Lecture 5. Univariate Prediction Problems

1. The Wiener-Kolmogorov Prediction Formulas

As shown in Lecture 3, the solution to the problem

\[
\min_{\tilde{z} \in H_t(x)} E\left( x_{t+k} - \tilde{z} \right)^2
\]

is equal to the projection of \( x_{t+k} \) onto \( H_t(x) \), i.e. \( x_{t+k \dagger} \). Once again letting \( \sum_{j=0}^{\infty} \alpha_{j} e_{t+k-j} \) denote the fundamental MA representation of \( x_t \), from the equivalence of \( H_t(x) \) and \( H_t(e) \), it is immediate that \( x_{t+k \dagger} \) may be written as

\[
x_{t+k \dagger} = \sum_{j=0}^{\infty} \alpha_{j} e_{t-j}
\]

To achieve a more parsimonious expression, define the annihilation operator for lag polynomials.

Definition 5.1 Annihilation operator
Let \( \pi(L) \) be a polynomial lag operator of the form
\[ \ldots + \pi_2L^2 + \pi_1L^{-1} + \pi_0 + \pi_1L + \pi_2L^2 + \ldots \]
The annihilation operator \( L \cdot \) eliminates all negative terms, i.e.
\[ \pi(L) \cdot = \pi_0 + \pi_1L + \pi_2L^2 + \ldots \quad (5.3) \]

We can therefore re-express the optimal predictor as
\[ x_{t+k} = \left( \frac{\alpha(L)}{L^k} \right)_+ \varepsilon_t \quad (5.4) \]

When the MA polynomial is invertible, \( \alpha(L)^{-1}x_i = \varepsilon_i \), so that one can explicitly express the predictor as a weighted average of current and lagged \( x_i \)'s, (4.4) implies
\[ x_{t+k} = \left( \frac{\alpha(L)}{L^k} \right)_+ \alpha(L)^{-1}x_i \quad (5.5) \]

Equations (5.4) and (5.5) are known as the Wiener-Kolmogorov prediction formulas.

**Example 5.1. MA(1) process.**

Let \( x_i = \varepsilon_i - \rho \varepsilon_{i-1} \). The Wiener-Kolmogorov formula implies that optimal linear predictions of the process is
\[ x_{t+k} = \left( \frac{1-\rho L}{L^k} \right)_+ (1-\rho L)^{-1}x_i \quad (5.6) \]
which equals 0 for \( j > 1 \), as one would expect.

**Example 5.2. AR(1) process.**

If \( x_t = \rho x_{t-1} + \epsilon_t \), then the Wiener-Kolmogorov formula is

\[
x_{t+k} = \left( \sum_{i=0}^{\infty} \rho^{i+k} L^i \right) (1 - \rho L) x_t = \rho^k x_t
\]

(5.7)

2. The Hansen-Sargent formula for geometric distributed leads

The Wiener-Kolmogorov formula has played a critical role in the development of modern empirical macroeconomics. The reason for this is many dynamic aggregate models imply that one time series represents a combination of forecasts of others. In a pair of seminal papers (both of which are on the reading list), Lars Hansen and Thomas Sargent developed the machinery to understand how such expectations-based relationships lead to testable implications of macroeconomic theories. I develop the basic Hansen-Sargent formula in the context of the constant discount dividend stock price model. This is an example of a “geometric distributed leads” model in that one variable is a geometrically weighted sum of expected values of future levels of another variable

Defining the variables \( D_t = \text{dividends (measured in real terms)} \), \( P_t = \text{stock price (measured in real terms)} \), and \( \beta = \text{fixed discount rate} \), the constant discount dividend stock price asserts that the stock price series obeys

\[
P_t = \sum_{i=0}^{\infty} \beta^i E(D_{t+i}) = E\left( P^*_t \right)
\]

(5.8)

where
What is the economic interpretation of this model? The model can be thought of as follows. Suppose agents are risk neutral (i.e. utility from consumption at \( t \) is linear in the level) and discount utility geometrically at rate \( \beta \). One can think of ownership of a unit of stock as providing a stream of real consumption (via the dividends) if held forever. The opportunity cost of ownership is the consumption foregone today by buying a unit of stock. The price formula expresses the price at which an expected utility maximizer is indifferent between buying a unit or not, which is what the price must do in equilibrium.

How can one test whether stock prices are consistent with this theory? Suppose that

\[
D_t = \alpha(L)\epsilon_t. \tag{5.10}
\]

Then, one may test the theory by computing the projection of \( P_t \) onto \( H_t(D) \) that is implied by the theory, and comparing this projection to the actual projection \( P_t \) onto \( H_t(D) \). Put differently, the stock price model that has been described implies that

\[
\text{proj}(P_t | H_t(D)) = \sum_{j=0}^{\infty} \beta^j \text{proj}(D_{t+j} | H_t(D)) \tag{5.11}
\]

If one knows the process for the dividend series, one can derive the right hand side expression and compare it to the left hand side formula.

In order to derive the projection on the right hand side of this expression, one proceeds in two steps. First one projects \( P_t \) onto \( H_t(\epsilon) \) to generate a time series \( \pi(L)\epsilon_t \) and second, one inverts the \( \epsilon_t \)'s to generate \( D_t \)'s. The derivation relies on the implication of the model that \( P_t \) may also be expressed as:

\[
P_t' = \sum_{i=0}^{\infty} \beta^i D_{t+i} \tag{5.9}
\]
\[ E(P_i | H_i (D)) = \beta E(P_{i+1} | H_i (D)) + D_i \] (5.12)

which implies that

\[ \pi (L) \varepsilon_i = \beta \left( \frac{\pi(L)}{L} \right) \varepsilon_i + \alpha(L) \varepsilon_i \] (5.13)

or

\[ \pi (L) \varepsilon_i = \beta \left( \frac{\pi(L)}{L} - \frac{\pi_0}{L} \right) \varepsilon_i + \alpha(L) \varepsilon_i \] (5.14)

Algebraic manipulation yields

\[ \left( 1 - \beta L^{-1} \right) \pi (L) = \alpha(L) - \beta \pi_0 L^{-1} \] (5.15)

If \( L = \beta \), then \( \pi_0 = \alpha(\beta) \). Substituting into this expression and rearranging yields

\[ \pi (L) = \frac{\alpha(L) - \beta \alpha(\beta) L^{-1}}{1 - \beta L^{-1}} \] (5.16)

so that

\[ E(P_i | H_i (D)) = \frac{\alpha(L) - \beta \alpha(\beta) L^{-1}}{1 - \beta L^{-1}} \varepsilon_i = \frac{1 - \beta \alpha(\beta) L^{-1}}{1 - \beta L^{-1}} D_i \] (5.17)

The relationship between \( \pi (L) \) and \( \alpha(L) \) is complicated. For example, if one writes
\[
E(P_i|H_i(D)) = \sum_{i=0}^{\infty} \beta^i \left( \frac{\alpha(L)}{L'} \right) \epsilon_i 
\]

(5.18)

then it is apparent that

\[
\pi_k = \sum_{i=k}^{\infty} \beta^{i-k} \alpha_i 
\]

(5.19)

3. Some properties of forecasts

A number of tests of macroeconomic models may be constructed using properties of forecasts. For example, suppose that a variable \( x_t \) is, if a certain model holds, the optimal forecast of another variable \( y_t \) given an information set \( F_t \). Such a claim requires that the forecast error \( \eta_t = y_t - x_t \) is orthogonal to \( F_t \). This provides a simple way of testing such models. Here are some examples of implications that have been exploited in the empirical literature.

i. One of the elements of \( F_t \) is \( x_t \), so one implication of this is that forecasts must be orthogonal to forecast errors.

ii. Suppose that \( y_t \) is an element of \( F_{t+1} \). This will mean that the forecast errors \( \eta_t \) are uncorrelated.

iii. Since \( \text{var}(y_t) = \text{var}(x_t + \eta_t) = \text{var}(x_t) + \text{var}(\eta_t) \) when \( x_t \) is a forecast of \( y_t \), one can test this hypothesis by evaluating the implication that \( \text{var}(y_t) > \text{var}(x_t) \). This is known as an excess volatility test. These tests were introduced by Stephen LeRoy and Robert Shiller. Notice that \( \text{var}(y_t) = \text{var}(x_t + \eta_t) = \text{var}(x_t) + \text{var}(\eta_t) + \text{cov}(x_t, \eta_t) \), so that \( \text{var}(y_t) > \text{var}(x_t) \)
implies that \( \text{var}(\eta_t) > -2 \text{cov}(\eta_t, x_t) \), which may be rewritten as
\[
\frac{\text{cov}(\eta_t, x_t)}{\text{var}(\eta_t)} < -\frac{1}{2}.
\]
Hence an excess volatility test looks for negative nonlocal correlations between forecasts and forecast errors.

4. Bubbles

Returning to the dividend stock price model, the basic equation (5.8) may be rewritten
\[
P_t = \sum_{i=0}^{\infty} \beta^i E_t (D_{t+i}) = D_t + \sum_{i=1}^{\infty} \beta^i E_t (D_{t+i})
\]
\[
= D_t + \beta \sum_{i=0}^{\infty} \beta^i E_t (D_{t+i+1}) = D_t + \beta E_t \sum_{i=0}^{\infty} \beta^i E_{t+1} (D_{t+i+1})
\]
\[
= D_t + \beta E_t P_{t+1}
\]
(5.20)

This means that the “excess holding return”
\[
\beta P_{t+1} + D_t - P_t
\]
(5.21)
is unpredictable given all information available at time \( t \). A nature question is whether testing this unpredictability property is equivalent to testing the nonlinear restrictions embedded in the Hansen-Sargent formula.

One way to answer this question is to ask whether (5.21) imposes fewer restrictions on the data that (5.8). It turns out the answer is yes. Suppose that one appends to a \( P_t \) a process \( B_t \) with structure
\[
B_t = \beta^{-1} B_{t-1} + \mu_t
\]
(5.22)
where $\mu_t$ is unpredictable given information at $t-1$. This new process will violate (5.8) but not (5.21). Such a process is known as a bubble. Tests of the unpredictability of excess holding returns cannot detect bubbles as they do not place any restrictions on the source of the returns.

Notice that $B_t$ is an explosive process. When present, prices will not have a finite variance; the assumptions required to construct a Hilbert space around current and lagged prices are violated. Projections of prices onto current and lagged dividends are no longer defined. This has implications for the construction of tests for the presence of bubbles.