beta^3\mathcal{Z}_b(\beta)\varepsilon_{t+3} + \beta^2\mathcal{Z}_b(\beta)\varepsilon_{t+2} + \beta\mathcal{Z}_b(\beta)\varepsilon_{t+1} \end{aligned}$$

You can see that these are exactly the forward-looking terms in (7.7). By subtracting these terms, we neatly subtract off all the forward terms $\varepsilon_{t+1}, \varepsilon_{t+2}$, etc. from the ex-post present value and find the expected present value.

You can check the continuous-time Hansen–Sargent formula in the same way. Express the ex-post forward looking present value $\int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau}\, d\tau$ in moving average representation, collect all the $dB_{t-\tau}$ terms in one place for each τ, then notice that the second half of the Hansen–Sargent formula neatly eliminates all the $dB_{t+\tau}$ terms. Start with

$$\begin{aligned}\frac{\mathcal{L}_b(D)}{r-D}\sigma D B_t &= \int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau}\, d\tau \\ &= \int_{\tau=0}^{\infty} e^{-r\tau}\left(\int_{s=0}^{\infty} b(s)\sigma dB_{t+\tau-s}\right) d\tau\end{aligned}$$

We transform to an integral over $q = \tau - s$ that counts each dB_q once, and separate past dB_q from future dB_q. To find the limits of the definite integrals, when $q < 0$ (past), then $\tau \geq 0$ means $s \geq -q$. When $q > 0$ (future), then s starts at 0.

$$\begin{aligned}&\int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau}\, d\tau \\ &\quad= \int_{q=-\infty}^{\infty}\int_{s=\max(0,-q)}^{\infty} e^{-rq}e^{-rs}b(s)\sigma dB_{t+q}\, ds \\ &\quad= \int_{q=0}^{\infty}\int_{s=0}^{\infty} e^{-rq}e^{-rs}b(s)\sigma dB_{t+q} ds \\ &\qquad+ \int_{q=-\infty}^{0}\int_{s=-q}^{\infty} e^{-rq}e^{-rs}b(s)\sigma dB_{t+q} ds \\ &\quad= \int_{q=0}^{\infty} e^{-rq}\left(\int_{s=0}^{\infty} e^{-rs}b(s)ds\right)\sigma dB_{t+q} \\ &\qquad+ \int_{q=-\infty}^{0} e^{-rq}\left(\int_{\tau=0}^{\infty} e^{rq}e^{-r\tau}b(\tau-q)d\tau\right)\sigma dB_{t+q}\end{aligned}$$

$$= \left(\int_{s=0}^{\infty} e^{-rs} b(s) ds\right) \int_{q=0}^{\infty} e^{-rq} \sigma dB_{t+q}$$
$$+ \int_{q=-\infty}^{0} \left(\int_{s=0}^{\infty} e^{-r\tau} b(\tau - q) ds\right) \sigma dB_{t+q}.$$

To take expectations, we just drop the first term, so the second term is the expected value we are looking for. Translating the first two terms to operator notation, we have

$$\frac{\mathcal{L}_b(D)}{r - D} \sigma DB_t = \frac{\mathcal{L}_b(r)}{r - D} \sigma DB_t + E_t \left(\int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau} d\tau\right).$$

8

Integration and Cointegration

So far, I have assumed that the series x_t is stationary in levels. We study differences dx_t because that is more convenient in continuous time. Here I take up the possibility that x_t contains unit roots; that dx_t is stationary but x_t is not. I describe the transformation from differences to levels, and the unit root and cointegrated representations of difference-stationary series.

8.1 Difference-stationary Series

So far, we have been looking at differenced specifications simply because the differential operator is more convenient in continuos time, though the level of the series is stationary, with the AR(1) $dx_t = -\phi x_t + \sigma dB_t$ as the canonical example. Often, we will model series whose differences are stationary, but the levels are not, such as dB_t itself. Hence it is worth writing down what specifications based purely on differences look like.

The moving average is

$$\begin{aligned} Dx_t &= \mathcal{L}_c(D)\sigma DB_t \\ dx_t &= \int_{\tau=0}^{\infty} c(\tau)\sigma dB_{t-\tau} + \sigma dB_t \end{aligned}$$

As before, I assume that $c(\tau)$ has a δ function at $c(0) = 1$ to generate the Laplace transform $\mathcal{L}_c(D)$. Reiterating, we normalize so a unit shock σdB_t has a unit effect on dx_t,

$$\lim_{D\to\infty} \mathcal{L}_c(D) = 1.$$

A corresponding "autoregressive" representation is

$$\mathcal{L}_c(D)^{-1} Dx_t = \sigma DB_t$$

We make sense of these expressions with the usual manipulations. For example, a first-order polynomial model is

$$Dx_t = \frac{D+\theta}{D+\lambda}\sigma DB_t.$$

Its moving-average representation can be written as

$$\begin{aligned} Dx_t &= \left(1 + \frac{\theta-\lambda}{D+\lambda}\right)\sigma DB_t \\ dx_t &= (\theta-\lambda)\left(\int_{\tau=0}^{\infty} e^{-\lambda\tau}\sigma dB_{t-\tau}\right)dt + \sigma dB_t. \end{aligned}$$

The autoregressive representation is

$$\begin{aligned} \frac{D+\lambda}{D+\theta} Dx_t &= \sigma DB_t \\ \left(1 + \frac{\lambda-\theta}{D+\theta}\right) Dx_t &= \sigma DB_t \\ dx_t &= -(\lambda-\theta)\left(\int_{\tau=0}^{\infty} e^{-\theta\tau}\, dx_{t-\tau}\right)dt + \sigma dB_t. \end{aligned}$$

Here you see that we forecast future changes using past *changes* $dx_{t-\tau}$, as we normally would run an autoregression in first differences for series like stock returns or GDP growth.

8.2 Differences to Levels in Discrete Time; Beveridge and Nelson

Above, we studied the transition from levels to differences. Next, we study the converse operation. We want to get from

$$(1-L)x_t = \mathcal{Z}_c(L)\varepsilon_t \tag{8.1}$$

to something like

$$x_t = \mathcal{Z}_b(L)\varepsilon_t.$$

Lag operator notation suggests that we construct $\mathcal{Z}_b(L)$ as

$$\mathcal{Z}_b(L) = \frac{\mathcal{Z}_c(L)}{1-L} = c_0 + (c_0 + c_1)L + (c_0 + c_1 + c_2)L^2 + \cdots \tag{8.2}$$

However, this operation only produces a stationary process if $\sum_{j=0}^{\infty} c_j = \mathcal{Z}_c(1) = 0$. That condition need not hold. In general, a process (8.1) is not stationary in levels.

We can handle this situation by defining an initial value x_0 and a process $\varepsilon_t = 0$ for all $t \leq 0$. Now

$$x_t - x_0 = \mathcal{Z}_b(L)\varepsilon_t = (1-L)^{-1}\mathcal{Z}_c(L)\varepsilon_t$$

is finite, though nonstationary.

A more convenient way to handle this possibility is to decompose x_t in to stationary and random walk components via the Beveridge and Nelson (1981) decomposition. We rearrange the terms of $\mathcal{Z}_c(L)$ as

$$(1-L)x_t = \mathcal{Z}_c(L)\varepsilon_t = [\mathcal{Z}_c(1) + (1-L)\mathcal{Z}_b(L)]\varepsilon_t \tag{8.3}$$

where

$$\mathcal{Z}_b(L) = \sum_{j=0}^{\infty} b_j L^j \text{ with } b_j = -\sum_{k=j+1}^{\infty} c_k. \tag{8.4}$$

From (8.3) we can write x_t as the sum of two components,

$$x_t = z_t + w_t,$$

where

$$z_t = z_{t-1} + \mathcal{Z}_c(1)\varepsilon_t$$
$$w_t = \mathcal{Z}_b(L)\varepsilon_t.$$

Now, if $\mathcal{Z}_c(1) = 0$, then we have $x_t = w_t = \mathcal{Z}_b(L)\varepsilon_t$, the representation in levels we are looking for, and x_t is stationary. If $\mathcal{Z}_c(1) \neq 0$, we have the next best thing; we express x_t as an interesting combination of a stationary series w_t plus a pure random walk z_t component.

To verify the Beveridge–Nelson decomposition by brute force, just write out $\mathcal{Z}_b(L)$ as defined by (8.4):

$$\begin{aligned}\mathcal{Z}_b(L) = &-(c_1 + c_2 + c_3 + \cdots) - (c_2 + c_3 + c_4 + \cdots)L \\ &-(c_3 + c_4 + c_5 + \cdots)L^2 - \cdots\end{aligned}$$

then note

$$\begin{aligned}(1-L)\mathcal{Z}_b(L) &= -c_0 - (c_1 + c_2 + c_3 + \cdots) + c_0 + c_1 L + c_2 L^2 + \cdots \\ &= -\mathcal{Z}_c(1) + \mathcal{Z}_c(L).\end{aligned}$$

Since the $\{c_j\}$ are square summable, so are the $\{b_j\}$. This is a key observation, $\mathcal{Z}_b(L)\varepsilon_t$ defines a level-stationary process.

In operator notation, the decomposition (8.3) consists of just adding and subtracting $\mathcal{Z}_c(1)$:

$$(1-L)x_t = \mathcal{Z}_c(L)\varepsilon_t = \mathcal{Z}_c(1)\varepsilon_t + (1-L)\left[\frac{\mathcal{Z}_c(L) - \mathcal{Z}_c(1)}{1-L}\right]\varepsilon_t. \quad (8.5)$$

Then, we define $\mathcal{Z}_b(L)$ by

$$\mathcal{Z}_b(L) = \frac{\mathcal{Z}_c(L) - \mathcal{Z}_c(1)}{1-L}$$

to arrive at (8.3). This looks too easy — could you add and subtract anything, and multiply and divide by $(1-L)$? But the fact that makes it work is that $\mathcal{Z}_b(L) = [\mathcal{Z}_c(L) - \mathcal{Z}_c(1)]/(1-L)$ is a legitimate lag polynomial of a stationary process. All its poles lie outside the unit circle. (Following usual practice, I do not normalize so $\mathcal{Z}_b(0) = 1$ in this case.)

The Beveridge–Nelson trend z_t has the property

$$z_t = \lim_{j\to\infty} E_t(x_{t+j}), \quad (8.6)$$

which follows simply from the fact that w_t is stationary so $\lim_{j\to\infty} E_t w_{t+j} = 0$. This can also be used as the defining property to derive the Beveridge–Nelson decomposition, which is a longer but more satisfying since you construct the answer rather than verify it. Thinking in this way, we can derive the Beveridge–Nelson decomposition as

a case of the Hansen–Sargent formula (7.2) evaluated at $\beta = 1$:

$$\begin{aligned} z_t &= \lim_{j\to\infty} E_t(x_{t+j}) = x_t + E_t \sum_{j=1}^{\infty} \Delta x_{t+j} \\ &= x_t + \left(\frac{\mathcal{Z}_c(1) - \mathcal{Z}_c(L)}{1-L}\right)\varepsilon_t, \\ (1-L)z_t &= (1-L)\mathcal{Z}_c(L)\varepsilon_t + (1-L)\left(\frac{\mathcal{Z}_c(1) - \mathcal{Z}_c(L)}{1-L}\right)\varepsilon_t \\ &= \mathcal{Z}_c(1)\varepsilon_t \end{aligned}$$

Defining w_t as detrended x_t,

$$\begin{aligned} w_t &= x_t - z_t \\ (1-L)w_t &= (1-L)x_t + (1-L)z_t \\ (1-L)w_t &= [\mathcal{Z}_c(L) + \mathcal{Z}_c(1)]\varepsilon_t. \end{aligned}$$

8.3 Differences to Levels in Continuous Time

The same operations have natural analogues in continuous time. Before, we found the differenced moving average representation of a level-stationary series, in (5.7). Now we want to ask the converse question. Suppose you have a differential representation,

$$Dx_t = \mathcal{L}_c(D)\sigma DB_t.$$

How do you find $\mathcal{L}_b(D)$ or $b(\tau)$ in

$$x_t = \mathcal{L}_b(D)\sigma DB_t? \tag{8.7}$$

The fly in the ointment, as in discrete time, is that the process x_t may not be stationary in levels, so the latter integral does not make sense. As a basic example, if you start with simple Brownian motion

$$dx_t = \sigma dB_t,$$

you can not invert that to

$$x_t = \sigma B_t = \int_{\tau=0}^{\infty} \sigma dB_{t-\tau},$$

because the latter integral blows up. For this reason, we usually express the level of pure Brownian motion as an integral that only looks back to an initial level,

$$x_t = x_0 + \int_{\tau=0}^{t} \sigma dB_{t-\tau} = x_0 + \sigma(B_t - B_0).$$

As in this example, we can ignore the nonstationarity, use (8.7) directly, and think of a nonstationary process that starts at time 0 with $dB_t = 0$ for all $t < 0$. (Hansen and Sargent (1983), last paragraph.)

Alternatively, we can handle this situation as in discrete time, with the continuous-time Beveridge–Nelson decomposition that isolates the nonstationarity to a pure random walk component. We rearrange the terms of $\mathcal{L}_c(D)$,

$$Dx_t = \mathcal{L}_c(D)\sigma DB_t = [\mathcal{L}_c(0) + D\mathcal{L}_b(D)]\sigma DB_t. \tag{8.8}$$

I will show in a moment how to construct $\mathcal{L}_b(D)$, and verify that it is the differential-operator function of a valid stationary process. Once that is done, though, we can write this last equation as

$$Dx_t = Dz_t + Dw_t$$

and hence

$$x_t = z_t + w_t$$

where z is a pure random walk

$$Dz_t = \mathcal{L}_c(0)\sigma DB_t$$

and w_t is stationary in levels,

$$w_t = \mathcal{L}_b(D)\sigma DB_t.$$

Now, if $\mathcal{L}_c(0) = 0$ we have $x_t = w_t$ stationary. If $\mathcal{L}_c(0) \neq 0$, then we isolate the nonstationarity to a pure random walk component z_t and put all the dynamics in a level-stationary stochastically detrended component w_t.

Now, how do we construct $\mathcal{L}_b(D)$ given $\mathcal{L}_c(D)$? The operator derivation is nearly trivial. By construction,

$$\mathcal{L}_c(D) = \mathcal{L}_c(0) + D\left[\frac{\mathcal{L}_c(D) - \mathcal{L}_c(0)}{D}\right].$$

Therefore, we just define

$$\mathcal{L}_b(D) = \frac{\mathcal{L}_c(D) - \mathcal{L}_c(0)}{D}. \tag{8.9}$$

Adding and subtracting $\mathcal{L}_c(0)$ and multiplying and dividing by D looks artificial, but the key is that $(\mathcal{L}_c(D) - \mathcal{L}_c(0))/D$ is a valid level-stationary process, since $-\mathcal{L}_c(0)$ removes the pole at 0. Equivalently, it produces a new difference operator function $\mathcal{L}_{c^*}(D) = \mathcal{L}_c(D) - \mathcal{L}_c(0)$, which does have the property $\mathcal{L}_{c^*}(0) = 0$ and hence $\mathcal{L}_b(D) = \mathcal{L}_{c^*}(D)/D$ is a proper level-stationary process.

We can construct the terms $b(\tau)$ by integrating $c(\tau)$

$$b(\tau) = -\int_{s=\tau}^{\infty} c(s)ds.$$

This is the obvious inverse to our construction of terms $c(\tau)$ by differentiating $b(\tau)$ in (5.7), and it mirrors the discrete-time formula (8.4). To see where this expression comes from, let us write

$$\mathcal{L}_c(D) = c(0) + \int_{\tau=0}^{\infty} e^{-D\tau} c(\tau)d\tau.$$

Then,

$$\begin{aligned}
\mathcal{L}_b(D) &= \frac{\mathcal{L}_c(D) - \mathcal{L}_c(0)}{D} \\
&= \frac{c(0) + \int_{s=0}^{\infty} e^{-Ds} c(s)ds - \left[c(0) + \int_{s=0}^{\infty} c(s)ds\right]}{D} \\
&= \frac{\int_{s=0}^{\infty} [e^{-Ds} - 1]c(s)ds}{D} = -\int_{s=0}^{\infty} \left[\int_{\tau=0}^{s} e^{-D\tau} d\tau\right] c(s)ds \\
&= -\int_{\tau=0}^{\infty} e^{-D\tau} \left[\int_{s=\tau}^{\infty} c(s)ds\right] d\tau.
\end{aligned} \tag{8.10}$$

In sum, as we used the identity (5.7)

$$\mathcal{L}_c(D) = D\mathcal{L}_b(D) = b(0) + \mathcal{L}_{b'}(D)$$

to construct $\mathcal{L}_c(D)$ from a given $\mathcal{L}_b(D)$, here we use the identity

$$\mathcal{L}_b(D) = \frac{\mathcal{L}_c(D) - \mathcal{L}_c(0)}{D} = \mathcal{L}_{\int c}(D)$$

where I use the notation $\mathcal{L}_{\int c}(D)$ to refer to the transform in (8.10)

The random walk component z_t has the property

$$z_t = \lim_{T\to\infty} E_t(x_{t+T}),$$

and this property can be used to derive the decomposition. Doing so as a case of the Hansen–Sargent prediction formula (7.4) with $r = 0$ provides more intuition for the operator definition (8.9). We write

$$z_t = x_t + E_t\left(\int_{s=0}^{\infty} dx_{t+s}\right) = x_t + \left(\frac{\mathcal{L}_c(D) - \mathcal{L}_c(0)}{-D}\right)\sigma DB_t$$

Therefore,

$$\begin{aligned} Dz_t &= Dx_t - [\mathcal{L}_c(D) - \mathcal{L}_c(0)]\,\sigma DB_t \\ &= \{\mathcal{L}_c(D) - [\mathcal{L}_c(D) - \mathcal{L}_c(0)]\}\,\sigma DB_t \\ &= \mathcal{L}_c(0)\sigma DB_t. \end{aligned}$$

Defining $w_t = x_t - z_t$ we recover the decomposition.

8.4 Cointegration

Cointegration is really a vector generalization of the differences-to-levels issues. Here, I translate the basic representation theorems, such as Engle and Granger (1987). Let x_t now denote a vector of N time series, and dB_t a vector of N independent standard Brownian motions. The moving average representations such as

$$\begin{aligned} x_t &= \int_{\tau=0}^{\infty} b(\tau)\,VdB_{t-\tau} \\ dx_t &= \left(\int_{\tau=0}^{\infty} c(\tau)\,VdB_{t-\tau}\right)dt + VdB_t \end{aligned}$$

or in operator notation

$$\begin{aligned} x_t &= \mathcal{L}_b(D)\,VDB_t \\ Dx_t &= \mathcal{L}_c(D)\,VDB_t \end{aligned} \tag{8.11}$$

represent matrix operations with $\mathcal{L}(D)$ and V denoting $N \times N$ matrices.

For a scalar, the level/difference issue is whether $\mathcal{L}_c(0) = 0$ or not. For a vector, we have the additional possibility that $\mathcal{L}_c(0)$ may be nonzero but not full rank. In that case, the elements of x_t are cointegrated. To keep the discussion simple I will mostly consider the case $N = 2$, and $x_t = [x_{1t}\ x_{2t}]'$.

Cointegrated series have a *common trend representation*: if $\mathcal{L}_c(0)$ has rank 1, then we can write

$$\mathcal{L}_c(0) = \delta\beta' \tag{8.12}$$

where δ and β are 2×1 vectors. Building on the Beveridge–Nelson decomposition (8.8), define a scalar random walk component z_t,

$$Dz_t = \beta' V\ DB_t,$$

and we can write the two components of x_t as a sum of this shared random walk component and stationary components,

$$x_t = \delta z_t + w_t,$$

i.e.

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} z_t + \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix}. \tag{8.13}$$

To derive this representation in operator notation, we proceed exactly as we did in deriving the Beveridge–Nelson decomposition, interpreting the symbols as matrices, and introducing $\mathcal{L}_c(0) = \delta\beta'$ at the right time. From (8.11),

$$\begin{aligned} Dx_t &= \mathcal{L}_c(0)V\ DB_t + D\left[\frac{\mathcal{L}_c(D) - \mathcal{L}_c(0)}{D}\right] V\ DB_t \\ Dx_t &= \delta\beta' V\ DB_t + D\mathcal{L}_b(D)V\ DB_t \\ Dx_t &= \delta Dz_t + D\mathcal{L}_b(D)V\ DB_t \end{aligned}$$

or, in levels

$$x_t = \delta z_t + \mathcal{L}_b(D)DB_t.$$

The *cointegrating vector* gives the linear combination of x_t that is stationary in levels, though the individual components of x_t are not.

Since $\mathcal{L}_c(0) = \delta\beta'$ is singular by assumption, we can find α such that $\alpha'\delta = 0$, and $\alpha'\mathcal{L}_c(0) = 0$, and we can find a ϕ such that $\beta'\phi = 0$ and $\mathcal{L}_c(0)\phi = 0$. Then, from (8.13),

$$\alpha' x_t = \alpha' w_t,$$

i.e., $\alpha' x_t$ is stationary. To get there directly, we can just write

$$\begin{aligned} Dx_t &= \delta\beta' V\ DB_t + D\mathcal{L}_b(D)V\ DB_t \\ \alpha' Dx_t &= (\alpha'\delta)\beta' V\ DB_t + \alpha' D\mathcal{L}_b(D)V\ DB_t \\ \alpha' Dx_t &= \alpha' D\mathcal{L}_b(D)V\ DB_t \\ \alpha' x_t &= \alpha'\mathcal{L}_b(D)V\ DB_t. \end{aligned}$$

The *error-correction representation* is also very useful. For example, forecasting regressions of stock returns and dividend growth on dividend yields, or consumption and income growth on the consumption/income ratio are good examples of useful error-correction representations.

A useful form of the error correction representation is

$$dx_t = -\phi(\alpha' x_t)dt + \left[\int_{\tau=0}^{\infty} e^{-D\tau} d(\tau) V\ dB_{t-\tau}\right] dt + V\ dB_t.$$

Here ϕ is a 2×1 vector which shows how the lagged cointegrating vector affects changes in each of the two differences. I allow extra stationary components in the middle term, expressed as moving averages or "serially correlated errors" in discrete-time parlance. We could also follow the discrete-time VAR literature and write these as lags of dx_t which help to forecast dx_t. The cointegrated AR(1) is a useful special case, in which the middle term is missing. Finally, we have the shock term.

In operator notation, this error correction representation is

$$Dx_t = -\phi(\alpha' x_t) + \mathcal{L}_d(D)V\ DB_t. \tag{8.14}$$

The cointegrated AR(1) is the special case $\mathcal{L}_d(D) = I$.

Applying α' to both sides, the cointegrating vector itself follows

$$D(\alpha' x_t) = -(\alpha'\phi)(\alpha' x_t) + \alpha'\mathcal{L}_b(D)V\ DB_t.$$

Note $\alpha'\phi$ is a scalar (in general a full-rank matrix). Therefore, the scalar process $\alpha' x_t$ is stationary in levels, and has the moving-average representation

$$(\alpha' x_t) = \frac{1}{D + \alpha'\phi}\alpha'\mathcal{L}_b(D)V \ DB_t. \tag{8.15}$$

For the cointegrated AR(1) special case, this is just a scalar AR(1).

Now, let us connect the error correction representation to the above moving-average characterizations. We can substitute (8.15) back in to (8.14) to obtain the moving average differential operator $\mathcal{L}_c(D)$,

$$Dx_t = \left(I - \frac{\phi\alpha'}{D + \alpha'\phi}\right)\mathcal{L}_b(D)V \ DB_t = \mathcal{L}_c(D)V \ dB_t.$$

Since $\mathcal{L}_b(0) = I$, this moving average operator obeys

$$\mathcal{L}_c(0) = I - \phi(\alpha'\phi)^{-1}\alpha'.$$

This is a rank 1 idempotent matrix, confirming the condition (8.12) that defines cointegration, and generalizing the usual special cases $\mathcal{L}_c(0) = 0$ (stationary in levels) and $\mathcal{L}_c(0) = I$ (stationary in differences.) Furthermore,

$$\begin{aligned}\alpha'\mathcal{L}_c(0) &= \alpha'(I - \phi(\alpha'\phi)^{-1}\alpha') = 0\\ \mathcal{L}_c(0)\phi &= (I - \phi(\alpha'\phi)^{-1}\alpha')\phi = 0\end{aligned}$$

so the cointegrating vector α defined by the error-correction mechanism is the same as that which results from the condition $\alpha'\mathcal{L}_c(0) = 0$.

9

Summary

- Basic operators.

$$\begin{aligned} L^{\tau} x_t &= x_{t-\tau} \\ Dx_t &= \frac{1}{dt} dx_t \\ L &= e^{-D};\ D = -\log(L). \end{aligned}$$

- Lag operators, differential operators, Laplace transforms, moving average representation.

$$\begin{aligned} x_t &= \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \mathcal{Z}_b(L)\varepsilon_t;\ \mathcal{Z}_b(L) = \sum_{j=0}^{\infty} b_j L^j;\ b_0 = 1 \\ x_t &= \int_{\tau=0}^{\infty} b(\tau)\sigma dB_{t-\tau} = \mathcal{L}_b(D)\sigma DB_t; \\ \mathcal{L}_b(D) &= \int_{\tau=0}^{\infty} e^{-D\tau} b(\tau) d\tau;\ b(0) = 1. \end{aligned}$$

- The AR(1).

$$x_{t+1} = \rho x_t + \varepsilon_t \Rightarrow x_t = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}$$

$$dx_t = -\phi x_t \, dt + \sigma dB_t \Rightarrow x_t = \int_{\tau=0}^{\infty} e^{-\phi\tau} \, dB_{t-\tau}.$$

- Operators and inverting the AR(1).

$$(1 - \rho L)x_t = \varepsilon_t \Rightarrow$$

$$x_t = \frac{1}{1 - \rho L}\varepsilon_t = \left(\sum_{j=0}^{\infty} \rho^j L^j \right) \varepsilon_t$$

$$(D + \phi)x_t = DB_t \Rightarrow$$

$$x_t = \frac{1}{D + \phi} DB_t = \left(\int_{\tau=0}^{\infty} e^{-\phi\tau} e^{-D\tau} \, d\tau \right) \frac{1}{dt} \, dB_t.$$

- Forward-looking operators.

$$\begin{aligned} \|\rho\| > 1 \Rightarrow \left(\frac{1}{1 - \rho L} \right) \varepsilon_t &= -\left(\frac{\rho^{-1} L^{-1}}{1 - \rho^{-1} L^{-1}} \right) \varepsilon_t \\ &= -\left(\sum_{j=1}^{\infty} \rho^{-j} L^{-j} \right) \varepsilon_t \\ &= -\sum_{j=1}^{\infty} \rho^{-j} \varepsilon_{t+j} \end{aligned}$$

$$\begin{aligned} \|\phi\| > 0 \Rightarrow \frac{1}{D - \phi} \sigma DB_t &= -\left(\int_{\tau=0}^{\infty} e^{-\phi\tau} e^{+D\tau} \, d\tau \right) \sigma DB_t \\ &= -\int_{\tau=0}^{\infty} e^{-\phi\tau} \sigma dB_{t+\tau}. \end{aligned}$$

- Moving averages and moments.

$$\sigma^2(x_t) = \int_{\tau=0}^{\infty} b^2(\tau)\sigma^2 \, d\tau,$$

$$\operatorname{cov}(x_t, x_{t-k}) = \int_{\tau=0}^{\infty} b(\tau)b(t+k)\sigma^2 \, d\tau$$

$$
\begin{aligned}
S_x(\omega) &= \int_{\tau=-\infty}^{\infty} \operatorname{cov}(x_t x_{t-\tau}) d\tau = \mathcal{L}_b(i\omega)\mathcal{L}_b(-i\omega)\sigma^2 \\
\operatorname{cov}(x_t, x_{t-k}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega k} S_x(\omega) d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega k}\mathcal{L}_b(i\omega)\mathcal{L}_b(-i\omega)\sigma^2 d\omega.
\end{aligned}
$$

- Polynomial models and autoregressive representations.

$$x_t = \frac{(D+\theta_1)(D+\theta_2)\cdots}{(D+\lambda_1)(D+\lambda_2)(D+\lambda_3)\cdots}\sigma DB_t.$$

Moving average in partial fractions form

$$x_t = \left[\frac{A}{D+\lambda_1} + \frac{B}{D+\lambda_2} + \frac{C}{D+\lambda_3} + \cdots\right]\sigma DB_t.$$

Autoregressive form

$$\left[D + A + \frac{B}{D+\theta_1} + \frac{C}{D+\theta_2} + \cdots\right] x_t = \sigma DB_t.$$

- "AR(2)."

$$x_t = \frac{(D+\theta_1)}{(D+\lambda_1)(D+\lambda_2)}\sigma DB_t.$$

Moving average

$$
\begin{aligned}
x_t &= \frac{1}{\lambda_1-\lambda_2}\left(\frac{\lambda_1-\theta_1}{D+\lambda_1} - \frac{\lambda_2-\theta_1}{D+\lambda_2}\right)\sigma DB_t \\
&= \frac{\lambda_1-\theta_1}{\lambda_1-\lambda_2}\int_{\tau=0}^{\infty} e^{-\lambda_1\tau}\sigma dB_{t-\tau} + \frac{\lambda_2-\theta_1}{\lambda_2-\lambda_1}\int_{\tau=0}^{\infty} e^{-\lambda_2\tau}\sigma dB_{t-\tau}.
\end{aligned}
$$

Autoregression

$$\left[D + (\lambda_1+\lambda_2-\theta_1) + \frac{(\theta_1-\lambda_1)(\theta_1-\lambda_2)}{D+\theta_1}\right] x_t = \sigma DB_t$$

$$
\begin{aligned}
dx_t &= -[(\lambda_1+\lambda_2)-\theta_1]x_t\,dt \\
&\quad -\left((\theta_1-\lambda_1)(\theta_1-\lambda_2)\int_{\tau=0}^{\infty} e^{-\theta_1\tau}x_{t-\tau}\,d\tau\right)dt + \sigma DB_t.
\end{aligned}
$$

- Moving average representations for differences.

  $$(1-L)x_t = \mathcal{Z}_c(L)\varepsilon_t = (1-L)\mathcal{Z}_b(L)\varepsilon_t = [1+\mathcal{Z}_{\Delta b}(L)]\varepsilon_t$$

  The representation:

  $$\begin{aligned} dx_t &= \left(\int_{\tau=0}^{\infty} c(\tau)\sigma dB_{t-\tau}\right)dt + \sigma dB_t \\ Dx_t &= \mathcal{L}_c(D)\sigma DB_t. \end{aligned}$$

  Finding $\mathcal{L}_c(D)$ from $\mathcal{L}_b(D)$:

  $$\begin{aligned} \mathcal{L}_c(D) &= D\mathcal{L}_d(D) = 1 + \mathcal{L}_{b'}(D) \\ dx_t &= \left(\int_{\tau=0}^{\infty} b'(\tau)\sigma dB_{t-\tau}\right)dt + \sigma dB_t. \end{aligned}$$

  The AR(1):

  $$\begin{aligned} Dx_t &= \frac{D}{D+\phi}\sigma DB_t = \left(1 - \frac{\phi}{D+\phi}\right)\sigma DB_t \\ dx_t &= -\phi\left(\int_{\tau=0}^{\infty} e^{-\phi\tau}\sigma dB_{t-\tau}\right)dt + \sigma dB_t. \end{aligned}$$

  Polynomials:

  $$\mathcal{L}_c(D) = 1 - \frac{\lambda_1 A}{D+\lambda_1} - \frac{\lambda_2 B}{D+\lambda_2} - \cdots$$

- Impulse–response functions and multipliers.

  $$\begin{aligned} (E_t - E_{t-1})x_{t+j} &= b_j\varepsilon_t. \\ \text{“}\lim_{\Delta\to 0}(E_{t+\Delta} - E_t)\text{”} \quad x_{t+\tau} &= b(\tau)\sigma dB_t, \end{aligned}$$

  meaning, if $y_t = E_t x_{t+\tau}$, then

  $$dy_t = ()dt + b(\tau)\sigma dB_t.$$

  Impact multiplier:

  $$\begin{aligned} b_0 &= \mathcal{Z}_b(0) = 1 \\ b(0) &= \lim_{D\to\infty}[D\mathcal{L}_b(D)] = 1 \\ c(0) &= \lim_{D\to\infty}[\mathcal{L}_c(D)] = 1. \end{aligned}$$

Final multiplier:

$$\begin{aligned} b_\infty &= \mathcal{Z}_b(\infty) \\ b(\infty) &= \lim_{D\to 0}[D\mathcal{L}_b(D)]. \end{aligned}$$

These should be zero for a stationary x_t.

Cumulative response of $\int_{\tau=0}^{\infty} x_{t+\tau}\ d\tau$:

$$\begin{aligned} \mathcal{Z}_b(1) &= \sum_{j=0}^{\infty} b_j \\ \mathcal{L}_b(0) &= \int_{\tau=0}^{\infty} b(\tau)d\tau. \end{aligned}$$

Cumulative response of $x_t = \int_{\tau=0}^{\infty} dx_{t+\tau}$:

$$\begin{aligned} \mathcal{Z}_c(1) &= \sum_{j=0}^{\infty} c_j \\ \mathcal{L}_c(0) &= 1 + \int_{\tau=0}^{\infty} c(\tau)d\tau. \end{aligned}$$

These should be zero for a stationary x_t.

- Hansen–Sargent prediction formulas.

$$E_t \sum_{j=0}^{\infty} \beta^j x_{t+j} = \left(\frac{L\mathcal{Z}_b(L) - \beta\mathcal{Z}_b(\beta)}{L-\beta}\right)\varepsilon_t$$

$$E_t \sum_{j=1}^{\infty} \beta^{j-1} x_{t+j} = \left(\frac{\mathcal{Z}_b(L) - \mathcal{Z}_b(\beta)}{L-\beta}\right)\varepsilon_t$$

$$(E_t - E_{t-1})\sum_{j=0}^{\infty} \beta^j x_{t+j} = \mathcal{Z}_b(\beta)\varepsilon_t.$$

$$E_t \int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau}\ d\tau = \left(\frac{\mathcal{L}_b(D) - \mathcal{L}_b(r)}{r-D}\right)\sigma DB_t.$$

$$\text{“}\lim_{\Delta\to 0}(E_{t+\Delta} - E_t)\text{”} \int_{\tau=0}^{\infty} e^{-r\tau} x_{t+\tau}\ d\tau = \mathcal{L}_b(r)\sigma dB_t.$$

- Difference-stationary processes.

$$Dx_t = \mathcal{L}_c(D)\sigma DB_t$$
$$dx_t = \int_{\tau=0}^{\infty} c(\tau)\sigma dB_{t-\tau} + \sigma dB_t.$$

Polynomial example. In moving average form:

$$Dx_t = \frac{D+\theta}{D+\lambda}\sigma DB_t$$
$$Dx_t = \left(1 + \frac{\theta-\lambda}{D+\lambda}\right)\sigma DB_t$$
$$dx_t = (\theta-\lambda)\left(\int_{\tau=0}^{\infty} e^{-\lambda\tau}\sigma dB_{t-\tau}\right)dt + \sigma dB_t.$$

In autoregressive form:

$$\frac{D+\lambda}{D+\theta}\, Dx_t = \sigma DB_t$$
$$\left(1 + \frac{\lambda-\theta}{D+\theta}\right) Dx_t = \sigma DB_t$$
$$dx_t = -(\lambda-\theta)\left(\int_{\tau=0}^{\infty} e^{-\theta\tau}\, dx_{t-\tau}\right)dt + \sigma dB_t.$$

- Transforming from differences to levels, Beveridge–Nelson decompositions.

$$\mathcal{Z}_c(L) = \mathcal{Z}_c(1) + (1-L)\mathcal{Z}_b(L);\ b_j = -\sum_{k=j+1}^{\infty} c_k$$

implies

$$x_t = z_t + w_t;$$
$$(1-L)z_t = \mathcal{Z}_c(1)\varepsilon_t; \quad w_t = \mathcal{Z}_b(L)\varepsilon_t.$$

In continuous time,

$$Dx_t = \mathcal{L}_c(D)\sigma DB_t = [\mathcal{L}_c(0) + D\mathcal{L}_b(D)]\sigma DB_t$$

implies

$$x_t = z_t + w_t;$$
$$Dz_t = \mathcal{L}_c(0)\sigma DB_t; w_t = \mathcal{L}_b(D).$$

Constructing $\mathcal{L}_b(D)$:

$$\mathcal{L}_b(D) = \frac{\mathcal{L}_c(D) - \mathcal{L}_c(0)}{D}$$
$$b(\tau) = -\int_{s=\tau}^{\infty} c(s)ds.$$

z_t has the "trend" property

$$z_t = \lim_{T\to\infty} E_t\left(x_{t+T}\right) = x_t + E_t \int_{\tau=0}^{\infty} dx_{t+\tau}.$$

- Cointegration. Given the moving average representation,

$$Dx_t = \mathcal{L}_c(D)V\ DB_t$$

x_t are *cointegrated* if $\mathcal{L}_c(0)$ has rank less than N. Then

$$\mathcal{L}_c(0) = \delta\beta'$$

and there exist α, ϕ :

$$\alpha'\delta = \alpha'\mathcal{L}_c(0) = 0; \quad \beta'\phi = \mathcal{L}_c(0)\phi = 0.$$

The *common trend representation*

$$Dz_t = \beta' V DB_t$$
$$x_t = \delta z_t + w_t.$$

The *cointegrating vector* $\alpha' x_t = \alpha' w_t$ is stationary.
The *error correction representation* is

$$Dx_t = -\phi(\alpha' x_t) + \mathcal{L}_d(D)V\ DB_t.$$

Acknowledgments

I thank George Constantinides for helpful comments. I acknowledge research support from CRSP.

References

Beveridge, S. and C. R. Nelson (1981), 'A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle". *Journal of Monetary Economics* **7**, 151–174. http://dx.doi.org/10.1016/0304-3932(81)90040-4.

Cochrane, J. H. (2005a), *Asset Pricing: Revised Edition*. Princeton: Princeton University Press. http://press.princeton.edu/titles/7836.html.

Cochrane, J. H. (2005b), 'Time series for macroeconomics and finance'. Manuscript, University of Chicago. http://faculty.chicagobooth.edu/john.cochrane/research/papers/time_series_book.pdf.

Cochrane, J. H. (2012), 'A continous-time asset pricing model with habits and durability'. Manuscript, University of Chicago. http://faculty.chicagobooth.edu/john.cochrane/research/papers/linquad_asset_price_example.pdf.

Engle, R. F. and C. W. J. Granger (1987), 'Co-integration and error correction: Representation, estimation, and testing'. *Econometrica* **55**, 251–276.

Hansen, L. P. and T. J. Sargent (1980), 'Formulating and estimating dynamic linear rational-expectations models'. *Journal of Economic Dynamics and Control* **2**, 7–46.

Hansen, L. P. and T. J. Sargent (1981), 'A note on Wiener–Kolmogorov prediction formulas for rational expectations models'. *Economics Letters* **8**, 255–260.

Hansen, L. P. and T. J. Sargent (1991), 'Prediction formulas for continuous-time linear rational expectations models'. Chapter 8 of *Rational Expectations Econometrics*. https://files.nyu.edu/ts43/public/books/TOMchpt.8.pdf.

Heaton, J. (1993), 'The interaction between time-nonseparable preferences and time aggregation'. *Econometrica* **61**, 353–385. http://www.jstor.org/stable/2951555.