Finance 520,
Time Series Analysis
Christiano

Solving Projection Problems Using Spectral Analysis

This note describes the use of the tools of spectral analysis to solve projection problems. The four tools used are the Wold decomposition theorem, the 'annihilation operator', the cross spectrum and partial fractions expansions.

1. The Problem

The problem is to solve:

$$P\left[y_t | x_t, x_{t-1}, x_{t-2}, ...\right] = \sum_{j=0}^{\infty} d_j x_{t-j}, \tag{1.1}$$

where y_t and x_t are two purely indeterministic, covariance stationary stochastic processes. The projection problem in (1.1) can arise in various ways. For example, this is a standard forecasting problem if $y_t = x_{t+k}$, $k > 0$. Another example is a classic from the early days of rational expectations, due to Muth (1960). In Muth's problem an agent is interested in tracking a variable, y_t, given a signal, x_t, which is corrupted by measurement error:

$$\begin{aligned} x_t &= y_t + \varepsilon_t & (1.2) \\ y_t &= \rho y_{t-1} + u_t. & (1.3) \end{aligned}$$

Here, u_t is *iid* and orthogonal to y_{t-s}, $s > 0$. Also, ε_t is orthogonal to u_t at all leads and lags.

At first glance, the solution to (1.1) seems intractable. The presence of an infinity of right hand 'explanatory variables' in (1.1) requires, in effect, inverting an infinite ordered matrix to solve for the d_j's. To see this, note that the orthogonality condition associated with the projection problem requires

$$c_{yx}(k) = \sum_{j=0}^{\infty} d_j c_{xx}(k-j), \tag{1.4}$$

where

$$c_{yx}(k) \equiv E y_t x_{t-k}, \;\; c_{xx}(l) \equiv E x_t x_{t-l}.$$

For illustration, suppose that the agent is only interested in using a finite record of past observations on $x_t, ..., x_{t-l}$ to estimate y_t. In this case, the ∞ in (1.4) is replaced by l and these equations are represented in matrix form as follows:

$$\begin{pmatrix} c_{yx}(0) \\ c_{yx}(1) \\ \vdots \\ c_{yx}(l) \end{pmatrix} = \begin{bmatrix} c_{xx}(0) & c_{xx}(-1) & \cdots & c_{xx}(-l) \\ c_{xx}(1) & c_{xx}(0) & \cdots & c_{xx}(1-l) \\ \vdots & \vdots & \ddots & \vdots \\ c_{xx}(l) & c_{xx}(l-1) & \cdots & c_{xx}(0) \end{bmatrix} \begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_l \end{pmatrix}.$$

To solve for the d_j's evidently requires inverting the $l+1 \times l+1$ matrix in square brackets. When $l \to \infty$ as in (1.1) and (1.4), this means inverting an infinite dimensional matrix. This definitely looks difficult!

2. Getting Help from the Wold Decomposition Theorem

The problem with the matrix inversion just described is that autocorrelation among the x_t's ensures that the matrix to be inverted grows non-trivially larger as l gets large. If, for example, the matrix were simply diagonal for all l, the inversion would not be a problem, no matter how large l is. The Wold decomposition theorem in effect allows us to replace the projection (1.1) with a projection on non-autocorrelated variables and thus solve the matrix inversion problem. In particular, according to the Wold decomposition theorem any purely indeterministic process can be linearly decomposed in terms of its history of one-step-ahead forecast errors, which are themselves not autocorrelated over time. Denoting the one-step-ahead forecast error in x_t by w_t, it follows that any variable that can be constructed by some linear function of $[x_t, x_{t-1}, ...]$ can also be recovered by a linear function of $[w_t, w_{t-1}, ...]$. This is because, according to the Wold decomposition theorem, there is a square summable sequence, $\{d_0, d_1, ...\}$, such that

$$x_t = d(L) w_t.$$

The operator, $d(L)$, is easy to find for our example. Apply $(1 - \rho L)$ to both sides of (1.2) and take into account (1.3) to obtain:

$$(1 - \rho L) x_t = u_t + \varepsilon_t - \rho \varepsilon_{t-1}.$$

The variance and lag-one autocovariance of the term on the right side of the equality is $\sigma_u^2 + \left(1 + \rho^2\right) \sigma_\varepsilon^2$, and $-\rho\sigma_\varepsilon^2$, respectively. We obtain the Wold representation for this term by choosing $|\lambda| \leq 1$ and σ_w^2 to solve

$$\frac{\rho\sigma_\varepsilon^2}{\sigma_u^2 + (1 + \rho^2)\sigma_\varepsilon^2} = \frac{\lambda}{1 + \lambda^2},\ \sigma_w^2 = \frac{\rho\sigma_\varepsilon^2}{\lambda}.$$

Then,

$$d(L) = \frac{1 - \lambda L}{1 - \rho L},\ Ew_t^2 = \sigma_w^2.$$

Note that if the measurement errors, σ_ε^2, are large compared with σ_u^2 then $\rho = \lambda$ and $\sigma_w^2 = \sigma_\varepsilon^2$ so that $x_t = \varepsilon_t$, as we would expect. In the other extreme, when $\sigma_\varepsilon^2 \to 0$ we can show that $x_t = y_t$ and $w_t = u_t$. Note that in our example, w_t can be expressed as a square summable sum of current and past x_t's:

$$w_t = \frac{1 - \rho L}{1 - \lambda L} x_t = \frac{u_t + \varepsilon_t - \rho\varepsilon_{t-1}}{1 - \lambda L}. \tag{2.1}$$

The expression on the right of the equality shows how a realization of current and past shocks, u_t and ε_t, maps into a realization of Wold errors.[1]

To gain intuition into (2.1), suppose that $\sigma_\varepsilon^2/\sigma_u^2$ is large, so that $\rho \simeq \lambda$. In this case, (2.1) indicates that an ε_t shock generates virtually no persistence in w_t.[2] This is not surprising since in this case x_t basically *is* ε_t. It is also not surprising that when $\sigma_\varepsilon^2/\sigma_u^2$ is large and there is a positive shock to u_t, this generates a long sequence of positive Wold forecast errors, w_t, when ρ is large. To understand this, note from (1.2) and (1.3) that if x_t jumps due to u_t when $\sigma_\varepsilon^2/\sigma_u^2$ is large, then the optimal response is to treat this as a jump in ε_t and not revise up one's forecast of x_t. But, when u_t jumps, y_t remains high for a long time when ρ is large, and this translates into a persistent

[1] It is a useful exercise to verify that w_t is not autocorrelated over time.

[2] This may not be obvious at first sight, but reflects that $1 - \rho L$ and 1-λL roughly cancel. A simple simulation will verify that a perturbation in ε_t generates a virtually completely transitory response in w_t when $\rho \simeq \lambda$.

rise in x_t. This persistent rise in x_t will at first be assumed to be a persistent sequence of positive realizations in ε_t. At some point the improbability of this assumption becomes overwhelming and it is optimal to start thinking that the observed rise in x_t was due to u_t and not ε_t.

It is convenient to consider the moving average representation of the vector process, $X_t = (y_t, w_t)'$, as well as its spectral density. Using (1.3) and (2.1):

$$\begin{aligned} X_t &= C(L)\begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix}, \\ \begin{pmatrix} y_t \\ w_t \end{pmatrix} &= \begin{bmatrix} \frac{1}{1-\rho L} & 0 \\ \frac{1}{1-\lambda L} & \frac{1-\rho L}{1-\lambda L} \end{bmatrix}\begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix} \end{aligned}$$

The spectral density of X_t is:

$$S_X\left(e^{-i\omega}\right) = C\left(e^{-i\omega}\right) V C\left(e^{i\omega}\right)',\ V = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix}.$$

As in result #1 in the class handout on spectral analysis, we have

$$\begin{aligned} S_X\left(e^{-i\omega}\right) &= c(0) + c(1)e^{-i\omega} + c(2)e^{-i2\omega} + ... \\ &\quad + c(1)'e^{i\omega} + c(2)'e^{i2\omega} + ... , \end{aligned}$$

where

$$c(j) = EX_tX_{t-j}' = \begin{bmatrix} c_{yy}(j) & c_{yw}(j) \\ c_{wy}(j) & c_{ww}(j) \end{bmatrix},$$

and

$$c_{yw}(j) = Ey_tw_{t-j},\ c_{wy}(j) = Ew_ty_{t-j},\ c_{ww}(j) = Ew_tw_{t-j}.$$

Similarly,

$$S_X\left(e^{-i\omega}\right) = \begin{bmatrix} S_{yy}\left(e^{-i\omega}\right) & S_{yw}\left(e^{-i\omega}\right) \\ S_{wy}\left(e^{-i\omega}\right) & S_{ww}\left(e^{-i\omega}\right) \end{bmatrix} = \begin{bmatrix} \frac{\sigma_u^2}{(1-\rho e^{-i\omega})(1-\rho e^{i\omega})} & \frac{\sigma_u^2}{(1-\rho e^{-i\omega})(1-\lambda e^{i\omega})} \\ \frac{\sigma_u^2}{(1-\lambda e^{-i\omega})(1-\rho e^{i\omega})} & \frac{\sigma_u^2}{(1-\rho e^{-i\omega})(1-\rho e^{i\omega})} + \frac{\sigma_\varepsilon^2}{(1-\lambda e^{-i\omega})(1-\lambda e^{i\omega})} \end{bmatrix}$$

Consider the following projection:

$$P\left[y_t|w_t, w_{t-1}, ...\right] = \sum_{j=0}^{\infty}\phi_j w_{t-j}.$$

The orthogonality condition which is necessary and sufficient to solve the projection problem is:

$$E\left[y_t - \sum_{j=0}^{\infty}\phi_j w_{t-j}\right]w_{t-k} = 0,\ k = 0, 1, ... ,$$

which is satisfied by setting the ϕ_j's as follows:

$$\phi_j = \frac{Ey_tw_{t-j}}{\sigma_w^2},\ \text{for all } j,\ \sigma_w^2 \equiv Ew_t^2.$$

The $z-$transform of the ϕ_j's is

$$\phi(z) = \sum_{j=0}^{\infty} \phi_j z^j = \frac{\sum_{j=0}^{\infty} c_{yw}(j) z^j}{\sigma_w^2} = \frac{[S_{yw}(z)]_+}{\sigma_w^2} = \frac{1}{\sigma_w^2} \left[\frac{\sigma_u^2}{(1-\rho z)(1-\lambda z^{-1})} \right]_+ .$$

Here, the '+' subscript indicates the 'annihilation operator', which means 'ignore terms in negative powers of z'. Our task now is to develop a simple expression for $[\cdot]_+$.

3. Finishing Off the Problem Using the Annihilation Operator and Partial Fractions Expansions

It is convenient to develop the partial fraction expansion of the term inside the annihilation operator. This is expressed as follows:

$$S_{yw}(z) = \frac{\sigma_u^2}{(1-\rho z)(1-\lambda z^{-1})} = \frac{z\sigma_u^2}{(1-\rho z)(z-\lambda)} = \frac{A_1}{1-\rho z} + \frac{A_2}{z-\lambda},$$

where A_1 and A_2 are constants whose values are to be determined. In the usual undetermined coefficient style, the values of A_1 and A_2 are determined by the requirement that the third equality in the last expression be satisfied. After multiplying both sides of that equality by $(1-\rho z)(z-\lambda)$, we obtain:

$$\begin{aligned} z\sigma_u^2 &= A_1(z-\lambda) + A_2(1-\rho z) \\ &= A_2 - A_1\lambda + (A_1 - \rho A_2) z. \end{aligned}$$

For the left and right sides to be equal requires

$$\begin{aligned} A_2 &= A_1\lambda \\ A_1 - \rho A_2 &= \sigma_u^2, \end{aligned}$$

or,

$$A_1 = \frac{\sigma_u^2}{1-\rho\lambda},\ A_2 = \frac{\lambda\sigma_u^2}{1-\rho\lambda}.$$

We write the partial fraction expansion of the cross spectrum as follows:

$$S_{yw}(z) = \frac{\sigma_u^2}{(1-\rho z)(1-\lambda z^{-1})} = \frac{\sigma_u^2}{1-\rho\lambda}\frac{1}{1-\rho z} + \frac{\lambda\sigma_u^2}{1-\rho\lambda}\frac{z^{-1}}{1-\lambda z^{-1}}. \tag{3.1}$$

The two terms in the partial fractions expansion are:

$$\begin{aligned} \frac{\sigma_u^2}{1-\rho\lambda}\frac{1}{1-\rho z} &= \frac{\sigma_u^2}{1-\rho\lambda}\left[1 + \rho z + \rho z^2 + ...\right] \\ \frac{\lambda\sigma_u^2}{1-\rho\lambda}\frac{z^{-1}}{1-\lambda z^{-1}} &= \frac{\lambda\sigma_u^2}{1-\rho\lambda}\left[z^{-1} + \lambda z^{-2} + \lambda^2 z^{-3} + ...\right]. \end{aligned}$$

Before continuing, we pause to take a closer look at some steps taken above which may at first seem arbitrary. For example, we might have considered the following expansion:

$$\frac{\sigma_u^2}{(1-\rho z)(1-\lambda z^{-1})} = \frac{\tilde{A}_1}{1-\rho z} + \frac{\tilde{A}_2}{1-\lambda z^{-1}}.$$

It is easy to see that there does not exist an expansion like this, because there do not exist constants, $\tilde{A}_1$ and $\tilde{A}_2$, that satisfy this equation. Multiply both sides by $(1-\rho z)\left(1-\lambda z^{-1}\right)$ to obtain:

$$\sigma_u^2 = \tilde{A}_1\left(1-\lambda z^{-1}\right) + \tilde{A}_2\left(1-\rho z\right),$$

and note that there are no values for $\tilde{A}_1$ and $\tilde{A}_2$ so that the expression on the right is independent of z. Another step taken above may also appear arbitrary at first glance. In particular, we could have adopted the following expansion:

$$\frac{A_2}{z-\lambda} = \frac{-\lambda^{-1}}{-\lambda^{-1}}\frac{A_2}{z-\lambda} = \frac{A_2\lambda^{-1}}{1-\lambda^{-1}z} = A_2\lambda^{-1}\left[1+\lambda^{-1}z+\lambda^{-2}z^2+\lambda^{-3}z^3+...\right]. \tag{3.2}$$

We did not adopt this expansion because, although it is technically valid, it is not relevant for our purposes. We seek the expansion of $S_{yw}(z)$, the $z-$transform of a cross-covariance function, where the cross-covariances are square summable:

$$S_{yw}(z) = \sum_{\tau=-\infty}^{\infty} c_{yw}(\tau)\, z^{\tau}.$$

Although the polynomial representation of $S_{yw}(z)$ technically has many expansions in powers of z, only one corresponds to the one we are interested in. The expansion that interests us is the one whose coefficients on powers of z are square summable. From this perspective, (3.2) is not of interest because the coefficients on powers of z are exploding. It is our special interest in the expansion that involves square summable coefficients that drove us to adopt the expansion of $A_2/(z-\lambda)$ that is implicit in (3.1).

In sum, we have expressed $S_{yw}(z)$ in the form of (3.1), with the understanding that the first polynomial represents an expansion in positive powers of z while the second polynomial represents an expansion in purely negative powers of z. This representation of $S_{yw}(z)$ puts us in a perfect position to evaluate the annihilation operator. Using the obvious fact that $g(z) = a(z) + b(z)$ implies

$$[g(z)]_+ = [a(z)]_+ + [b(z)]_+,$$

we obtain:

$$\begin{aligned} [S_{yw}(z)]_+ &= \left[\frac{\sigma_u^2}{1-\rho\lambda}\frac{1}{1-\rho z}\right]_+ + \left[\frac{\lambda\sigma_u^2}{1-\rho\lambda}\frac{z^{-1}}{1-\lambda z^{-1}}\right]_+ \\ &= \frac{\sigma_u^2}{1-\rho\lambda}\frac{1}{1-\rho z}. \end{aligned}$$

We conclude that

$$\begin{aligned} P\left[y_t|w_t, w_{t-1}, ...\right] &= \sum_{j=0}^{\infty}\phi_j w_{t-j} == \frac{[S_{yw}(z)]_+}{\sigma_w^2} w_t \\ &= \frac{\sigma_u^2/\sigma_w^2}{1-\rho\lambda}\frac{1}{1-\rho z} w_t \\ &= \frac{\sigma_u^2/\sigma_w^2}{1-\rho\lambda}\frac{1}{1-\lambda L} x_t, \end{aligned}$$

where the last term makes use of (2.1). Thus,

$$P\left[y_t|x_t, x_{t-1}, ...\right] = \frac{\sigma_u^2/\sigma_w^2}{1-\rho\lambda}\sum_{j=0}^{\infty}\lambda^j x_{t-j}.$$

This is a pretty simple solution to what at first may have appeared to be a formidable projection problem. The expression says that when data, y_t, are corrupted by noise as in (1.2)-(1.3), it makes sense to 'smooth' the observed data when using them to formulate a guess about y_t. For a more complete discussion of the material in this handout, see Sargent (1979, section 17, page 275) and Whittle (1983, section 3.7, page 42).

4. Discussion

The tools described here are not just useful for solving the projection problem in (1.1). They are also useful in solving for the equilibrium in certain economic models. In these models, the noise which corrupts observations is idiosyncratic to individual agents and the way they do their forecasting helps determine the structure of the stochastic processes driving the variables. Riccardo Masolo studies a simple example of this in his thesis work. In his problem, there is a process denoted l_t, which is analogous to y_t above and which evolves according to an $AR(1)$:

$$l_t = \rho l_{t-1} + \eta_t.$$

Here, l_t is an aggregate variable observed with noise by individual agents. The noise is idiosyncratic across different agents and agent h observes:

$$s_t^h = l_t + \varepsilon_t^h.$$

Here, ε_t^h is a measurement error, which averages out to zero across the large number of agents in the economy. Based on observing the history of signals, the h^{th} agent makes a decision, denoted w_t^h according to following rule:

$$w_t^h = \alpha_1 P\left[w_t|s_t^h, s_{t-1}^h, s_{t-2}^h, ...\right] + \alpha_2 P\left[l_t|s_t^h, s_{t-1}^h, s_{t-2}^h, ...\right].$$

Here, w_t is the aggregate of all agent decisions. Agents do not see this aggregate, since they only see their own history of s_t^h. The actions of individual agents depend on the time series representation of w_t. At the same time, the time series representation of w_t is determined by those same actions. This fixed point problem can easily be solved with an extension of the methods described in this handout (this is a point that was particularly stressed by Kasa (2000).)

References

[1] Kasa, Kenneth, 2000, "Forecasting the Forecasts of Others in the Frequency Domain," *Review of Economic Dynamics* 3, 726-56.

[2] Muth, John F., 1960, "Optimal Properties of Exponentially Weighted Forecasts,' *Journal of the American Statistical Association*, 55, 299-306.

[3] Sargent, Thomas, 1979, *Macroeconomic Theory*.

[4] Whittle, Peter, 1983, *Prediction and Regulation by Linear Least-Squares Methods*, second edition.