1

Partitioned Regression and the Frisch–Waugh–Lovell Theorem

This chapter introduces the reader to important background material on the partitioned regression model. This should serve as a refresher for some matrix algebra results on the partitioned regression model as well as an introduction to the associated Frisch–Waugh–Lovell (FWL) theorem. The latter is shown to be a useful tool for proving key results for the fixed effects model in Chapter 2 as well as artificial regressions used in testing panel data models such as the Hausman test in Chapter 4.

Consider the partitioned regression given by

$$y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u \tag{1.1}$$

where y is a column vector of dimension $(n \times 1)$ and X is a matrix of dimension $(n \times k)$. Also, $X = [X_1, X_2]$ with X_1 and X_2 of dimension $(n \times k_1)$ and $(n \times k_2)$, respectively. One may be interested in the least squares estimates of β_2 corresponding to X_2, but one has to control for the presence of X_1 which may include seasonal dummy variables or a time trend; see Frisch and Waugh (1933) and Lovell (1963). For example, in a time-series setting, including the time trend in the multiple regression is equivalent to detrending each variable first, by residualing out the effect of time, and then running the regression on these residuals. Davidson and MacKinnon (1993) denote this result more formally as the FWL theorem.

The ordinary least squares (OLS) normal equations from (1.1) are given by:

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_{1,\mathrm{OLS}} \\ \widehat{\beta}_{2,\mathrm{OLS}} \end{bmatrix} = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} \tag{1.2}$$

Exercise 1.1 *(Partitioned regression).* Show that the solution to (1.2) yields

$$\widehat{\beta}_{2,\mathrm{OLS}} = (X_2'\overline{P}_{X_1}X_2)^{-1}X_2'\overline{P}_{X_1}y \tag{1.3}$$

where $P_{X_1} = X_1(X_1'X_1)^{-1}X_1'$ is the projection matrix on X_1, and $\overline{P}_{X_1} = I_n - P_{X_1}$.

Solution

Write (1.2) as two equations:

$$(X_1'X_1)\widehat{\beta}_{1,\mathrm{OLS}} + (X_1'X_2)\widehat{\beta}_{2,\mathrm{OLS}} = X_1'y$$
$$(X_2'X_1)\widehat{\beta}_{1,\mathrm{OLS}} + (X_2'X_2)\widehat{\beta}_{2,\mathrm{OLS}} = X_2'y$$

Solving for $\widehat{\beta}_{1,\text{OLS}}$ in terms of $\widehat{\beta}_{2,\text{OLS}}$ by multiplying the first equation by $(X_1'X_1)^{-1}$, we get

$$\widehat{\beta}_{1,\text{OLS}} = (X_1'X_1)^{-1}X_1'y - (X_1'X_1)^{-1}X'X_1'X_2\widehat{\beta}_{2,\text{OLS}} = (X_1'X_1)^{-1}X_1'(y - X_2\widehat{\beta}_{2,\text{OLS}})$$

Substituting $\widehat{\beta}_{1,\text{OLS}}$ in the second equation, we get

$$X_2'X_1(X_1'X_1)^{-1}X_1'y - X_2'P_{X_1}X_2\widehat{\beta}_{2,\text{OLS}} + (X_2'X_2)\widehat{\beta}_{2,\text{OLS}} = X_2'y$$

Collecting terms, we get $(X_2'\overline{P}_{X_1}X_2)\widehat{\beta}_{2,\text{OLS}} = X_2'\overline{P}_{X_1}y$. Hence, $\widehat{\beta}_{2,\text{OLS}} = (X_2'\overline{P}_{X_1}X_2)^{-1}X_2'\overline{P}_{X_1}y$ as given in (1.3). $\overline{P}_{X_1}$ is the orthogonal projection matrix of X_1 and $\overline{P}_{X_1}X_2$ generates the least squares residuals of each column of X_2 regressed on all the variables in X_1. In fact, if we write $\widetilde{X}_2 = \overline{P}_{X_1}X_2$ and $\widetilde{y} = \overline{P}_{X_1}y$, then

$$\widehat{\beta}_{2,\text{OLS}} = (\widetilde{X}_2'\widetilde{X}_2)^{-1}\widetilde{X}_2'\widetilde{y} \tag{1.4}$$

using the fact that $\overline{P}_{X_1}$ is idempotent. For a review of idempotent matrices, see Abadir and Magnus (2005, p.231). This implies that $\widehat{\beta}_{2,\text{OLS}}$ can be obtained from the regression of $\widetilde{y}$ on $\widetilde{X}_2$. In words, the residuals from regressing y on X_1 are in turn regressed upon the residuals from each column of X_2 regressed on all the variables in X_1. If we premultiply (1.1) by $\overline{P}_{X_1}$ and use the fact that $\overline{P}_{X_1}X_1 = 0$, we get

$$\overline{P}_{X_1}y = \overline{P}_{X_1}X_2\beta_2 + \overline{P}_{X_1}u \tag{1.5}$$

Exercise 1.2 *(The Frisch–Waugh–Lovell theorem).* Prove that:
(a) the least squares *estimates* of β_2 from equations (1.1) and (1.5) are numerically identical;
(b) the least squares *residuals* from equations (1.1) and (1.5) are identical.

Solution

(a) Using the fact that $\overline{P}_{X_1}$ is idempotent, it immediately follows that OLS on (1.5) yields $\widehat{\beta}_{2,\text{OLS}}$ as given by (1.3). Alternatively, one can start from (1.1) and use the result that

$$y = P_Xy + \overline{P}_Xy = X\widehat{\beta}_{\text{OLS}} + \overline{P}_Xy = X_1\widehat{\beta}_{1,\text{OLS}} + X_2\widehat{\beta}_{2,\text{OLS}} + \overline{P}_Xy \tag{1.6}$$

where $P_X = X(X'X)^{-1}X'$ and $\overline{P}_X = I_n - P_X$. Premultiplying (1.6) by $X_2'\overline{P}_{X_1}$ and using the fact that $\overline{P}_{X_1}X_1 = 0$, one gets

$$X_2'\overline{P}_{X_1}y = X_2'\overline{P}_{X_1}X_2\widehat{\beta}_{2,\text{OLS}} + X_2'\overline{P}_{X_1}\overline{P}_Xy \tag{1.7}$$

But $P_{X_1}P_X = P_{X_1}$. Hence, $\overline{P}_{X_1}\overline{P}_X = \overline{P}_X$. Using this fact along with $\overline{P}_XX = \overline{P}_X[X_1, X_2] = 0$, the last term of (1.7) drops out yielding the result that $\widehat{\beta}_{2,\text{OLS}}$ from (1.7) is identical to the expression in (1.3). Note that no partitioned inversion was used in this proof. This proves part (a) of the FWL theorem. To learn more about partitioned and projection matrices, see Chapter 5 of Abadir and Magnus (2005).

(b) Premultiplying (1.6) by $\overline{P}_{X_1}$ and using the fact that $\overline{P}_{X_1}\overline{P}_X = \overline{P}_X$, one gets

$$\overline{P}_{X_1} y = \overline{P}_{X_1} X_2 \widehat{\beta}_{2,\mathrm{OLS}} + \overline{P}_X y \tag{1.8}$$

Note that $\widehat{\beta}_{2,\mathrm{OLS}}$ was shown to be numerically identical to the least squares estimate obtained from (1.5). Hence, the first term on the right-hand side of (1.8) must be the fitted values from equation (1.5). Since the dependent variables are the same in equations (1.8) and (1.5), $\overline{P}_X y$ in equation (1.8) must be the least squares residuals from regression (1.5). But $\overline{P}_X y$ is the least squares residuals from regression (1.1). Hence, the least squares residuals from regressions (1.1) and (1.5) are numerically identical. This proves part (b) of the FWL theorem. Several applications of the FWL theorem will be given in this book.

Exercise 1.3 *(Residualing the constant).* Show that if X_1 is the vector of ones indicating the presence of a constant in the regression, then regression (1.8) is equivalent to running $(y_i - \overline{y})$ on the set of variables in X_2 expressed as deviations from their respective sample means.

Solution

In this case, $X = [\iota_n, X_2]$ where ι_n is a vector of ones of dimension n. $P_{X_1} = \iota_n(\iota_n'\iota_n)^{-1}\iota_n' = \iota_n\iota_n'/n = J_n/n$, where $J_n = \iota_n\iota_n'$ is a matrix of ones of dimension n. But $J_n y = \sum_{i=1}^n y_i$ and $J_n y/n = \overline{y}$. Hence, $\overline{P}_{X_1} = I_n - P_{X_1} = I_n - J_n/n$ and $\overline{P}_{X_1} y = (I_n - J_n/n)y$ has a typical element $(y_i - \overline{y})$. From the FWL theorem, $\widehat{\beta}_{2,\mathrm{OLS}}$ can be obtained from the regression of $(y_i - \overline{y})$ on the set of variables in X_2 expressed as deviations from their respective means, i.e., $\overline{P}_{X_1} X_2 = (I_n - J_n/n)X_2$. From the solution of Exercise 1.1, we get

$$\begin{aligned}\widehat{\beta}_{1,\mathrm{OLS}} &= (X_1'X_1)^{-1}X_1'(y - X_2\widehat{\beta}_{2,\mathrm{OLS}}) = (\iota_n'\iota_n)^{-1}\iota_n'(y - X_2\widehat{\beta}_{2,\mathrm{OLS}})\\ &= \frac{\iota_n'}{n}(y - X_2\widehat{\beta}_{2,\mathrm{OLS}}) = \overline{y} - \overline{X}_2'\widehat{\beta}_{2,\mathrm{OLS}}\end{aligned}$$

where $\overline{X}_2' = \iota_n'X_2/n$ is the vector of sample means of the independent variables in X_2.

Exercise 1.4 *(Adding a dummy variable for the ith observation).* Show that including a dummy variable for the ith observation in the regression is equivalent to omitting that observation from the regression. Let $y = X\beta + D_i\gamma + u$, where y is $n \times 1$, X is $n \times k$ and D_i is a dummy variable that takes the value 1 for the ith observation and 0 otherwise. Using the FWL theorem, prove that the least squares estimates of β and γ from this regression are $\widehat{\beta}_{\mathrm{OLS}} = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$ and $\widehat{\gamma}_{\mathrm{OLS}} = y_i - x_i'\widehat{\beta}_{\mathrm{OLS}}$, where X^* denotes the X matrix without the ith observation, y^* is the y vector without the ith observation and (y_i, x_i') denotes the ith observation on the dependent and independent variables. Note that $\widehat{\gamma}_{\mathrm{OLS}}$ is the forecasted OLS residual for the ith observation obtained from the regression of y^* on X^*, the regression which excludes the ith observation.

Solution

The dummy variable for the ith observation is an $n \times 1$ vector $D_i = (0, 0, \ldots, 1, 0, \ldots, 0)'$ of zeros except for the ith element which takes the value 1. In this case,

$P_{D_i} = D_i(D_i'D_i)^{-1}D_i' = D_iD_i'$ which is a matrix of zeros except for the ith diagonal element which takes the value 1. Hence, $I_n - P_{D_i}$ is an identity matrix except for the ith diagonal element which takes the value zero. Therefore, $(I_n - P_{D_i})y$ returns the vector y except for the ith element which is zero. Using the FWL theorem, the OLS regression

$$y = X\beta + D_i\gamma + u$$

yields the same estimates as $(I_n - P_{D_i})y = (I_n - P_{D_i})X\beta + (I_n - P_{D_i})u$ which can be rewritten as $\widetilde{y} = \widetilde{X}\beta + \widetilde{u}$ with $\widetilde{y} = (I_n - P_{D_i})y$, $\widetilde{X} = (I_n - P_{D_i})X$. The OLS normal equations yield $(\widetilde{X}'\widetilde{X})\widehat{\beta}_{\text{OLS}} = \widetilde{X}'\widetilde{y}$ and the ith OLS normal equation can be ignored since it gives $0'\widehat{\beta}_{\text{OLS}} = 0$. Ignoring the ith observation equation yields $(X^{*\prime}X^*)\widehat{\beta}_{\text{OLS}} = X^{*\prime}y^*$, where X^* is the matrix X without the ith observation and y^* is the vector y without the ith observation. The FWL theorem also states that the residuals from $\widetilde{y}$ on $\widetilde{X}$ are the same as those from y on X and D_i. For the ith observation, $\widetilde{y}_i = 0$ and $\widetilde{x}_i = 0$. Hence the ith residual must be zero. This also means that the ith residual in the original regression with the dummy variable D_i is zero, i.e., $y_i - x_i'\widehat{\beta}_{\text{OLS}} - \widehat{\gamma}_{\text{OLS}} = 0$. Rearranging terms, we get $\widehat{\gamma}_{\text{OLS}} = y_i - x_i'\widehat{\beta}_{\text{OLS}}$. In other words, $\widehat{\gamma}_{\text{OLS}}$ is the forecasted OLS residual for the ith observation from the regression of y^* on X^*. The ith observation was excluded from the estimation of $\widehat{\beta}_{\text{OLS}}$ by the inclusion of the dummy variable D_i.

The results of Exercise 1.4 can be generalized to including dummy variables for several observations. In fact, Salkever (1976) suggested a simple way of using dummy variables to compute forecasts and their standard errors. The basic idea is to augment the usual regression in (1.1) with a matrix of observation-specific dummies, i.e., a dummy variable for each period where we want to forecast:

$$\begin{bmatrix} y \\ y_o \end{bmatrix} = \begin{bmatrix} X & 0 \\ X_o & I_{T_o} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} u \\ u_o \end{bmatrix} \tag{1.9}$$

or

$$y^* = X^*\delta + u^* \tag{1.10}$$

where $\delta' = (\beta', \gamma')$. X^* has in its second part a matrix of dummy variables, one for each of the T_o periods for which we are forecasting.

Exercise 1.5 *(Computing forecasts and forecast standard errors)*

(a) Show that OLS on (1.9) yields $\widehat{\delta}' = (\widehat{\beta}', \widehat{\gamma}')$, where $\widehat{\beta} = (X'X)^{-1}X'y$, $\widehat{\gamma} = y_o - \widehat{y}_o$, and $\widehat{y}_o = X_o\widehat{\beta}$. In other words, OLS on (1.9) yields the OLS estimate of β without the T_o observations, and the coefficients of the T_o dummies, i.e., $\widehat{\gamma}$, are the forecast errors.
(b) Show that the first n residuals are the usual OLS residuals $e = y - X\widehat{\beta}$ based on the first n observations, whereas the next T_o residuals are all zero. Conclude that the mean square error of the regression in (1.10), s^{*2}, is the same as s^2 from the regression of y on X.

(c) Show that the variance–covariance matrix of $\widehat{\delta}$ is given by

$$s^2(X^{*\prime}X^*)^{-1} = s^2\begin{bmatrix}(X'X)^{-1} & \\ & [I_{T_o} + X_o(X'X)^{-1}X_o']\end{bmatrix} \tag{1.11}$$

where the off-diagonal elements are of no interest. This means that the regression package gives the estimated variance of $\widehat{\beta}$ and the estimated variance of the forecast error in one stroke.

(d) Show that if the forecasts rather than the forecast errors are needed, one can replace y_o by zero, and I_{T_o} by $-I_{T_o}$ in (1.9). The resulting estimate of γ will be $\widehat{y}_o = X_o\widehat{\beta}$, as required. The variance of this forecast will be the same as that given in (1.11).

Solution

(a) From (1.9) one gets

$$X^{*\prime}X^* = \begin{bmatrix}X' & X_o' \\ 0 & I_{T_o}\end{bmatrix}\begin{bmatrix}X & 0 \\ X_o & I_{T_o}\end{bmatrix} = \begin{bmatrix}X'X + X_o'X_o & X_o' \\ X_o & I_{T_o}\end{bmatrix}$$

and

$$X^{*\prime}y^* = \begin{bmatrix}X'y + X_o'y_o \\ y_o\end{bmatrix}$$

The OLS normal equations yield

$$X^{*\prime}X^*\begin{bmatrix}\widehat{\beta}_{\text{OLS}} \\ \widehat{\gamma}_{\text{OLS}}\end{bmatrix} = X^{*\prime}y^*$$

or $(X'X)\widehat{\beta}_{\text{OLS}} + (X_o'X_o)\widehat{\beta}_{\text{OLS}} + X_o'\widehat{\gamma}_{\text{OLS}} = X'y + X_o'y_o$ and $X_o\widehat{\beta}_{\text{OLS}} + \widehat{\gamma}_{\text{OLS}} = y_o$. From the second equation, it is obvious that $\widehat{\gamma}_{\text{OLS}} = y_o - X_o\widehat{\beta}_{\text{OLS}}$. Substituting this in the first equation yields

$$(X'X)\widehat{\beta}_{\text{OLS}} + (X_o'X_o)\widehat{\beta}_{\text{OLS}} + X_o'y_o - X_o'X_o\widehat{\beta}_{\text{OLS}} = X'y + X_o'y_o$$

which upon cancellation gives $\widehat{\beta}_{\text{OLS}} = (X'X)^{-1}X'y$. Alternatively, one could apply the FWL theorem using $X_1 = \begin{bmatrix}X \\ X_o\end{bmatrix}$ and $X_2 = \begin{bmatrix}0 \\ I_{T_o}\end{bmatrix}$. In this case, $X_2'X_2 = I_{T_o}$ and

$$P_{X_2} = X_2(X_2'X_2)^{-1}X_2' = X_2X_2' = \begin{bmatrix}0 & 0 \\ 0 & I_{T_o}\end{bmatrix}$$

This means that

$$\overline{P}_{X_2} = I_{n+T_o} - P_{X_2} = \begin{bmatrix}I_n & 0 \\ 0 & 0\end{bmatrix}$$

Premultiplying (1.9) by $\overline{P}_{X_2}$ is equivalent to omitting the last T_o observations. The resulting regression is that of y on X, which yields $\widehat{\beta}_{\text{OLS}} = (X'X)^{-1}X'y$ as obtained above.

(b) Premultiplying (1.9) by $\overline{P}_{X_2}$, the last T_o observations yield zero residuals because the observations on both the dependent and independent variables are zero. For this to be true in the original regression, we must have $y_o - X_o\widehat{\beta}_{\mathrm{OLS}} - \widehat{\gamma}_{\mathrm{OLS}} = 0$. This means that $\widehat{\gamma}_{\mathrm{OLS}} = y_o - X_o\widehat{\beta}_{\mathrm{OLS}}$ as required. The OLS residuals of (1.9) yield the usual least squares residuals

$$e_{\mathrm{OLS}} = y - X\widehat{\beta}_{\mathrm{OLS}}$$

for the first n observations and zero residuals for the next T_o observations. This means that $e^{*\prime} = (e'_{\mathrm{OLS}}, 0')$ and $e^{*\prime}e^* = e'_{\mathrm{OLS}}e_{\mathrm{OLS}}$ with the same residual sum of squares. The number of observations in (1.9) are $n + T_o$ and the number of parameters estimated is $k + T_o$. Hence, the new degrees of freedom in (1.9) are $(n + T_o) - (k + T_o) = (n - k) =$ the degrees of freedom in the regression of y on X. Hence, $s^{*2} = e^{*\prime}e^*/(n - k) = e'_{\mathrm{OLS}}e_{\mathrm{OLS}}/(n - k) = s^2$.

(c) Using partitioned inverse formulas on $(X^{*\prime}X^*)$ one gets

$$(X^{*\prime}X^*)^{-1} = \begin{bmatrix} (X'X)^{-1} & -(X'X)^{-1}X'_o \\ -X_o(X'X)^{-1} & I_{T_o} + X_o(X'X)^{-1}X'_o \end{bmatrix}$$

Hence, $s^{*2}(X^{*\prime}X^*)^{-1} = s^2(X^{*\prime}X^*)^{-1}$ and is given by (1.11).

(d) If we replace y_o by 0 and I_{T_o} by $-I_{T_o}$ in (1.9), we get

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} X & 0 \\ X_o & -I_{T_o} \end{bmatrix}\begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} u \\ u_o \end{bmatrix}$$

or $y^* = X^*\delta + u^*$. Now

$$X^{*\prime}X^* = \begin{bmatrix} X' & X'_o \\ 0 & -I_{T_o} \end{bmatrix}\begin{bmatrix} X & 0 \\ X_o & -I_{T_o} \end{bmatrix} = \begin{bmatrix} X'X + X'_oX_o & -X'_o \\ -X_o & I_{T_o} \end{bmatrix}$$

and $X^{*\prime}y^* = \begin{bmatrix} X'y \\ 0 \end{bmatrix}$. The OLS normal equations yield

$$(X'X)\widehat{\beta}_{\mathrm{OLS}} + (X'_oX_o)\widehat{\beta}_{\mathrm{OLS}} - X'_o\widehat{\gamma}_{\mathrm{OLS}} = X'y$$

and

$$-X_o\widehat{\beta}_{\mathrm{OLS}} + \widehat{\gamma}_{\mathrm{OLS}} = 0$$

From the second equation, it immediately follows that $\widehat{\gamma}_{\mathrm{OLS}} = X_o\widehat{\beta}_{\mathrm{OLS}} = \widehat{y}_o$, the forecast of the T_o observations using the estimates from the first n observations. Substituting this in the first equation yields

$$(X'X)\widehat{\beta}_{\mathrm{OLS}} + (X'_oX_o)\widehat{\beta}_{\mathrm{OLS}} - X'_oX_o\widehat{\beta}_{\mathrm{OLS}} = X'y$$

which gives $\widehat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'y$. Alternatively, one could apply the FWL theorem using $X_1 = \begin{bmatrix} X \\ X_o \end{bmatrix}$ and $X_2 = \begin{bmatrix} 0 \\ -I_{T_o} \end{bmatrix}$. In this case, $X_2'X_2 = I_{T_o}$ and $P_{X_2} = X_2X_2' = \begin{bmatrix} 0 & 0 \\ 0 & -I_{T_o} \end{bmatrix}$ as before. This means that $\overline{P}_{X_2} = I_{n+T_o} - P_{X_2} = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$.

As in part (a), premultiplying by $\overline{P}_{X_2}$ omits the last T_o observations and yields $\widehat{\beta}_{\mathrm{OLS}}$ based on the regression of y on X from the first n observations only. The last T_o observations yield zero residuals because the dependent and independent variables for these T_o observations have zero values. For this to be true in the original regression, it must be true that $0 - X_o\widehat{\beta}_{\mathrm{OLS}} + \widehat{\gamma}_{\mathrm{OLS}} = 0$, which yields $\widehat{\gamma}_{\mathrm{OLS}} = X_o\widehat{\beta}_{\mathrm{OLS}} = \widehat{y}_o$ as expected. The residuals are still $(e'_{\mathrm{OLS}}, 0')$ and $s^{*2} = s^2$ for the same reasons given above. Also, using partitioned inverse, one gets

$$(X^{*\prime}X^*)^{-1} = \begin{bmatrix} (X'X)^{-1} & (X'X)^{-1}X_o' \\ X_o(X'X)^{-1} & I_{T_o} + X_o(X'X)^{-1}X_o' \end{bmatrix}$$

Hence, $s^{*2}(X^{*\prime}X^*)^{-1} = s^2(X^{*\prime}X^*)^{-1}$ and the diagonal elements are as given in (1.11).