3. ***Dynamic Models***

3.1 Introduction

CHANGE OF NOTATION

Consider $y_{it} = \gamma\, y_{it-1} + \mathbf{x_{it}}\boldsymbol{\beta} + c_i + u_{it}$ $\quad t=2,\ldots,T \qquad (1)$

The assumption is (weak exog):

$$E(y_{it}|\mathbf{x_{it}}, \mathbf{y_{it-1}}, \ldots, \mathbf{x_{i2}}, y_{i1}|c_i) = E(y_{it}|\mathbf{x_{it}}, y_{it-1}, c_i) = \gamma\, y_{it-1} + \mathbf{x_{it}}\boldsymbol{\beta} + c_i + u_{it} \qquad (2)$$

Can't justify strict exog assumption! **WHY?**

<u>Notes</u>:

1. Above does not require u_{it} to be uncorrelated with the future values of **x**. i.e Feedback is allowed for from y_{it} to $\mathbf{x}_{it+1},\ldots, \mathbf{x}_{iT}$. If necessary we can impose additional orthogonality conditions to make u uncorrelated with past, current and future **x**s.

   <u>Example</u>:

   Static model with feedback: [x is w.exog]

   $y_{it} = \delta x_{it} + c_i + u_{it}$ and $x_{it} = \alpha y_{it-1} + \phi c_i + v_{it}$

Here, strict exog assumption for x will not work!

Consider $E(x_{it+1}\ u_{it}) = \alpha\ E(y_{it}u_{it}) = \alpha\ E(u_{it}^2) \neq 0$ even if other terms are 0.

2. With short panels we cannot calculate individual time series auto-covariances using something like $\frac{1}{T-1}\sum_{t=2}^{T} y_t y_{t-1}$. What we calculate is cross-sectional auto-covariance using $\frac{1}{N}\sum_{i=1}^{N} y_{i1} y_{i2}$ for example. **very important**!

3. Asymptotics based on large N and fixed T. So NO need to bother with assumption about γ. Identification of parameters and small sample properties of the estimators - depend on the time series properties of the series. Influence of the initial observation cannot be ignored when T is small.

4. Persistence can come from two different sources…. Heterogeneity or state dependence…..

Have true state dependence when $\gamma \neq 0$. When $\gamma=0$, we still observe non-zero correlation (see RE model). When this happens, we say that we have spurious state dependence.

5. IDENTIFICATION

T=2 (becomes a cross-section model)

Consider Model 1: $y_{it} = c + u_{it}$, $u_{it}=\rho u_{it-1}+\varepsilon_{it}$

Vs Model 2: $y_{it} = c_i + u_{it}$

In Model 1: $\text{corr}(y_{i1}, y_{i2})= \text{corr}(c + u_{i1}, c + u_{i2})= \rho$

In Model 2: $\text{corr}(y_{i1}, y_{i2}) = \sigma^2_c / (\sigma^2_c + \sigma^2_u)$

Can't distinguish between $\rho=0.4$ and $\sigma^2_c / (\sigma^2_c + \sigma^2_u) = 0.4$ for eg.

Need at least T=3 here……

5. y_{it-1} and $c_{i.}$ are correlated.

6. Bias calculations for various estimators are based on the assumption about the initial condition of the process.

7. <u>Estimation</u>: WG transformation to eliminate the c_i will not help. The LDV in mean deviation will be correlated with the u_{it} in mean deviation via the

means. Thus WG is inconsistent (Nickell bias). Inconsistency vanishes as $T \rightarrow \infty$. i.e. need large T and large N. with large σ^2_c and fixed y_{i1} the bias in the WG est will be small. WG biased downwards.

8. Pooled OLS is also biased and inconsistent. Here the bias is upwards.

9. FD est is also biased downwards.

10. MLE depends on the assumption about y_{i1} [fixed or stochastic, correlation with c_i or not, stationarity….].

11. So better to use IV

3.2 Estimation

Take first differences and then use instrumental variable estimation.

Consider the simpler model: $\Delta y_{it} = \gamma \, \Delta y_{it-1} + \Delta u_{it}$ $\quad t=2,..,T$ (3)

Choice of instruments depends on the assumptions we make.

Assump (2) $E(u_{it} | \, y_{it-1},..,y_{i1},c_i) = 0$ implies the following:

1. $E(y_{\mathbf{it-1}} \, u_{it})=0$
2. $E(y_{is} \, \Delta u_{it})=0$ $\quad s=1,.,t-2; \; t>2;$ [note: $\boldsymbol{\Delta y_{it} - \gamma \, \Delta y_{it-1} = \Delta u_{it}}$]
3. $E(u_{it} \, u_{it-j} | y_{i1},..,y_{it-1},c_i)=0$ $\quad j>0$ (serially uncorrelated conditionally)

4. $E(u_{it}\, u_{it-j})=0$ (serially uncorrelated unconditionally too).

5. $E(c_i\, u_{it}) = 0$ for all t.

Thus, $(y_{i1},\ldots,y_{it-2})$ are valid instruments at time t.

Instruments (Anderson & Hsiao (1981)): $(y_{it-2} - y_{it-3})$ or y_{it-2} for Δy_{it-1};

$$\Delta y_{it} = \gamma\, \Delta y_{it-1} + \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it} \qquad (4)$$

In practice, the choice will obviously depend on the relevant correlations.

Arellano & Bond: use all available moment condition and do **GMM**.

For fixed T (>2) and assuming predetermined y_{i1} (can be correlated with c_i):

Moments for t=3: $E[(u_{i3}-u_{i2})y_{i1}]=0$

For t=4; $E[(u_{i4}-u_{i3})y_{i2}]=0$ & $E[(u_{i4}-u_{i3})y_{i1}]=0$; and so on….

Giving the matrix of instruments (when there are no **x**s):

$$\mathbf{Z}_i = \begin{bmatrix} [y_{i1}] & 0 & 0 & 0 \\ 0 & [y_{i1}, y_{i2}] & 0 & 0 \\ . & 0 & ... & 0 \\ 0 & 0 & . & [y_{i1}, y_{i2}, .. y_{iT-2}] \end{bmatrix}$$

T-2 rows & $\frac{1}{2}$ (T-1)(T-2) cols

So the set of moment conditions: $E[\mathbf{Z_i}'\Delta \mathbf{u}_i]=0$ **LINEAR in γ**

In the simple model:

$$\hat{\boldsymbol{\gamma}}_{gmm} = \left[\left(\sum_i \Delta\mathbf{y}_{-1}'\mathbf{Z}_i\right)\mathbf{W}\left(\sum_i \mathbf{Z}_i'\Delta\mathbf{y}_{-1}\right)\right]^{-1}\left[\left(\sum_i \Delta\mathbf{y}_{-1}'\mathbf{Z}_i\right)\mathbf{W}\left(\sum_i \mathbf{Z}_i'\Delta\mathbf{y}\right)\right] \quad (5)$$

with the optimal weighting matrix given by:

$\mathbf{W} = E\left[\mathbf{Z}_i'\Delta\mathbf{u}_i\Delta\mathbf{u}_i'\mathbf{Z}_i\right]^{-1}$ and is estimated by the sample equivalent

$$\hat{\mathbf{W}} = \left[\frac{1}{N}\sum_i \mathbf{Z}_i'\Delta\hat{\mathbf{u}}_i\Delta\hat{\mathbf{u}}_i'\mathbf{Z}_i\right]^{-1} \quad (6)$$

- AVAR $\hat{\gamma} = \left[\left(\mathbf{X'Z} \right) \left(\sum_i \mathbf{Z_i}'\mathbf{\hat{u}_i}\mathbf{\hat{u}_i}'\mathbf{Z_i} \right)^{-1} \left(\mathbf{Z'X} \right) \right]^{-1}$ (N cancels) (7)

- T=3 Model is just identified; Over identified for T>3. Can test for this using Sargan's (Hansen's) test.

- Δu_{it} is MA(1); Under homoskedasticity assumption on u_{it},

$$\mathbf{W}_{1N} = \left[\frac{1}{N} \sum \mathbf{Z}_i' \mathbf{H} \mathbf{Z}_i \right]^{-1} \quad \text{where} \qquad (8)$$

$$\mathbf{H} = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix} \quad \textbf{Does not depend on any param.} \qquad \textbf{(9)}$$

- If Δu_{it} is MA(1), the second order serial correlation should be zero. This can be tested using Arellano-Bond Test which uses the sample serial correlation coefficient.

- Use the first-step GMM with W=H to get the residuals for the two-step GMM.
- Stop with the second step (not a lot of efficiency gains in continuing).
- Simulation studies show - very little efficiency gain in doing 2-step GMM even in the presence of considerable heteroskedasticity.
- Additionally, since 2-step GMM depends on 1^{st} step coeff est, the std. error calculations tend to be too small…. Windmeijer provides a correction (now implemented in some software – PcGive).

- If we have st exog covariates **x** we can have further moment conditions.
- Can end up with too many instruments and get inefficiency. So might want to reduce the set of instruments – for eg use A-H. Small sample biases when too many instruments are used can be large.
- With additional assumptions, there are other moment conditions that can be used to improve the efficiency of the GMM. Some of these are quadratic and some are linear in γ.

3.2.1 **Weak instruments** (Blundell and Bond)

If {y} is very close to being a random walk or when σ^2_c/σ^2_u is becomes large, correlation between level of y and the difference of y will be weak. IV methods will not work properly (weak instruments).

- IV suffers from very serious finite sample bias when the instruments are weak.

- Solution: Under homoskedasticity assumption and that y_{i1} satisfies mean stationarity [$E(y_{i1})=c_i/(1-\gamma)$ for each i], can use Δy_{it-1} as an instrument for the levels equation. To see this:

Write $y_{i1} = c_i/(1-\gamma) + \varepsilon_{i1}$ (under mean stationarity)

This gives, $E(c_i \varepsilon_{i1})=0$

$y_{i2} = \gamma y_{i1} + c_i + u_{i2} = \gamma c_i/(1-\gamma) + \gamma\varepsilon_{i1} + c_i + u_{i2} = c_i/(1-\gamma) + \gamma\varepsilon_{i1} + u_{i2}$

Thus, $\Delta y_{i2} = \gamma\varepsilon_{i1} + u_{i2}$ giving you $E(\Delta y_{i2} c_i)=0$

Which in turn will give you (with the previous assumptions)

$E[(y_{it-1}-y_{it-2})(c_i+u_{it})] = E(\Delta y_{it-1}(c_i + u_{it}))=0 \qquad t=3,4,\ldots,T$

Hence differences can be used as instruments for the levels giving you an additional (**T-2**) **LINEAR** moment conditions.

[The conditions regarding the initial observation translates to other periods because Δy_{it} can be written as a function of Δy_{i2} plus errors using repeated substitution.]

In summary, we use: $E(\mathbf{Z_i}'\Delta u_i)=0$ & $E[\Delta y_{it-1}(c_i + u_{it})]=0$.

These make the quadratic moment conditions redundant.

- The estimation in the above case requires a combination of equations in FD as well as in levels. The use of both reduces the problem of weak instruments.

- For large T case, the consistency depends on $T/N \rightarrow \text{constant} < \infty$.

Ref: Arellano, M & Bond, S. (1991) – "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies*, 58, 277-297.

3.3 <u>Other points</u>

- MLE and GLS require some assumptions regarding y_{i1} (stochastic or non-stochastic, correlated with c or not, etc..).

3.4 <u>Test for over identifying restrictions</u> [Sargan/ Hansen]

Moment condition: $E(\mathbf{Z_i}'\Delta\mathbf{u_i})=0$ L instruments for K covariates

Min: $$Q=\left[\frac{1}{N}\sum_i \mathbf{Z_i}'\Delta\mathbf{u_i}\right]'\mathbf{W}\left[\frac{1}{N}\sum_i \mathbf{Z_i}'\Delta\mathbf{u_i}\right]$$

When L=K, $Q(\hat{\boldsymbol{\beta}})=0$; **[β refers to all the coefficients now!]**

When L>K, then $Q(\hat{\boldsymbol{\beta}})>0$ although $Q(\hat{\boldsymbol{\beta}})\rightarrow 0$ in probability.

So use this to derive the test by comparing the value of the criterion function Q with its expected value under the null that the restrictions are valid.

This is simple when the optimal weighting matrix is used:

$N\,Q(\hat{\boldsymbol{\beta}})$ dist asym as a $\chi^2(L-K)$ under H_0.

Only use the two-step GMM results.

If you only suspect that L_1 are ok but L_2 are not (where $L=L_1 + L_2$) then Can use $N(Q-Q_1)$ is asymp $\chi^2(L_2)$ where Q_1 is the minimand when the non-suspect instruments L_1 are used.

Sargan's test uses the minimand from the 2-step GMM. This has a convenient chi sq regardless of heterosk.

Should use the same est of the optimal weighting matrix in the calculation.

Use different set of moment restrictions to see how the ests change since having too many inst can cause small sample bias and also ineff. Also sargan's test also has low power when too many inst are used.

3.5 Test for serial correlation

Should not have serial correlation in the Δu_{it}. Use the sample correlation coefficients to test for H_0: no correlation. Arellano-Bond test. Complicated formula. Test statistic is std normally distributed under the null of no serial correlation

3.6 STATA

In STATA you can do the following:

- 1-step GMM – this uses the H matrix on page 14 to calculate the estimates. You can do this in 1-step since the H matrix does not depend on any parameters (see (8) and (9)).

- 1-step GMM with robust covar matrix, estimates the parameters using the 1-step method and then, instead of estimating the covar matrix as

$$\text{AVAR}\ \hat{\boldsymbol{\beta}} = \left[\left(\mathbf{X'Z}\right) \left(\sum_i \mathbf{Z_i}'\Delta\hat{\mathbf{u}}_\mathbf{i}\Delta\hat{\mathbf{u}}_\mathbf{i}'\mathbf{Z_i} \right)^{-1} \left(\mathbf{Z'X}\right) \right]^{-1} \quad \text{(N cancels)} \quad (10)$$

it uses the residuals to calculate (10).

- 2-step GMM uses the optimal weighting matrix and calculates the AVAR as in (10). i.e uses the residuals from step 1 to calculate the optimal weighting matrix as in (6) and recalculates (5) using this new W. The AVAR is then calculated as in (10) above using the new set of residuals.