LECTURE NOTE 4

PANEL DATA

Consider a population of families. Choose one at random. For each family member t (or for a subset of the family members), there is an outcome variable Y_t and a predictor variable Z_t. There are also variables W and A whose values are the same for all the family members. 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, Z_t, and W, but A is not observed. Our objective is to measure a (predictive) effect of Z_t on Y_t, holding constant W and A. We shall develop assumptions that, combined with the family structure of the data, make this feasible.

Consider a population of firms. Choose one at random. For each year from $t = 1, \ldots, T$ there is an output variable Y_t and an input variable Z_t. There is also an input variable A that does not vary over time (but does vary across firms). 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 and Z_t, but A is not observed. Our objective is to measure a (predictive) effect of Z_t on Y_t holding A constant.

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.

These three examples have much in common, and I shall refer to them as panel data.

The last two examples are a special case called longitudinal data.

1. REGRESSION SYSTEMS

We shall work with the random variables

$$(Y_1, \dots, Y_T, Z_1, \dots, Z_T, W, A),$$

which have a joint distribution. For example, we could have a randomly drawn family, from a population of families that have T siblings. The siblings are indexed by $t = 1, \dots, T$; Y_t is the (adult) earnings of sibling t, Z_t is the education of sibling t, and W is parents' income. The unobserved variable A could be some other measure of family background, such as parents' education.

Consider the regression function for Y_t given $Z_1, \dots, Z_T, W, A$. We shall assume that Z_s is relevant only for $s = t$:

$$E(Y_t \mid Z_1, \dots, Z_T, W, A) = g_t(Z_t, W, A),$$

so we have exclusion restrictions on Z_s for $s \neq t$. In addition, we shall impose the following functional form restriction:

$$E(Y_t \mid Z_1, \dots, Z_T, W, A) = \theta_{0t} + \theta_{1t} Z_t + \theta_{2t} W + \theta_{3t} A,$$

so that the variables enter in a simple linear fashion, with no interactions or higher order polynomial terms. We shall begin with the case in which the coefficients do not depend upon t:

$$E(Y_t \mid Z_1, \dots, Z_T, W, A) = \theta_0 + \theta_1 Z_t + \theta_2 W + \theta_3 A, \tag{1}$$

and relax that restriction later.

We shall refer to the regression function in equation (1) as a structural regression function. Here "structural" just means that the coefficients in (1) are of direct interest. In particular, θ_1 is the partial (predictive) effect of education on earnings, holding constant

W and A. This regression function does not have a sample counterpart, since A is not observed. So we shall develop linear predictors that do have sample counterparts. Define

$$X' = (\, Z_1 \quad \ldots \quad Z_T \quad 1 \quad W \quad X_{T+2} \quad \ldots \quad X_K \,),$$

where $X_{T+2}, \ldots, X_K$ can include functions (such as polynomials) of the observed variables $Z_1, \ldots, Z_T, W$. The linear predictor of Y_t given X is

$$E^*(Y \mid X) = \theta_0 + \theta_1 Z_t + \theta_2 W + \theta_3 E^*(A \mid X).$$

Our notation for the linear predictor of A given X is

$$E^*(A \mid X) = \gamma_1 X_1 + \cdots + \gamma_K X_K.$$

To see how this works, suppose that $T = 2$:

$$E^*(Y_1 \mid X) = (\theta_1 + \theta_3\gamma_1)Z_1 + \theta_3\gamma_2 Z_2 + R, \tag{2}$$

$$E^*(Y_2 \mid X) = \theta_3\gamma_1 Z_1 + (\theta_1 + \theta_3\gamma_2)Z_2 + R, \tag{3}$$

with

$$R = (\theta_0 + \theta_3\gamma_3) + (\theta_2 + \theta_3\gamma_4)W + \theta_3\gamma_5 X_5 + \cdots + \theta_3\gamma_K X_K.$$

If we look only at the coefficients in the Y_1 predictor, then we cannot identify θ_1. But if we use the Y_1 and Y_2 predictors together, then can obtain θ_1 by subtracting the Z_1 coefficient in predicting Y_2 from the Z_1 coefficient in predicting Y_1. We can also obtain θ_1 by subtracting the Z_2 coefficient in predicting Y_1 from the Z_2 coefficient in predicting Y_2. This is the key to our approach: work with the system of linear predictors for Y_t. (If these were conditional expectation functions, they would be a regression system or multivariate regression.)

The coefficient θ_2 on W in the structural regression function only appears in (2) and (3) in the term $(\theta_2 + \theta_3\gamma_4)$. Since γ_4 does not appear anywhere else, we cannot identify θ_2.

Likewise, we cannot identify θ_0. Also, γ_t only appears multiplied by θ_3. So we can only identify $\theta_3\gamma_1$, $\theta_3\gamma_2$, and $\theta_3\gamma_5, \ldots, \theta_3\gamma_K$.

With general T, we can simplify notation by defining

$$\lambda' \equiv (\,\theta_3\gamma_1 \quad \ldots \quad \theta_3\gamma_T \quad (\theta_0 + \theta_3\gamma_{T+1}) \quad (\theta_2 + \theta_3\gamma_{T+2}) \quad \theta_3\gamma_{T+3} \quad \ldots \quad \theta_3\gamma_K\,)\,.$$

Define

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix}, \quad 1_T = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

(1_T is a $T \times 1$ matrix of ones.) Then

$$E^*(Y \mid X) = \begin{pmatrix} E^*(Y_1 \mid X) \\ \vdots \\ E^*(Y_T \mid X) \end{pmatrix} = \Pi X,$$

with

$$\Pi = (\,\theta_1 I_T \quad 0\,) + 1_T\lambda'. \tag{4}$$

$E^*(Y \mid X)$ is our notation for the system of linear predictors (or multivariate linear predictor). The predictor coefficients are arranged in the $T \times K$ matrix Π. (I_T is our notation for the $T \times T$ identity matrix.) Define

$$\alpha' \equiv (\,\theta_1 \quad \lambda_1 \quad \ldots \quad \lambda_T\,)\,.$$

Note that α is unrestricted; there are no restrictions connecting θ_1 and the λ_t. Equation (4) expresses the $T \cdot K$ elements of Π in terms of the $K+1$ elements of $\alpha' = (\,\theta_1 \quad \lambda'\,)$.

With $T = 3$, we have

$$\Pi = \begin{pmatrix} \theta_1 + \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \ldots & \lambda_K \\ \lambda_1 & \theta_1 + \lambda_2 & \lambda_3 & \lambda_4 & \ldots & \lambda_K \\ \lambda_1 & \lambda_2 & \theta_1 + \lambda_3 & \lambda_4 & \ldots & \lambda_K \end{pmatrix}. \tag{5}$$

2. DIFFERENCING TRANSFORMATIONS

We can obtain a simpler system of linear predictors by applying a differencing transformation. The key is a matrix D such that

$$D 1_T = 0.$$

Then

$$E^*(DY \mid X) = DE^*(Y \mid X) = D\Pi X = \theta_1 DZ, \tag{6}$$

where

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_T \end{pmatrix}.$$

For example, let D be the $(T-1) \times T$ matrix

$$D = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix}.$$

Then DY and DZ give first differences:

$$DY = \begin{pmatrix} Y_2 - Y_1 \\ \vdots \\ Y_T - Y_{T-1} \end{pmatrix}, \quad DZ = \begin{pmatrix} Z_2 - Z_1 \\ \vdots \\ Z_T - Z_{T-1} \end{pmatrix},$$

and (6) gives

$$E^*(Y_t - Y_{t-1} \mid X) = \theta_1 (Z_t - Z_{t-1}) \qquad (t = 2, \dots, T).$$

For a second example, let D be the $T \times T$ matrix

$$D = I_t - \frac{1}{T} 1_T 1_T'.$$

Then DY and DZ give deviations from the means:

$$DY = Y - \bar{Y} 1_T = \begin{pmatrix} Y_1 - \bar{Y} \\ \vdots \\ Y_T - \bar{Y} \end{pmatrix}, \quad DZ = Z - \bar{Z} 1_T = \begin{pmatrix} Z_1 - \bar{Z} \\ \vdots \\ Z_T - \bar{Z} \end{pmatrix},$$

with $\bar{Y} = \sum_{t=1}^{T} Y_t/T$ and $\bar{Z} = \sum_{t=1}^{T} Z_t/T$. Equation (6) gives

$$E^*(Y_t - \bar{Y} \mid X) = \theta_1(Z_t - \bar{Z}) \qquad (t = 1, \ldots, T).$$

3. IMPOSING RESTRICTIONS

Go back to equation (4). This expresses the $T \cdot K$ elements of Π in terms of the $K+1$ elements of $\alpha' = (\,\theta_1 \quad \lambda'\,)$. So there are restrictions on Π, as we can see in the display in (5). We are going to express the elements of Π as a linear function of α. This leads in the next section to a minimum distance estimator for imposing the restrictions on sample data.

The transpose of Π is the $K \times T$ matrix

$$\Pi' = \theta_1 \begin{pmatrix} I_T \\ 0 \end{pmatrix} + \lambda 1_T' = \theta_1\,(\,e_1 \quad \ldots \quad e_T\,) + \lambda 1_T',$$

where e_t is a $K \times 1$ matrix of zeros except for a one in row t. Let π be the $K \cdot T \times 1$ matrix formed by stacking the columns of Π':

$$\pi = \mathrm{stack}(\Pi') = \begin{pmatrix} \theta_1 e_1 + \lambda \\ \vdots \\ \theta_1 e_T + \lambda \end{pmatrix}.$$

Now we can express π as a linear function of α:

$$\pi = G \begin{pmatrix} \theta_1 \\ \lambda \end{pmatrix} = G\alpha,$$

where G is the $K \cdot T \times (K+1)$ matrix

$$G = \begin{pmatrix} e_1 & I_K \\ \vdots & \vdots \\ e_T & I_K \end{pmatrix}.$$

The restriction on π is that it is a linear combination of the columns of the known matrix G. So π is restricted to lie in the linear subspace generated by the columns of G; this is a known (given) subspace since G is known (given).

4. MINIMUM DISTANCE ESTIMATION AND GENERALIZED LEAST SQUARES

The data from a sample of $i = 1, \ldots, N$ families can be put in the matrices

$$y_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{pmatrix}, \quad z_t = \begin{pmatrix} z_{1t} \\ \vdots \\ z_{Nt} \end{pmatrix}, \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix} \qquad (t = 1, \ldots, T).$$

We use least squares to form an estimate $\hat{\Pi}$. A least-squares fit of y_t on $z_1, \ldots, z_T, 1, w$ gives the coefficients in row t of Π. (We could also construct data matrices $x_{T+3}, \ldots, x_K$, based on functions of z and w; then these would be included in the least-squares fit.)

Form $\hat{\pi}$ by stacking the columns of $\hat{\Pi}'$:

$$\hat{\pi} = \mathrm{stack}(\hat{\Pi}').$$

Recall that $\alpha' = (\,\theta_1 \quad \lambda'\,)$ and $\pi = G\alpha$. The minimum distance estimate of α is

$$\hat{\alpha} = \arg\min_{\alpha} ||\hat{\pi} - G\alpha||^2.$$

The motivation for this estimator is that the least-squares estimate $\hat{\pi}$ corresponds to the population π but does not impose the restriction that π is a linear combination of the columns of G. So we find the linear combination of the columns of G that gives the best fit to $\hat{\pi}$. The norm in the distance criterion corresponds to the inner product

$$\langle a, b\rangle = a'Cb,$$

where C is a positive definite, symmetric matrix. (A positive definite matrix C is a square matrix, say $J \times J$, such that if a is any nonzero $J \times 1$ matrix, then $a'Ca > 0$; C symmetric means that $C' = C$.) So

$$||\hat{\pi} - G\alpha||^2 = (\hat{\pi} - G\alpha)'C(\hat{\pi} - G\alpha).$$

There is a positive definite, symmetric matrix $C^{1/2}$ that provides a square root of C:

$$C^{1/2}C^{1/2} = C$$

(based on the spectral decomposition of C, from linear algebra). Define

$$\hat{\pi}^* = C^{1/2}\hat{\pi}, \quad G^* = C^{1/2}G.$$

Then $\hat{\alpha}$ can be obtained from a least-squares fit of $\hat{\pi}^*$ on G^* (and we use the matrix version from Section 3 of Note 3):

$$\begin{aligned}
\hat{\alpha} &= \arg\min_{\alpha}(\hat{\pi}^* - G^*\alpha)'(\hat{\pi}^* - G^*\alpha) \\
&= (G^{*\prime}G^*)^{-1}G^{*\prime}\hat{\pi}^* \\
&= (G'CG)^{-1}G'C\hat{\pi}. \qquad (7)
\end{aligned}$$

The expression for $\hat{\alpha}$ in (7) is known as *generalized least squares*.

We are free to choose the weight matrix C, and this makes the minimum distance estimator more flexible. If some components of $\hat{\pi}$ are estimated more precisely than others, then we may want to give more weight to those components. We will discuss an optimal choice for C in the inference part of the course, where we work out the distribution of $\hat{\pi}$ in repeated samples. For now, we can just set C equal to an identity matrix.

5. STACKING

There is another way to impose restrictions, in which the data matrices are stacked and then used in a least-squares fit. I'll use the difference transformations from Section 2 to illustrate. Define

$$\tilde{Y}_t = Y_t - Y_{t-1}, \quad \tilde{Z}_t = Z_t - Z_{t-1} \qquad (t = 2, \ldots, T).$$

Set up the corresponding data matrices

$$\tilde{y}_t = \begin{pmatrix} y_{1t} - y_{1,t-1} \\ \vdots \\ y_{Nt} - y_{N,t-1} \end{pmatrix}, \quad \tilde{z}_t = \begin{pmatrix} z_{1t} - z_{1,t-1} \\ \vdots \\ z_{Nt} - z_{N,t-1} \end{pmatrix} \qquad (t = 2, \ldots, T).$$

We have

$$E^*(\tilde{Y}_t \mid \tilde{Z}_t) = \pi_t \tilde{Z}_t,$$

with $\pi_t = \theta_1$ $(t = 2, \ldots, T)$. The sample counterpart is to obtain an estimate $\hat{\pi}_t$ from the least-squares fit of $\tilde{y}_t$ on $\tilde{z}_t$. Define

$$\pi = \begin{pmatrix} \pi_2 \\ \vdots \\ \pi_T \end{pmatrix}, \quad \hat{\pi} = \begin{pmatrix} \hat{\pi}_2 \\ \vdots \\ \hat{\pi}_T \end{pmatrix}.$$

Then

$$\pi = G\theta_1 \quad \text{with} \quad G = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1_{T-1},$$

and we can obtain a minimum distance estimate of θ_1 from

$$\hat{\theta}_1 = \arg\min_{\theta_1} ||\hat{\pi} - G\theta_1||^2 = (G'CG)^{-1}G'C\hat{\pi}. \tag{8}$$

Now stack up the data matrices as follows:

$$\tilde{y} = \begin{pmatrix} \tilde{y}_2 \\ \vdots \\ \tilde{y}_T \end{pmatrix}, \quad \tilde{z} = \begin{pmatrix} \tilde{z}_2 \\ \vdots \\ \tilde{z}_T \end{pmatrix}.$$

We can obtain an estimate of θ_1 from a least-squares fit of $\tilde{y}$ on $\tilde{z}$:

$$\hat{\theta}_1 = (\tilde{z}'\tilde{z})^{-1}\tilde{z}'\tilde{y} = \left(\sum_{t=2}^{T} \tilde{z}_t'\tilde{z}_t\right)^{-1} \sum_{t=2}^{T} \tilde{z}_t'\tilde{y}_t. \tag{9}$$

This stacked estimate equals the minimum distance estimate for a particular choice of the weight matrix C. Let

$$C = \operatorname{diag}(\tilde{z}_2'\tilde{z}_2, \ldots, \tilde{z}_T'\tilde{z}_T) = \begin{pmatrix} \tilde{z}_2'\tilde{z}_2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \tilde{z}_T'\tilde{z}_T \end{pmatrix}.$$

Because $\hat{\pi}_t = (\tilde{z}_t'\tilde{z}_t)^{-1}\tilde{z}_t'\tilde{y}_t$, we have

$$(G'CG)^{-1}G'C\hat{\pi} = \left(\sum_{t=2}^{T} \tilde{z}_t'\tilde{z}_t\right)^{-1} \sum_{t=2}^{T} \tilde{z}_t'\tilde{y}_t.$$

So for this choice of the weight matrix C, the minimum distance estimate in (8) equals the stacked estimate in (9).

Similar points apply with the deviations from means transformation. Now let

$$\tilde{Y}_t = Y_t - \bar{Y}, \quad \tilde{Z}_t = Z_t - \bar{Z} \qquad (t = 1, \ldots, T),$$

and set up the corresponding data matrices

$$\tilde{y}_t = \begin{pmatrix} y_{1t} - \bar{y}_1 \\ \vdots \\ y_{Nt} - \bar{y}_N \end{pmatrix}, \quad \tilde{z}_t = \begin{pmatrix} z_{1t} - \bar{z}_1 \\ \vdots \\ z_{Nt} - \bar{z}_N \end{pmatrix},$$

with

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{z}_i = \frac{1}{T}\sum_{t=1}^{T} z_{it} \qquad (i = 1, \ldots, N).$$

We have

$$E^*(\tilde{Y}_t \mid \tilde{Z}_t) = \pi_t \tilde{Z}_t,$$

with $\pi_t = \theta_1$ $(t = 1, \ldots, T)$. The sample counterpart is to obtain an estimate $\hat{\pi}_t$ from the least-squares fit of $\tilde{y}_t$ on $\tilde{z}_t$. Define

$$\hat{\pi} = \begin{pmatrix} \hat{\pi}_1 \\ \vdots \\ \hat{\pi}_T \end{pmatrix}.$$

Stack up the data matrices:

$$\tilde{y} = \begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_T \end{pmatrix}, \quad \tilde{z} = \begin{pmatrix} \tilde{z}_1 \\ \vdots \\ \tilde{z}_T \end{pmatrix}.$$

We can obtain an estimate of θ_1 from a least-squares fit of $\tilde{y}$ on $\tilde{z}$:

$$\hat{\theta}_1 = (\tilde{z}'\tilde{z})^{-1}\tilde{z}'\tilde{y} = \left(\sum_{t=1}^{T} \tilde{z}_t'\tilde{z}_t\right)^{-1} \sum_{t=1}^{T} \tilde{z}_t'\tilde{y}_t. \tag{10}$$

This stacked estimate equals the minimum distance estimate for a particular choice of the weight matrix C. Let

$$C = \operatorname{diag}(\tilde{z}_1'\tilde{z}_1, \dots, \tilde{z}_T'\tilde{z}_T) = \begin{pmatrix} \tilde{z}_1'\tilde{z}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \tilde{z}_T'\tilde{z}_T \end{pmatrix}.$$

Because $\hat{\pi}_t = (\tilde{z}_t'\tilde{z}_t)^{-1}\tilde{z}_t'\tilde{y}_t$, we have

$$(G'CG)^{-1}G'C\hat{\pi} = \left(\sum_{t=1}^{T} \tilde{z}_t'\tilde{z}_t\right)^{-1} \sum_{t=1}^{T} \tilde{z}_t'\tilde{y}_t.$$

So for this choice of the weight matrix C, the minimum distance estimate equals the stacked estimate in (10).

In general, the stacked estimate of θ_1 based on the first difference transformation is not equal to the stacked estimate based on the deviations from means transformation. (They are equal if $T = 2$.) The choice between these estimates depends on their sampling distributions and will be discussed in the inference part of the course.

6. PRODUCTION FUNCTION

For a randomly chosen farm, Q_t is output in year t, L_t is labor input in year t, F is a measure of soil quality and other location aspects that are not changing over time, and V_t is a measure of rainfall and other weather conditions in year t. The production function is

$$Q_t = L_t^{\theta} F V_t \qquad (t = 1, \dots, T)$$

with $0 < \theta < 1$. Data are available on output and labor input for N of these farms over T years; data on soil quality and weather conditions are not available. The farmer's objective is to choose the labor input to maximize the conditional expectation of profit, conditional on the information $\mathcal{J}_t$ available to him when the labor choice is made:

$$\max_L E[P_t Q_t - W_t L \mid \mathcal{J}_t].$$

The price of output (P_t) and the price of labor (W_t) are not affected by the farmer's choice. The first-order condition for

$$\max_L P_t[L^\theta F E(V_t \mid \mathcal{J}_t)] - W_t L$$

gives

$$\theta P_t L^{\theta-1} F E(V_t \mid J_t) = W_t.$$

The derived demand for labor is

$$\log L_t = \frac{1}{1-\theta}[\log\theta - \log\frac{W_t}{P_t} + \log F + \log E(V_t \mid \mathcal{J}_t)].$$

We can write the production function as

$$\log Q_t = \theta \log L_t + \log F + \log V_t,$$

or

$$Y_t = \theta Z_t + A + \log V_t,$$

with $Y_t = \log Q_t$, $Z_t = \log L_t$, and $A = \log F$. Note that A is correlated with Z_t through the derived demand for labor, and so a regression function for Y_t that does not include A will not have the production function elasticity θ as the coefficient on Z_t. This is the omitted variable bias motivation for the use of panel data.

Note that

$$E(Y_t \mid Z_1, \ldots, Z_T, A) = \theta Z_t + A + E(\log V_t \mid Z_1, \ldots, Z_T, A).$$

Our structural regression function in equation (1) in Section 1 has the form

$$E(Y_t \mid Z_1, \ldots, Z_T, A) = \theta Z_t + A + \text{constant}.$$

So to apply the results from Section 1, with θ interpreted as the production function elasticity, we need

$$E(\log V_t \mid Z_1, \ldots, Z_T, A) = \text{constant}.$$

This could fail to hold if there is correlation over time in the weather conditions V_t. For then lagged values of $\log V_t$ will enter the demand for labor, and so Z_{t+1} will be correlated with $\log V_t$.

7. TIME-VARYING COEFFICIENTS

Return to the structural regression function in equation (1) of Section 1, and allow for time-varying coefficients:

$$E(Y_t \mid Z_1, \ldots, Z_T, W, A) = \theta_{0t} + \theta_{1t} Z_t + \theta_{2t} W + \theta_{3t} A.$$

As before, let

$$X' = (\, Z_1 \quad \ldots \quad Z_T \quad 1 \quad W \quad X_{T+3} \quad \ldots \quad X_K \,),$$

$$E^*(A \mid X) = \gamma_1 X_1 + \cdots + \gamma_K X_K = \gamma' X.$$

Then the linear predictor of Y_t given X is

$$E^*(Y_t \mid X) = \theta_{0t} + \theta_{1t} Z_t + \theta_{2t} W + \theta_{3t} \gamma' X.$$

The multivariate linear predictor is

$$E^*(Y \mid X) = \Pi X \quad \text{with} \quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix}.$$

The matrix Π of linear predictor coefficients is $T \times K$. With $T = 3$,

$$\Pi = \begin{pmatrix} \theta_{11} + \theta_{31}\gamma_1 & \theta_{31}\gamma_2 & \theta_{31}\gamma_3 & \ldots \\ \theta_{32}\gamma_1 & \theta_{12} + \theta_{32}\gamma_2 & \theta_{32}\gamma_3 & \ldots \\ \theta_{33}\gamma_1 & \theta_{33}\gamma_2 & \theta_{13} + \theta_{33}\gamma_3 & \ldots \end{pmatrix}$$

(where we are displaying only the first three columns of Π). Note that

$$\frac{\pi_{12}}{\pi_{32}} = \frac{\theta_{31}\gamma_2}{\theta_{33}\gamma_2} = \frac{\theta_{31}}{\theta_{33}},$$

$$\pi_{31} = \theta_{33}\gamma_1,$$

and so

$$\pi_{11} - \frac{\pi_{12}}{\pi_{32}}\pi_{31} = \theta_{11}.$$

So θ_{11} is identified (provided that $\pi_{32} \neq 0$), and a similar argument shows that θ_{1t} is identified for $t = 1, 2, 3$.