LECTURE NOTE 5

AUTOREGRESSION IN PANEL DATA

1. STRUCTURAL REGRESSION MODEL

Consider a population of people. Choose a person at random. For each year from $t = 1, \ldots, T$ there is an earnings variable Y_t. There is also a characteristic A of the individual that does not vary over time (but does vary across individuals). We have access to data generated by a random sample of size N from this population. The data have realized values of the Y_t, but A is not observed. Our objective is to measure the (predictive) effect of Y_{t-1} on Y_t holding A constant.

The structural regression model is

$$E(Y_t \mid Y_1, \ldots, Y_{t-1}, A) = \lambda_t + \theta Y_{t-1} + A \qquad (t = 2, \ldots, T). \tag{1}$$

This is called autoregression because we are predicting Y_t using the past values of the same variable. Two sorts of restrictions are being imposed. We condition on all the past values from Y_1 to Y_{t-1}, but the assumption in (1) is that only Y_{t-1} matters (once we control for A). So there are exclusion restrictions. The other sort of restriction in (1) is that the functional form is very simple, with no interaction terms involving Y_{t-1} and A. The intercept λ_t allows for an unrestricted additive period effect. The partial effects of Y_{t-1} and A are assumed to be constant over time (but we could extend the analysis and allow for $\theta_{1t} Y_{t-1} + \theta_{2t} A$). Given the assumption that the partial effect of A is constant over time, it is not restrictive to set the coefficient on A equal to one. Because A is not observed, we can scale it so that the coefficient is one.

The sample data are y_{it} for $i = 1, \ldots, N$ individuals and $t = 1, \ldots, T$ periods. For example, y_{it} could be the earnings of person i in year t. Data on the variable A are not available.

2. SOLVING THE MODEL

We are going to solve out the lagged dependent variables and express Y_t as a function just of A and prediction errors. This will lead to a formula for the covariances between Y_t and Y_s $(s, t = 1, \ldots, T)$. We will see that θ can be expressed as a function of these covariances. Then we can use the sample covariances to obtain an estimate of θ.

To solve the model, we must do something about Y_1, because we no not observe Y_0. We handle this initial conditions problem by introducing a linear predictor for Y_1:

$$E^*(Y_1 \mid 1, A) = \delta_0 + \delta_1 A.$$

Note that this does not introduce additional restrictions. All the restrictions are expressed in the structural regression model in (1). We shall impose the normalization that

$$E(A) = 0.$$

This is not restrictive, because we could redefine A as $A - E(A)$, and redefine the period effect λ_t as $\lambda_t + E(A)$. The period effects would still be unrestricted.

Define the prediction errors

$$\begin{aligned} U_t &= Y_t - E^*(Y_t \mid 1, Y_1, \ldots, Y_{t-1}, A) \qquad (t = 2, \ldots, T), \\ V &= Y_1 - E^*(Y_1 \mid 1, A). \end{aligned}$$

Then we have

$$Y_1 = \delta_0 + \delta_1 A + V, \tag{2}$$

$$Y_t = \lambda_t + \theta Y_{t-1} + A + U_t \qquad (t = 2, \ldots, T). \tag{3}$$

By construction, the prediction error V is orthogonal to 1 and A. So $E(V) = 0$ and V is uncorrelated with A. Likewise, by construction, the prediction error U_t is orthogonal to $1, Y_1, \ldots, Y_{t-1}, A$. So $E(U_t) = 0$ and U_t is uncorrelated with $Y_1, \ldots, Y_{t-1}$, and with A. Because U_{t-1} is a linear function of Y_{t-1}, Y_{t-2}, and A, the covariance between U_t and U_{t-1} is zero. Likewise, U_t is uncorrelated with $U_{t-2}, \ldots, U_2$. In addition, since V is a linear function of Y_1 and A, the covariance between U_t and V is zero. So after we have solved out the lagged Y's, and have expressed the Y's as linear functions of $A, U_2, \ldots, U_T, V$, it will be straightforward to obtain a formula for the covariances between the Y's.

Substitute for Y_1 in the equation for Y_2 in (3):

$$\begin{aligned} Y_2 &= \lambda_2 + \theta[\delta_0 + \delta_1 A + V] + A + U_2 \\ &= (\lambda_2 + \theta\delta_0) + (1 + \theta\delta_1)A + \theta V + U_2. \end{aligned} \tag{4}$$

Recursively substitute for Y_2 from (4) into the equation for Y_3 in (3):

$$\begin{aligned} Y_3 &= \lambda_3 + \theta[(\lambda_2 + \theta\delta_0) + (1 + \theta\delta_1)A + \theta V + U_2] + A + U_3 \\ &= (\lambda_3 + \theta\lambda_2 + \theta^2\delta_0) + (1 + \theta + \theta^2\delta_1)A + \theta^2 V + U_3 + \theta U_2. \end{aligned}$$

We could continue with this recursive substitution, but it is more convenient to use matrix notation. With $T = 3$:

$$\begin{pmatrix} 1 & 0 & 0 \\ -\theta & 1 & 0 \\ 0 & -\theta & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \delta_0 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ 1 \\ 1 \end{pmatrix} A + \begin{pmatrix} V \\ U_2 \\ U_3 \end{pmatrix}.$$

We can write this as

$$B(\theta)Y = \begin{pmatrix} \delta_0 \\ \lambda \end{pmatrix} + \begin{pmatrix} \delta_1 \\ 1_2 \end{pmatrix} A + \begin{pmatrix} V \\ U \end{pmatrix}.$$

It is straightforward to check that

$$B(\theta)^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \theta & 1 & 0 \\ \theta^2 & \theta & 1 \end{pmatrix}.$$

The recursive substitution to solve out the lagged Y's corresponds to multiplying by $B(\theta)^{-1}$:

$$Y = B(\theta)^{-1} \begin{pmatrix} \delta_0 \\ \lambda \end{pmatrix} + B(\theta)^{-1} \left[\begin{pmatrix} \delta_1 \\ 1_2 \end{pmatrix} A + \begin{pmatrix} V \\ U \end{pmatrix} \right]. \tag{5}$$

3. COVARIANCE MATRIX

With

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix}$$

a $T \times 1$ matrix of random variables, define $E(Y)$ to be the $T \times 1$ matrix with t^{th} element equal to $E(Y_t)$. Define $\text{Cov}(Y)$ to be the $T \times T$ matrix with (s,t) element equal to $\text{Cov}(Y_s, Y_t)$. For notation, let $\mu = E(Y)$ and $\Sigma = \text{Cov}(Y)$. With $T = 3$:

$$\mu = E(Y) = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ E(Y_3) \end{pmatrix},$$

$$\Sigma = \text{Cov}(Y) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \quad \text{with} \quad \sigma_{st} = \text{Cov}(Y_s, Y_t) = \sigma_{ts}.$$

We shall focus on the population $E(Y)$ and covariance matrix $\text{Cov}(Y)$ because they have direct sample counterparts. Given the data y_{it} $(i = 1, \ldots, N;\ t = 1, \ldots, T)$, we can estimate the population means using sample means:

$$\hat{\mu}_t = \bar{y}_t = \frac{1}{N} \sum_{t=1}^{T} y_{it}.$$

Note that

$$\sigma_{st} = \text{Cov}(Y_s, Y_t) = E[(Y_s - \mu_s)(Y_t - \mu_t)] = E(Y_s Y_t) - \mu_s \mu_t.$$

This suggests the estimate

$$\hat{\sigma}_{st} = \frac{1}{N} \sum_{i=1}^{N} (y_{is} - \bar{y}_s)(y_{it} - \bar{y}_t) = \frac{1}{N} \sum_{i=1}^{N} y_{is} y_{it} - \bar{y}_s \bar{y}_t. \tag{6}$$

The following result makes it easy to use equation (5) express the covariance matrix of Y as an explicit function of θ and some other parameters.

Claim 1. Suppose that the random matrix Y is $T \times 1$; and the nonrandom matrices d_1 and d_2 are $M \times T$ and $M \times 1$. Then

$$\mathrm{Cov}(d_1 Y + d_2) = d_1 \mathrm{Cov}(Y) d_1'.$$

Proof. First note that with $\tilde{Y} = Y - E(Y)$,

$$\mathrm{Cov}(Y) = E(\tilde{Y}\tilde{Y}').$$

With $W = d_1 Y + d_2$, we have

$$\begin{aligned} E(W) &= d_1 E(Y) + d_2 \\ \tilde{W} = W - E(W) &= d_1[Y - E(Y)] = d_1 \tilde{Y}, \end{aligned}$$

and so

$$\mathrm{Cov}(W) = E(\tilde{W}\tilde{W}') = d_1 E(\tilde{Y}\tilde{Y}') d_1' = d_1 \mathrm{Cov}(Y) d_1'. \quad \diamond$$

Applying this result to equation (5) (with $T = 3$) gives

$$\Sigma = \mathrm{Cov}(Y) = B(\theta)^{-1}\left[\begin{pmatrix} \delta_1 \\ 1_2 \end{pmatrix} \sigma_A^2 \begin{pmatrix} \delta_1 & 1_2' \end{pmatrix} + \begin{pmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_{u_2}^2 & 0 \\ 0 & 0 & \sigma_{u_3}^2 \end{pmatrix}\right] B(\theta)^{-1\prime}, \qquad (7)$$

with $\sigma_A^2 = \mathrm{Var}(A)$, $\sigma_v^2 = \mathrm{Var}(V)$, and $\sigma_{u_t}^2 = \mathrm{Var}(U_t)$. With general T, we have

$$\Sigma = \mathrm{Cov}(Y) = B(\theta)^{-1}\left[\begin{pmatrix} \delta_1 \\ 1_{T-1} \end{pmatrix} \begin{pmatrix} \delta_1 & 1_{T-1}' \end{pmatrix} \sigma_A^2 + \mathrm{diag}(\sigma_v^2, \sigma_{u_2}^2, \ldots, \sigma_{u_T}^2)\right] B(\theta)^{-1\prime}, \quad (8)$$

with

$$B(\theta)^{-1} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \theta & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \theta^{T-1} & \theta^{T-2} & \ldots & 1 \end{pmatrix}.$$

4. MINIMUM DISTANCE ESTIMATION

With $T = 3$, let σ be a 6×1 matrix formed from the distinct elements of the symmetric 3×3 matrix Σ:

$$\sigma' = (\,\sigma_{11} \quad \sigma_{21} \quad \sigma_{31} \quad \sigma_{22} \quad \sigma_{32} \quad \sigma_{33}\,)\,.$$

Define the 6×1 matrix of parameters α:

$$\alpha' = (\,\theta \quad \delta_1 \quad \sigma_A^2 \quad \sigma_v^2 \quad \sigma_{u_2}^2 \quad \sigma_{u_3}^2\,)\,.$$

Equation (7) allows us to express σ as a known function $g(\cdot)$ of α:

$$\sigma = g(\alpha).$$

Use the sample covariances $\hat{\sigma}_{st}$ in (6) to obtain the estimate $\hat{\sigma}$. Then obtain an estimate $\hat{\alpha}$ by solving

$$\hat{\alpha} = \arg\min_{\alpha} ||\hat{\sigma} - g(\alpha)||^2.$$

Since $\hat{\sigma}$ are α are both 6×1, we expect to be able to find an $\hat{\alpha}$ that gives a perfect fit: $\hat{\sigma} = g(\hat{\alpha})$.

With general T, let σ be a $T(T+1)/2 \times 1$ matrix formed from the distinct elements of the symmetric $T \times T$ matrix Σ:

$$\sigma' = (\,\sigma_{11} \quad \ldots \quad \sigma_{T1} \quad \sigma_{22} \quad \ldots \quad \sigma_{T2} \quad \ldots \quad \sigma_{TT}\,)\,.$$

Define the $(T+3) \times 1$ matrix of parameters α:

$$\alpha' = (\,\theta \quad \delta_1 \quad \sigma_A^2 \quad \sigma_v^2 \quad \sigma_{u_2}^2 \quad \ldots \quad \sigma_{u_T}^2\,)\,.$$

Equation (8) allows us to express σ as a known function $g(\cdot)$ of α:

$$\sigma = g(\alpha). \tag{9}$$

Note that the parameters in α are unrestricted. When $T = 4$, there are 10 parameters in σ and only 7 parameters in α. So the structural regression model in equation (1) imposes restrictions on σ, and the minimum distance estimator provides a method for imposing those restrictions. As in Section 4 of Note 4, we are free to choose a positive definite, symmetric weight matrix C:

$$\hat{\alpha} = \arg\min_{\alpha} ||\hat{\sigma} - g(\alpha)||^2 = \arg\min_{\alpha} (\hat{\sigma} - g(\alpha))' C (\hat{\sigma} - g(\alpha)). \tag{10}$$

Later in the course we shall discuss an optimal choice for C; for now, we can use an identity matrix.

Once we have obtained an estimate of θ from $\hat{\alpha}$, we can use

$$E(Y) = B(\theta)^{-1} \begin{pmatrix} \delta_0 \\ \lambda_2 \\ \vdots \\ \lambda_T \end{pmatrix}$$

to obtain estimates of δ_0 and the λ_t:

$$\begin{pmatrix} \hat{\delta}_0 \\ \hat{\lambda}_2 \\ \vdots \\ \hat{\lambda}_T \end{pmatrix} = B(\hat{\theta}) \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_T \end{pmatrix}.$$

5. IDENTIFICATION AND EXPLICIT SOLUTIONS

We shall say that θ is identified if we can solve for θ given the population distribution of observed variables. The population distribution of Y determines $\Sigma = \mathrm{Cov}(Y)$, and so θ is identified if we can solve for θ from Σ.

Consider the differencing transformation:

$$Y_3 - Y_2 = \lambda_3 - \lambda_2 + \theta(Y_2 - Y_1) + U_3 - U_2.$$

This implies that

$$\mathrm{Cov}(Y_3 - Y_2 - \theta(Y_2 - Y_1), Y_1) = \mathrm{Cov}(\lambda_3 - \lambda_2 + U_3 - U_2, Y_1) = 0. \tag{11}$$

We are using

$$\mathrm{Cov}(U_3, Y_1) = \mathrm{Cov}(U_2, Y_1) = 0,$$

which follows from the definition of U_t as a prediction error. We can obtain from (11) an equation involving Σ and θ:

$$\sigma_{31} - \sigma_{21} - \theta(\sigma_{21} - \sigma_{11}) = 0.$$

So

$$\theta = \frac{\sigma_{31} - \sigma_{21}}{\sigma_{21} - \sigma_{11}}, \tag{12}$$

provided that $\sigma_{21} \neq \sigma_{11}$. Equation (12) suggests the following explicit estimator for θ:

$$\hat{\theta} = \frac{\hat{\sigma}_{31} - \hat{\sigma}_{21}}{\hat{\sigma}_{21} - \hat{\sigma}_{11}}.$$

For general T, the differencing transformation gives

$$Y_t - Y_{t-1} = \lambda_t - \lambda_{t-1} + \theta(Y_{t-1} - Y_{t-2}) + U_t - U_{t-1},$$

which implies that

$$\mathrm{Cov}(Y_t - Y_{t-1} - \theta(Y_{t-1} - Y_{t-2}), Y_{t-j}) = \mathrm{Cov}(\lambda_t - \lambda_{t-1} + U_t - U_{t-1}, Y_{t-j}) = 0$$

if $j \geq 2$. So we have

$$\sigma_{t,t-j} - \sigma_{t-1,t-j} - \theta(\sigma_{t-1,t-j} - \sigma_{t-2,t-j}) = 0,$$

and

$$\theta = \frac{\sigma_{t,t-j} - \sigma_{t-1,t-j}}{\sigma_{t-1,t-j} - \sigma_{t-2,t-j}} \qquad (t = 3, \ldots, T;\ j = 2, \ldots, t-1). \tag{13}$$

With $T = 4$, equation (13) gives three solutions for θ, from $j = 2$ with $t = 3$ and $j = 2, 3$ with $t = 4$.

There are additional solutions for θ that involve quadratic equations. Because

$$Y_t - \theta Y_{t-1} = \lambda_t + A + U_t,$$

we have

$$\mathrm{Cov}(Y_t - \theta Y_{t-1}, Y_s - \theta Y_{s-1}) = \sigma_A^2$$

if $t \neq s$. So

$$\mathrm{Cov}(Y_3 - \theta Y_2, Y_2 - \theta Y_1) = \sigma_A^2 = \mathrm{Cov}(Y_4 - \theta Y_3, Y_3 - \theta Y_2).$$

This implies that

$$\sigma_{32} - \theta\sigma_{22} - \theta\sigma_{31} + \theta^2\sigma_{21} = \sigma_{43} - \theta\sigma_{33} - \theta\sigma_{42} + \theta^2\sigma_{32},$$

which gives a quadratic equation to solve for θ.

An advantage of the minimum distance estimator in (10) is that it imposes all of the restrictions on Σ and so makes use of all the ways in which we can solve for θ from Σ.