

Lecture Notes 4. Univariate Forecasting

1. The Wiener-Kolmogorov Prediction Formulas

From previous notes, the solution to the problem

$$\min_{\xi \in H_t(x)} E(x_{t+k} - \xi)^2 \quad (4.1)$$

is equal to the projection of x_{t+k} onto $H_t(x)$, denoted as $x_{t+k|t}$. Once again letting

$\sum_{j=0}^{\infty} \alpha_j \varepsilon_{t+k-j}$ denote the fundamental MA representation of x_t , from the equivalence of $H_t(x)$ and $H_t(\varepsilon)$, it is immediate that $x_{t+k|t}$ may be written as

$$x_{t+k|t} = \sum_{j=0}^{\infty} \alpha_{j+k} \varepsilon_{t-j} \quad (4.2)$$

The reason for this is simple: at time t , one can condition forecasts on $\varepsilon_t, \varepsilon_{t-1}$, i.e.

$H_t(\varepsilon)$. The component of x_{t+k} that is determined by $\varepsilon_{t+k}, \dots, \varepsilon_{t+1}$, i.e. $\sum_{j=0}^k \alpha_j \varepsilon_{t+k-j}$ is orthogonal to $H_t(\varepsilon)$.

To achieve a more parsimonious expression of expressions such as (4.2), define the annihilation operator for lag polynomials.

Definition 4.1 Annihilation operator

Let $\pi(L)$ be a polynomial lag operator of the form $\dots + \pi_{-2}L^{-2} + \pi_{-1}L^{-1} + \pi_0 + \pi_1L + \pi_2L^2 + \dots$. The annihilation operator, denoted as $(\cdot)_+$ eliminates all lag terms with negative exponents, i.e.

$$(\pi(L))_+ \triangleq \pi_0 + \pi_1L + \pi_2L^2 + \dots \quad (4.3)$$

Using the annihilation operator, one can re-express the optimal predictor as

$$x_{t+k|t} = \left(\frac{\alpha(L)}{L^k} \right)_+ \varepsilon_t \quad (4.4)$$

When the MA polynomial is invertible, $\alpha(L)^{-1} x_t = \varepsilon_t$, so that one can explicitly express the predictor as a weighted average of current and lagged x_t 's, (4.4) implies

$$x_{t+k|t} = \left(\frac{\alpha(L)}{L^k} \right)_+ \alpha(L)^{-1} x_t \quad (4.5)$$

Equations (4.4) and (4.5) are known as the Wiener-Kolmogorov prediction formulas.

Example 4.1. MA(1) process.

Let $x_t = \varepsilon_t - \rho\varepsilon_{t-1}$. The AR Wiener-Kolmogorov formula imply that optimal linear predictions of the process is

$$x_{t+k|t} = \left(\frac{1-\rho L}{L^k} \right)_+ (1-\rho L)^{-1} x_t \quad (4.6)$$

which equals 0 for $j > 1$, as one would expect since the process is unaffected by shocks more than one period in the past.

Example 4.2. AR(1) process.

If $x_t = \rho x_{t-1} + \varepsilon_t$, then the AR Wiener-Kolmogorov formula is

$$x_{t+k|t} = \left(\sum_{i=0}^{\infty} \rho^{i+k} L^i \right) (1 - \rho L) x_t = \rho^k x_t \quad (4.7)$$

The AR(1) thus has a very simple structure: $x_{t+k+1|t} = \rho x_{t+k|t}$.

2. The Hansen-Sargent formula for geometrically discounted 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. This machinery has been of immense importance in the development of modern empirical macroeconomics. 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 $D_t =$ dividends (measured in real terms), $P_t =$ stock price (measured in real terms), and $\beta =$ fixed discount rate, the constant discount dividend stock price asserts that the stock price series obeys

$$P_t = \sum_{i=0}^{\infty} \beta^i \mathbb{E}(D_{t+i}) = \mathbb{E}(P_t^*) \quad (4.8)$$

where

$$P_t^* = \sum_{i=0}^{\infty} \beta^i D_{t+i} \quad (4.9)$$

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 β . 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 dividends have the fundamental moving average representation.

$$D_t = \alpha(L) \varepsilon_t. \quad (4.10)$$

This representation is recoverable from the data. One may use this representation to 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 direct projection P_t onto $H_t(D)$. Put differently, the stock price model that has been described implies that

$$\text{proj}(P_t | H_t(D)) = \sum_{j=0}^{\infty} \beta^j \text{proj}(D_{t+j} | H_t(D)) \quad (4.11)$$

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 $H_t(\varepsilon)$ to generate a time series $\pi(L)\varepsilon_t$ and second, one inverts the ε_t 's to generate D_t 's. The derivation relies on the implication of the model that P_t may also be expressed as:

$$E(P_t | H_t(D)) = \beta E(P_{t+1} | H_t(D)) + D_t \quad (4.12)$$

which implies that

$$\pi(L)\varepsilon_t = \beta \left(\frac{\pi(L)}{L} \right)_+ \varepsilon_t + \alpha(L)\varepsilon_t \quad (4.13)$$

or

$$\pi(L)\varepsilon_t = \beta \left(\frac{\pi(L)}{L} - \frac{\pi_0}{L} \right) \varepsilon_t + \alpha(L)\varepsilon_t \quad (4.14)$$

Algebraic manipulation yields

$$(1 - \beta L^{-1})\pi(L) = \alpha(L) - \beta\pi_0 L^{-1} \quad (4.15)$$

To manipulate this expression in order to express $\pi(L)$ in terms of $\alpha(L)$, notice that $L = \beta$, then $\pi_0 = \alpha(\beta)$. Substituting into this expression and rearranging yields

$$\pi(L) = \frac{\alpha(L) - \beta\alpha(\beta)L^{-1}}{1 - \beta L^{-1}} \quad (4.16)$$

which leads to

$$E(P_t | H_t(D)) = \frac{\alpha(L) - \beta\alpha(\beta)L^{-1}}{1 - \beta L^{-1}} \varepsilon_t = \frac{1 - \beta\alpha(\beta)\alpha(L)^{-1}L^{-1}}{1 - \beta L^{-1}} D_t \quad (4.17)$$

This relationship between $\pi(L)$ and $\alpha(L)$ is complicated. For example, if one writes

$$E(P_t | H_t(D)) = \sum_{i=0}^{\infty} \beta^i \left(\frac{\alpha(L)}{L^i} \right)_+ \varepsilon_t \quad (4.18)$$

then it is apparent that

$$\pi_k = \sum_{i=k}^{\infty} \beta^{i-k} \alpha_i \quad (4.19)$$

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 , i.e.

$$x_t = \text{proj}(y_t | F_t) \quad (4.20)$$

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 η_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 claimed relationship by evaluating the implication that the variances of realizations should be greater than the variance of rational forecasts of the realizations, i.e.

$$\text{var}(y_t) > \text{var}(x_t) \quad (4.21)$$

(The two are equal when forecasts errors are always equal to 0; I ignore this uninteresting case.) Evaluation of whether this inequality holds for a data series is known as an excess volatility test. These tests were introduced by Stephen LeRoy and Robert Shiller. One can construct many forms of excess volatility tests. Suppose that $z_t \in F_t$. Then it is immediate that

$$\text{var}(y_t - z_t) > \text{var}(x_t - z_t) \quad (4.22)$$

Notice that $\text{var}(y_t) = \text{var}(x_t + \eta_t) = \text{var}(x_t) + \text{var}(\eta_t) + 2\text{cov}(x_t, \eta_t)$, 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. Excess volatility can thus be tested based on the regression

$$x_t = \eta_t b + \xi_t \quad (4.23)$$

Excess volatility requires that the coefficient in this regression is less than $-\frac{1}{2}$. This regression has an unusual form as it regresses the forecast against the forecast errors, rather than the forecast error against the forecast, which is the regression suggested by the requirement that the forecast errors generated by rational forecasts are unpredictable.

Put differently, excess volatility implies predictability of forecast errors, although predictability of forecast errors does not imply excess volatility.

4. Bubbles

Returning to the dividend stock price model, the basic equation (4.8) may be rewritten

$$\begin{aligned}
 P_t &= \sum_{i=0}^{\infty} \beta^i E_t(D_{t+i}) = D_t + \sum_{i=1}^{\infty} \beta^i E_t(D_{t+i}) \\
 &= D_t + \beta \sum_{i=0}^{\infty} \beta^i E_t(D_{t+i+1}) = D_t + \beta E_t \sum_{i=0}^{\infty} \beta^i E_{t+1}(D_{t+i+1}) \\
 &= D_t + \beta E_t P_{t+1}
 \end{aligned} \tag{4.24}$$

This means that the “excess holding return”

$$\beta P_{t+1} + D_t - P_t \tag{4.25}$$

is unpredictable given all information available at time t as it is simply a particular forecast error. A natural question is whether testing this unpredictability property is equivalent to testing the nonlinear restrictions embedded in the Hansen-Sargent formula which relates the level of stock prices to the dividend process.

One way to answer this question is to ask whether (4.21) imposes fewer restrictions on the data than (4.8). It turns out the answer is yes. Suppose that one appends to a P_t a process B_t with structure

$$B_t = \beta^{-1} B_{t-1} + \mu_t \quad (4.26)$$

where μ_t is unpredictable given information at $t-1$. This new process will violate (4.8) but not (4.21). 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 and in fact the assumptions required to construct a Hilbert space around current and lagged prices are violated. Projections of prices onto current and lagged dividends will no longer be defined. This has implications for the construction of tests for the presence of bubbles.