Panel Data Models - part II

(Wooldridge, chapter 11)

1 Relaxing the strict exogeneity assumption

- The strict exogeneity assumption can be very strong and is often violated in economic problems. It states that,

$$E(u_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = 0$$

- We will start by considering the implications of non-strictly exogenous regressors on the WG estimator (see also Nickell, 1981, Econometrica)

- Take the simple model with a single explanatory variable, the lagged dependent variable ($x_{it} = y_{it-1}$):

$$y_{it} = \alpha y_{it-1} + f_i + u_{it} \tag{1}$$

  and therefore

$$E(u_{it}|x_{it+1}) = E(u_{it}|y_{it}) = E(u_{it}|\alpha y_{it-1} + f_i + u_{it}) \neq 0$$

- However, if u_{it} is serially uncorrelated for all i, then

$$E(u_{it}|x_{i1},\ldots,x_{it}) = E(u_{it}|y_{i0},\ldots,y_{it-1}) = 0$$

  since y_{ij} is a function of $u_{i1},\ldots,u_{ij}$ but not of $u_{ij+1},\ldots,u_{iT}$ and therefore u_{it} is mean independent of the explanatory variables up to time t, $y_{i0},\ldots,y_{it-1}$.

2 The WG estimator when the strict exogeneity assumption does not hold

- We start by applying the WG transformation to model (1):

$$y_{it} - \overline{y}_{i+} = \alpha(y_{it-1} - \overline{y}_{i-}) + (u_{it} - \overline{u}_{i+})$$

  where,

$$\overline{y}_{i+} = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \qquad \overline{y}_{i-} = \frac{1}{T}\sum_{t=0}^{T-1} y_{it}, \qquad \overline{u}_{i+} = \frac{1}{T}\sum_{t=1}^{T} u_{it}$$

  and will now check what the consequences are of relaxing the strict exogeneity assumption.

- Suppose $\alpha > 0$ (you can perform a similar analysis for the case $\alpha < 0$). Then the past values of u will have a positive effect on y_{it} while, from the model, the contemporaneous u_{it} partially determines y_{it}.

- Let $v_{it} = f_i + u_{it}$. Successive replacements yield,

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + v_{it} \\ &= \alpha^2 y_{it-2} + v_{it} + \alpha v_{it-1} \\ &= \dots \\ &= \alpha^t y_{i0} + \sum_{s=0}^{t-1} \alpha^s v_{it-s} \\ &= \alpha^t y_{i0} + \sum_{s=0}^{t-1} \alpha^s f_i + \sum_{s=0}^{t-1} \alpha^s u_{it-s} \end{aligned}$$

- Take $j > 0$. The model implies

$$\begin{aligned} E(u_{ij} y_{it}) &= E\left[u_{ij}\left(\alpha^t y_{i0} + \sum_{s=0}^{t-1} \alpha^s f_i + \sum_{s=0}^{t-1} \alpha^s u_{it-s}\right)\right] \\ &= \begin{cases} E\left[\alpha^{t-j} u_{ij}^2\right] = \alpha^{t-j}\sigma_u^2 > 0 & \text{if } j < t \\ E\left[\alpha^{t-t} u_{it}^2\right] = \sigma_u^2 > 0 & \text{if } j = t \\ 0 & \text{if } j > t \end{cases} \end{aligned}$$

- But then, $\overline{y}_{i-}$ will be positively affected by u_{it}:

$$E(u_{ij}\overline{y}_{i-}) = \frac{1}{T}\sum_{s=j}^{T-1} \alpha^{s-j}\sigma_u^2 > 0$$

- As a consequence, u_{ij} will be negatively related with the regressor $y_{it-1} - \overline{y}_{i-}$ when $j > t-1$,

$$\begin{aligned} E\left(u_{ij}\left(y_{it-1} - \overline{y}_{i-}\right)\right) &= -E\left(u_{ij}\overline{y}_{i-}\right) \\ &= -\frac{1}{T}\sum_{s=j}^{T-1} \alpha^{s-j}\sigma_u^2 < 0 \end{aligned}$$

  and in particular for $j = t$,

$$E\left(u_{it}\left(y_{it-1} - \overline{y}_{i-}\right)\right) = -\frac{1}{T}\sum_{s=t}^{T-1} \alpha^{s-t}\sigma_u^2 < 0$$

  implying that the WG estimator is downward biased.

- However, as T grows, the bias vanishes (under the assumption that $|\alpha| < 1$). But this is typically not of much use in the practical applications in microeconometrics as T is usually small.

3 The Anderson Hsiao Estimator: first differencing methods

- We need an alternative to within groups.
- Take again the basic model in the general notation,

$$y_{it} = \mathbf{x}_{it}\beta + f_i + u_{it} \quad \text{for } i = 1, ...N \text{ and } t = 1, ...T$$

- We can take first differences over time to obtain,

$$\Delta y_{it} = \Delta \mathbf{x}_{it}\beta + \Delta u_{it}$$

where $\Delta y_{it} = y_{it} - y_{it-1}$, $\Delta \mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{it-1}$ and $\Delta u_{it} = u_{it} - u_{it-1}$.

- Notice that by differencing the regression equation, we got rid of the fixed effect.
- However, we now have to deal with *past values* of the error term.
- In the absence of strict exogeneity, which was what lead us here, it is possible that past shocks predict contemporaneous regressors. In such case, $E(u_{it-1}|\mathbf{x}_{it}) \neq 0$.
- Hence in general $E(\Delta u_{it}|\Delta \mathbf{x}_{it}) \neq 0$.
- *Solution: instrumental variables.*

4 The Instrumental Variables approach

- We consider the system of equations with N iid observations, $\{y_{i1}, \dots, y_{iT}, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}\}$

$$y_{it} = \mathbf{x}_{it}\beta + u_{it} \qquad \text{for } t = 1, \dots, T$$

- We can write it in the system of equations format,

$$y_i = X_i\beta + u_i$$

4.1 Classical IV

Just identified case

- This occurs when the number of instruments equals the number of regressors (notice that Z includes the exogenous variables in X).
- In this case, the IV assumptions are,

IV1: Rank condition: $E(Z_i'X_i)$ is of order K.

IV2: Exclusion restriction: $E(Z_i'u_i) = 0$.

and the IV estimator is,

$$\beta^{IV} = (Z'X)^{-1}Z'Y$$

Over identified case

- This occurs when the number of instruments is higher than the number of regressors and the IV method is the 2SLS.
- The IV assumptions are,

  IV1': Rank condition: $E(Z_i'X_i)$ is of order K, which is the number of regressors in X, and $E(Z_i'Z_i)$ is of order L, which is the number of instruments in Z and where $L \geqslant K$.

  IV2': Exclusion restriction: $E(Z_iu_i) = 0$.

- The IV estimator explicitly uses the linear projection of X onto Z,

$$\widehat{X} = Z(Z'Z)^{-1}Z'X$$

- We then notice that $E(\widehat{X}'u) = 0$ since $E(Z'u) = 0$. This means that $\widehat{X}$ excludes the part of X that is related with the error term.
- But then, $\widehat{X}$ can be used as an instrument in the second step regression to yield,

$$\begin{aligned}
\beta^{2SLS} &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'Y \\
&= \left(\widehat{X}'X\right)^{-1} \widehat{X}'Y \\
&= \left(\widehat{X}'\widehat{X}\right)^{-1} \widehat{X}'Y \\
&= \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{x}_{it}'\widehat{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{x}_{it}'y_{it}\right)
\end{aligned}$$

  That is, this is the 2SLS estimator for pooled regressions.

The asymptotic distribution of the IV estimator

- The 2SLS estimator is

$$\begin{aligned}
\beta^{2SLS} &= \beta + \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'u \\
&= \beta + \left[\left(\sum_{i=1}^{N}\frac{1}{N}X_i'Z_i\right)\left(\sum_{i=1}^{N}\frac{1}{N}Z_i'Z_i\right)^{-1}\left(\sum_{i=1}^{N}\frac{1}{N}Z_i'X_i\right)\right]^{-1} \\
&\quad \left(\sum_{i=1}^{N}\frac{1}{N}X_i'Z_i\right)\left(\sum_{i=1}^{N}\frac{1}{N}Z_i'Z_i\right)^{-1}\left(\sum_{i=1}^{N}\frac{1}{N}Z_i'u_i\right)
\end{aligned}$$

- Under IV1' and IV2', 2SLS is consistent.

- Under the exclusion restriction we can apply the CLT to guarantee that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i'u_i \overset{a}{\sim} \mathcal{N}\left(0, E\left(Z_i'u_iu_i'Z_i\right)\right)$$

- As for the rest of the matrices composing the estimator β^{2SLS}, the rank condition ensures that

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} X_i'Z_i &= E(X_i'Z_i) = M_{XZ} \text{ of rank } K. \\
\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} Z_i'Z_i &= E(Z_i'Z_i) = M_{ZZ} \text{ of rank } L > K.
\end{aligned}$$

- Hence,

$$\sqrt{N}\left(\beta^{2SLS} - \beta\right) \overset{a}{\sim} \mathcal{N}(0, V)$$

where,

$$V = \left(M_{XZ}M_{ZZ}^{-1}M_{XZ}'\right)^{-1} M_{XZ}M_{ZZ}^{-1}E\left(Z_i'\Omega_iZ_i\right)M_{ZZ}^{-1}M_{XZ}'\left(M_{XZ}M_{ZZ}^{-1}M_{XZ}'\right)^{-1}$$

and $\Omega_i = E\left(u_iu_i'\right)$

A few notes about the 2SLS

- 2SLS is equivalent to GLS applied to the transformed model,

$$Z_i'y_i = Z_i'X_i\beta + Z_i'u_i$$

under the additional assumption that $E\left(Z_i'u_iu_i'Z_i\right) = \sigma_u^2 E\left(Z_i'Z_i\right)$.

- In panel data models, this is generally not the case. As a consequence, the estimate of the variance of β is not consistent as it is based on this assumption.

- One alternative is to use the White estimator (Econometrica, 1980) for the covariance matrix when there are suspicions of heteroscedasticity or serial correlation.

- The basic result for heteroscedasticity states that (in a cross section setup),

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} \widehat{u}_i^2 z_i' z_i = \underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} u_i^2 z_i' z_i$$

  where $\widehat{u}_i$ are the residuals obtained using a consistent estimator.

- In the context of panel data, this result extends to,

$$\begin{aligned} \underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} Z_i'\widehat{u}_i\widehat{u}_i' Z_i &= \underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} Z_i' u_i u_i' Z_i \\ &= E(Z_i'\Omega_i Z_i) \end{aligned}$$

- In practice, we need to estimate $E(Z_i'\Omega_i Z_i)$ and we will use the analogy principle as before,

$$\left(\widehat{\frac{Z'\Omega Z}{N}}\right) = \frac{1}{N}\sum_{i=1}^{N} Z_i'\widehat{u}_i\widehat{u}_i' Z_i$$

  where the $\widehat{u}$ are residuals from the consistent 2SLS procedure. This matrix should replace $E(Z_i'\Omega_i Z_i)$ in the covariance expression.

4.2 GMM estimation

- The 2SLS may not be efficient as it does not take into account possible correlations of the unobservable u over time for the same individual.

- To solve the endogeneity problem we transform our system of equations to,

$$Z_i' y_i = Z_i' X_i \beta + Z_i' u_i \tag{2}$$

  exploring the exclusion restriction IV2'.

- The 2SLS estimator is the GLS estimator under the assumption that

$$E\left(Z_i' u_i u_i' Z_i\right) = \sigma_u^2 E\left(Z_i' Z_i\right)$$

  which excludes serial correlation in u_{it}.

- The GMM estimator (3SLS) is the GLS estimator of model (2) without restricting the variance of $Z_i' u_i$.

- Let $\Omega = \text{var}(Z_i' u_i)$. Then

$$\beta^{GMM} = \left[\left(\sum_{i=1}^{N} X_i' Z_i\right)\Omega^{-1}\left(\sum_{i=1}^{N} Z_i' X_i\right)\right]^{-1}\left(\sum_{i=1}^{N} X_i' Z_i\right)\Omega^{-1}\left(\sum_{i=1}^{N} Z_i' y_i\right) \tag{3}$$

Two-step GMM

- However, Ω is not known at the start of the estimation procedure.
- So we can implement GMM in two steps. This is called feasible GMM (FGMM):
  1. Estimate the model using a consistent but not efficient estimator. A possibility is to use 2SLS. Then use the residuals from the first step to estimate Ω

     $$\widehat{\Omega} \;=\; \left[\frac{1}{N}\sum_{i=1}^{N} Z_i'\widehat{u}_i\widehat{u}_i{}'Z_i\right]$$

     with $\widehat{u}_{it} = y_{it} - \mathbf{x}_{it}\beta^{2SLS}$.
  2. Replace $\widehat{\Omega}$ in (3) to obtain

$$\beta^{FGMM} \;=\; \left[\left(\sum_{i=1}^{N} X_i'Z_i\right)\widehat{\Omega}^{-1}\left(\sum_{i=1}^{N} Z_i'X_i\right)\right]^{-1}\left(\sum_{i=1}^{N} X_i'Z_i\right)\widehat{\Omega}^{-1}\left(\sum_{i=1}^{N} Z_i'y_i\right)$$

Some notes of caution for two step GMM

- It has been shown that in finite samples one should avoid using too many orthogonality conditions. IV with too many orthogonality conditions for the sample size N produces estimates biased towards OLS (projecting X on Z when the number of instruments is large yields $\widehat{X}$ close to X).
- FGMM gives hopelessly biased estimates of the standard errors: they tend to come out much too small relative to the true variance of the estimator. Adjustments have been derived by Windmeijer (IFS working paper) and also based on the bootstrap.

5 The choice of the instrument(s) in dynamic panel data models

- Consider the simple dynamic model we have used before,

  $$y_{it} \;=\; \beta_0 + \beta_1 y_{it-1} + f_i + u_{it}$$

  and take first differences to get rid of the fixed effect,

  $$\Delta y_{it} \;=\; \beta_1 \Delta y_{it-1} + \Delta u_{it}$$

- If u_{it} are uncorrelated over time, Δu_{it} follows an MA(1) process and we can, in principle, choose as instrument $z_{it} = \Delta y_{it-2}$, and obtain,

$$\begin{aligned} \beta_1^{IV} &= \frac{\text{cov}(z_{it}, \Delta y_{it})}{\text{cov}(z_{it}, \Delta x_{it})} \\ &= \frac{\text{cov}(\Delta y_{it-2}, \Delta y_{it})}{\text{cov}(\Delta y_{it-2}, \Delta y_{it-1})} \end{aligned}$$

  which is the one-regressor/one-instrument estimator of β_1.

- The exclusion restriction holds under the serially uncorrelated error terms assumption, which implies

$$E\left(\Delta y_{it-2} \Delta u_{it}\right) = E\left[\left(\beta_1 \Delta y_{it-3} + \Delta u_{it-2}\right) \Delta u_{it}\right] = 0$$

- However, to guarantee the validity of this procedure we also need to make sure that the rank condition holds.

- In general, it does. But there are cases where it does not.

- Consider the special case of the model above where u_{it} are iid and $\beta_1 = 1$,

$$y_{it} = y_{it-1} + u_{it}$$

  and notice that,

$$\Delta y_{it} = u_{it}$$

- Thus, if we use Δy_{it-2} to instrument Δy_{it-1} the rank condition is not satisfied,

$$E\left(\Delta y_{it-2} \Delta y_{it-1}\right) = E\left(u_{it-2} u_{t-1}\right) = 0$$

- The estimator fails in this case because Δy_{it-2} fails to predict Δy_{it-1} since Δy_{it-1} is an unpredictable shock.

The efficient choice of instruments

- Suppose we wish to estimate efficiently (asymptotically) the following dynamic panel data model with iid errors and fixed effects,

$$y_{it} = \beta y_{it-1} + f_i + u_{it} \qquad \text{where } \beta \neq 1$$

- If u_{it} is serially uncorrelated, we have seen that Δy_{it-2} is a valid instrument in general.

- However, there may be more valid instruments available.

- Suppose $T > 2$. At each period t we have $t-2$ orthogonality (order) conditions,

$$\begin{aligned} E\left(y_{it-2}\Delta u_{it}\right) &= 0 \\ E\left(y_{it-3}\Delta u_{it}\right) &= 0 \\ &\vdots \\ E\left(y_{i1}\Delta u_{it}\right) &= 0 \end{aligned}$$

- Efficiency requires that we use all the information contained in the above relationship. To see exactly how many orthogonality conditions we have, suppose we have a panel of dimension $T = 5$. We lose two observations by lagging and differencing. Then we get,

$$\begin{array}{ccc} t=3 & t=4 & t=5 \\ E\left(y_{i1}\Delta u_{i3}\right)=0 & E\left(y_{1}\Delta u_{i4}\right)=0 & E\left(y_{i1}\Delta u_{i5}\right)=0 \\ & E\left(y_{2}\Delta u_{i4}\right)=0 & E\left(y_{i2}\Delta u_{i5}\right)=0 \\ & & E\left(y_{i3}\Delta u_{i5}\right)=0 \end{array}$$

- We have 6 orthogonality conditions and wish to use them all: each orthogonality condition defines a different estimator; the efficient estimator that combines all restrictions together is called GMM.

- To apply it, we need to construct Z such that $Z'\Delta u = \sum_{i=1}^{N} Z_i'\Delta u_i = 0$ is the collection of all orthogonality conditions.

- In our example we can write

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{bmatrix} \qquad \text{where} \qquad Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{i1} & y_{i2} & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & y_{i3} \end{bmatrix}$$

and

$$\Delta u_i = \begin{bmatrix} \Delta u_{i3} \\ \Delta u_{i4} \\ \Delta u_{i5} \end{bmatrix}$$

6 The cost of first differencing

- Taking differences is not an innocuous procedure. To see why we consider a model with one *strictly exogenous* regressor, no fixed effect and an iid error. Moreover, suppose we have just two time periods $T = 2$ for N individuals.

- More explicitly, the model is,

$$y_{it} = x_{it}\beta + u_{it} \quad \text{for } i = 1, \ldots, N \text{ and } t = 1, 2$$

  where, $u_{it} | (x_{i1}, x_{i2}) \sim \text{iid}\left(0, \sigma_u^2\right)$.

- In this model, OLS is the best estimator (BLUE), with variance being

$$\text{var}\left(\beta^{OLS}\right) = \frac{\sigma_u^2}{2N\text{var}\left(x_{it}\right)}$$

- The first differencing (FD) estimator is,

$$\beta^{FD} = \frac{\text{cov}(\Delta x_{it} \Delta y_{it})}{\text{var}(\Delta x_{it})}$$

  which has variance,

$$\begin{aligned} \text{var}(\beta^{FD}) &= \frac{2\sigma_u^2}{N\text{var}(\Delta x_{it})} \\ &= \frac{2\sigma_u^2}{N\left[\text{var}(x_{i1}) + \text{var}(x_{i2}) - 2\text{cov}(x_{i1}, x_{i2})\right]} \end{aligned}$$

- By first differencing we get twice the noise $(2\sigma_u^2)$. If $\text{cov}\left(x_{i1}, x_{i2}\right) > 0$ (as is likely for many economic variables) then we also get less signal. Finally we lose N observations.

- First differencing can reduce precision greatly. So it should be used only when properly justified.