# Economics 6003
# Quantitative Economics

## Classic Dynamic Panel Models // Difference and System GMM Estimators

John Bluedorn

Lecturer, Economics Division

Thursday, 7 May 2009 and Monday, 11 May 2009

# Classic linear panel model framework

- Previously, we considered how to estimate a classic, linear, unobserved effects panel model:

$$y_{i,t} = \underset{(1\times K)}{\beta}{}' \underset{(K\times 1)}{x_{i,t}} + \alpha_i + \varepsilon_{i,t}$$

- We saw that consistency and the accuracy of inference regarding the RE and FE estimators requires that a strict/strong exogeneity assumption be satisfied:

$$E\left(\varepsilon_{i,t}|\alpha_i, x_i\right) = 0$$

where $x_i = \{x_{i,s} \forall s = 1, \ldots, T\}$ denotes the set of explanatory variables at all points in time (future and past).

# Linear dynamic panel model framework

- Suppose that we would like for one of the explanatory variables to be a lagged dependent variable, so that the model has the form:

$$y_{i,t} = \rho y_{i,t-1} + \underset{(1\times K)}{\beta}{}' \underset{(K\times 1)}{x_{i,t}} + \alpha_i + \varepsilon_{i,t}$$

where all variables are defined as before.

- This is known as a dynamic panel model, since $y$ exhibits state dependence. Unfortunately, we cannot use our usual RE or FE estimators, because strict/strong exogeneity is violated – one of the regressors (the lagged dependent variable) is correlated with past values of the error term.
    - More generally, multiple (or different) lags of $y$ can be included as explanatory variables.

# Linear dynamic panel model framework

- Notice that an alternative, equivalent form of the dynamic panel model is given by:

$$\Delta y_{i,t} = (\rho - 1)\, y_{i,t-1} + \underset{(1\times K)}{\beta}{}' \underset{(K\times 1)}{x_{i,t}} + \alpha_i + \varepsilon_{i,t}.$$

- If $y$ is a logarithm of some kind, then the above form states that the growth rate of $y$ is a linear function of the initial level of $y$ and a set of regressors, plus the unobserved effect and error term.

# Exogeneity violations in a linear dynamic panel

- What exactly is the source of the strict exogeneity failure?
  - The idiosyncratic error term $\varepsilon_{i,t}$ is correlated with $y_{i,t+1}$ which will be the lagged dependent variable at time $t+2$:

$$E\left(\varepsilon_{i,t} | \alpha_i, x_i, y_{i,s-1} \forall s = 1, \ldots T\right) \neq 0$$

- Moreover, the weaker condition of zero contemporaneous correlation of the regressors with the composite error term $\alpha_i + \varepsilon_{i,t}$ is clearly violated, since the composite error exhibits serial correlation due to the time-invariant, panel-specific unobserved effect and one of the regressors is a lagged dependent variable.
  - If the time dimension $T$ of a sample is small, the inconsistency can be particularly severe. A larger cross-section dimension $N$ cannot help, since the number of unobserved effects increases one-for-one with $N$. It is sometimes known as **Nickell bias** (1981).

# Exogeneity violations in a linear dynamic panel

- Note that the endogeneity problem associated with the Nickell bias arises because of the omission of the unobserved effect as a regressor (with coefficient constrained to be 1). It is *not* due to simultaneity.
  - A proxy (such as a fixed effect) is not good enough here, since it is imprecisely and inconsistently estimated under $N$-asymptotics.

# Solutions for linear dynamic panel problems

- How might we address the endogeneity problem to get a consistent estimator and accurate inference?
  - The usual solution in econometrics would be to use instrumental variables (IV) of some form. Suppose that there are no instruments available other than those contained within the set of explanatory variables (including the lagged dependent variable).
  - A natural assumption which will allow us to locate valid instruments amongst this set is *sequential exogeneity*:

$$E\left(\varepsilon_{i,t}|\alpha_i, x_{i,t}, x_{i,t-1}, \ldots, x_{i,1}, y_{i,t-1}, y_{i,t-2}, \ldots y_{i,1}\right) = 0$$

  - Sequential exogeneity only requires that the idiosyncratic error term have mean zero *conditional* upon the current set of information (the predetermined and exogenous variables) *and* the unobserved effect.

# Solutions for linear dynamic panel problems

- Even under the sequential exogeneity assumption, we still have to deal with the correlation of the unobserved effect with the lagged dependent variable.
  - A natural transformation to use is the first-difference, which will eliminate the unobserved effect:

$$\Delta y_{i,t} = \rho \Delta y_{i,t-1} + \beta' \Delta x_{i,t} + \Delta \varepsilon_{i,t}$$

  - However, the first-differencing results in a *negative* correlation between the differenced, lagged dependent variable and the differenced, idiosyncratic error term. There is thus still a need to use an IV estimation strategy.

# Difference and System GMM

- The two most common estimation strategies for a linear, dynamic panel model of the form of equation 1 are known as:
  - difference GMM (D-GMM), due Arellano and Bond (1991), who built upon earlier work by Anderson and Hsiao (1982) and Holtz-Eakin, Newey, and Rosen (1988).
  - system GMM (S-GMM), due to Blundell and Bond (1998), using an insight from Arellano and Bover (1995). S-GMM augments the D-GMM estimation strategy (under additional assumptions).

# What is GMM?

- GMM (generalized method of moments) is an estimation approach that was first derived by Hansen (1982).
- The basic decision problem in GMM is:

$$\min_{\beta} M(\beta)' WM(\beta)$$

  where $M(\beta)$ is a column vector of empirical moment conditions which depend upon the parameter $\beta$.
- Thus, the GMM objective function is a quadratic form, with a weighting matrix for the moments given by $W$.
- The weighting matrix is crucial if we have more moment conditions than we do parameters to choose; it tells us how to trade-off the satisfaction of the empirical moments in choosing $\beta$.
- It ends up that GMM encompasses a lot of familiar estimation methods, such as OLS and TSLS.

# OLS/TSLS in GMM

- In the case of OLS/TSLS, the empirical moments for the GMM objective are defined by the OLS/TSLS first order conditions.
- Suppose that we have the following cross-section linear regression model:

$$y_i = \underset{(1\times K)}{\beta}{}' \underset{(K\times 1)}{x_i} + \varepsilon_i$$

- For this model, the OLS moment condition is given by:

$$\underset{(K\times N)}{X}{}' \underset{(N\times 1)}{\varepsilon} = X'\left(y - \beta' X\right) = 0$$

where $\underset{(N\times K)}{X} = \begin{bmatrix} \underset{(1\times K)}{x_i'} \end{bmatrix}$ is the stacked set of explanatory variables by observation and $\varepsilon$ is the column vector of errors by observation.

# OLS/TSLS in GMM

- Thus, under OLS, we have exactly as many moment conditions as we have parameters to estimate ($K$).
- If we have a set of instruments $z$ which we can use, the TSLS moment condition is:

$$Z'\varepsilon = Z'\left(y - \beta' X\right) = 0$$

where $Z$ is defined similar to $X$.

- If the number of instruments is greater than the number of parameters, then the system is said to be *overidentified* (dimensions of $Z$ are greater than the dimensions of $X$).

# Weighting matrix in GMM

- How is the weighting matrix $W$ chosen in GMM?
- In general, the optimal weighting matrix (in an efficiency sense) is to use the *inverse* of the variance/covariance matrix of the empirical moments.
- In the case of the TSLS estimator, this will be:

$$\text{Var}\left(Z'\varepsilon\right) = Z'\text{Var}\left(\varepsilon\right) Z = Z'\Omega Z$$

where $\Omega = E\left(\varepsilon\varepsilon'\right)$.

# Difference GMM (D-GMM)

- D-GMM proceeds by:
  1. Take first difference of the linear dynamic panel regression to remove the individual-specific, unobserved effect.
     - Arellano and Bover (1995) note that it is also feasible to use *forward orthogonal deviations*, where the average of future values of each variable are subtracted from the current-value variable rather than the once-lagged variable (this removes the individual-specific, unobserved effect also).
     - Forward orthogonal deviations may be useful in unbalanced panels, since it will preserve degrees-of-freedom (which are lost with lots of differencing).
  2. In the first-differenced linear dynamic panel regression, sequential exogeneity and zero serial and cross-section correlation of $\varepsilon$ implies that the following moment conditions hold:

$$E\left(y_{i,t-s}\Delta\varepsilon_{i,t}\right) = 0 \forall i, t \text{ and } s = 2, \ldots \infty.$$

# Difference GMM (D-GMM)

- Essentially, the past levels of the dependent variable act as instruments for the current first differences of the dependent variable (similar to a traditional two-stage least squares identification strategy).
- We can stack these moments up and then apply GMM.
- Notice how the number of instruments available in a sample rises with the advance of time (more past levels of the dependent variable are valid instruments under the assumptions). We can rapidly get a lot of instruments.
    - Notice how the use of this instrument set is still valid if we took the forward orthogonal deviations as opposed to the usual first difference.

# System GMM (S-GMM)

- S-GMM undertakes the D-GMM estimation procedure, but augments it through the introduction of an additional assumption which generates an additional set of moment conditions to leverage. The additional assumption is that:

$$E\left(\Delta y_{i,t-s}\left[\alpha_i + \varepsilon_{i,t}\right]\right) = 0 \forall i, t \text{ and } s = 1, \ldots \infty.$$

- Thus, S-GMM requires that lagged *changes* in the dependent variable are valid instruments for the *level* of the lagged dependent variable in the level equation.
  - In general, this will be true only if deviations from the long-run mean of $y$ *conditional* upon any exogenous variables are unrelated to the individual-specific, unobserved effect.
  - Sometimes this is translated as the requirement that the starting observations in the sample are not too far from their long-run means. This is not quite right, but the intuition is useful.

# System GMM (S-GMM)

- S-GMM requires more assumptions than D-GMM, but if the assumptions hold, it achieves a greater efficiency.
- Moreover, since it uses the level version of the dynamic panel model in addition to the differenced version, the effects of time-invariant regressors can be estimated (unlike D-GMM, where they get differenced out).

# Serial correlation and sequential exogeneity

- For either D-GMM or S-GMM, the degree of serial correlation of $\varepsilon$ will determine the validity of any instruments based upon the dependent variable.
- In the discussion here, we assumed that $\varepsilon$ has zero correlation.
  - Thus, under D-GMM, lags of the dependent variable greater than or equal to 2 are valid instruments for the differenced equation.
  - Under S-GMM, this also means that lags of the change in the dependent variable greater than or equal to 1 are valid instruments for the level equation.
- If $\varepsilon$ is serially correlated, then the set of valid instruments based upon the dependent variable changes.

# Serial correlation and sequential exogeneity

- If $\varepsilon$ is AR($p$), then the valid D-GMM moment conditions are:

$$E\left(y_{i,t-s}\Delta\varepsilon_{i,t}\right) = 0 \forall i, t \text{ and } s = \left(p+2\right), \ldots \infty.$$

- Similarly, the additional valid moment conditions for S-GMM are:

$$E\left(\Delta y_{i,t-s}\left[\alpha_i + \varepsilon_{i,t}\right]\right) = 0 \forall i, t \text{ and } s = \left(p+1\right), \ldots \infty.$$

- How can we diagnose such serial correlation?
  - Arellano and Bond (1991) devised a test of serial correlation based upon the D-GMM moment conditions. Under the null that $\varepsilon$ is not serially correlated, then $\Delta\varepsilon$ will be AR(1), but not AR(2).
    - If we reject no AR(2) serial correlation, then we need to use only moment conditions for D-GMM based upon lags of the dependent variable that are greater than or equal to 3 (as opposed to 2).

# Cross-section correlation and exogeneity

- D-GMM and S-GMM also require that there be no correlation of the unobserved effects and error terms *across* cross-section units.
- A simple method which can account for at least some correlation is to include *time dummies* in the linear regression. This is generally recommended. In Stata, this is easy to accomplish with `xi`.

# Many instruments and overidentification

- Given the surfeit of instruments, it is natural to consider *overidentification* tests
  - the classic Sargan test (if errors are homoskedastic).
  - Hansen's J-statistic test (if errors are not homoskedastic), based on the weighting matrix.
- The null under both of these tests is that *all* of the instruments are valid; the alternative is that some subset (not specified) are invalid.
- A difficulty with Hansen's J-statistic test is that its *size* is distorted as the number of instruments grows.
- The Sargan test is more reliable, but it is not appropriate if homoskedasticity fails.

# Many and weak instrument issues

- D-GMM and S-GMM generates many instruments, sometimes so many that there is a real danger of *overfitting* the endogenous variables.
  - We unintentionally fit the endogenous component of their variation with the huge number of instruments employed. Thus, we do not actually instrument successfully in a finite sample.
- It is also important to note that it can easily be the case that *all* of the instruments that we generate are extremely weak (they contain little information about the endogenous variables), which leads to the now well-known weak instrument problems (inconsistency, inaccurate inference, etc.). See Pagan (2007), *Weak Instruments: A Guide to the Literature*.
  - As an example, if $y$ is extremely persistent, the lagged levels of $y$ will be weak instruments for $\Delta y$ in D-GMM.

# D-GMM and S-GMM in Stata

- The best command for the estimation of linear dynamic panel models with unobserved effects by D-GMM and S-GMM is `xtabond2`, authored by Roodman (2007).
  - The general syntax is:

    ```
    xtabond2 depvar explvarlist, gmmstyle(varlist,
    laglimits(# #) collapse equation(diff | level |
      passthru)) ivstyle(varlist, equation(diff |
       level | both) passthru mz) robust twostep
                        orthogonal
    ```

  - There are additional options which are suppressed for brevity.
- All of the identification information is conveyed in the `gmmstyle` (for predetermined and endogenous variables) and `ivstyle` (for exogenous variables) options, which can be repeated as needed for different specifications.

# D-GMM and S-GMM in Stata

- The default in `xtabond2` is S-GMM. D-GMM is executed with the option `noleveleq`.
- Stata time series operators (e.g., `L` `D` or `S`) may be employed within the command.
- To specify HAC-robust standard error calculation, add the option `robust`.
- The default treatment of `gmmstyle` variables is to use all available lags for both the differenced and level equations.
  - This can be restricted with the `laglimits` option.
  - Instead of an additional instrument for each panel, time, and lag, the option `collapse` specifies that only instrument by panel and lag are to be used. These options can be useful to restrict the proliferation of instruments.
- The default treatment of `ivstyle` variables is to apply them across both the differenced and level equations.

# S-GMM application – LLB (2000)

- Levine, Loayza, and Beck (2000) investigated the effect of financial development upon growth in a linear dynamic panel model, where economic growth was a function of the initial level of income in a period and a set of regressors.
  - Notice how they employ the second form of the linear dynamic panel regression we discussed earlier.
- They use the system GMM estimator to account for the unobserved, country-specific effect (intercepts) in the presence of the lagged dependent variable.

# S-GMM application – LLB (2000)

- LLB (2000) also treat all of the other regressors, including financial development, as predetermined, endogeneous variables.
    - They are arguing that sufficiently deep lags of the predetermined variables are valid instruments for the changes in those predetermined variables.
    - This is a strong assumption $\rightarrow$ changes in financial development can be instrumented by the past levels of financial development. There is no correlation with omitted variables once the instrumentation has occurred.

# S-GMM results for LLB (2000)

Table 4
Financial intermediation and growth: dynamic panel regressions, summary[a]

| Estimator | Conditioning information set | LIQUID LIABILITIES | COMMERCIAL-CENTRAL BANK | PRIVATE CREDIT | Observa-tions |
|---|---|---|---|---|---|
| System estimator | Simple | 2.163 | 4.642 | 2.185 | 359 |
| | | (*0.001*) | (*0.001*) | (*0.001*) | |
| | | [0.313] | [0.278] | [0.183] | |
| | Policy | 2.952 | 2.437 | 1.522 | 359 |
| | | (*0.001*) | (*0.001*) | (*0.001*) | |
| | | [0.713] | [0.626] | [0.581] | |
| First differences | Simple | 1.135 | 2.007 | 1.699 | 285 |
| | | (*0.035*) | (*0.002*) | (*0.001*) | |
| | | [0.319] | [0.184] | [0.192] | |
| | Policy | 1.446 | 2.065 | 0.663 | 285 |
| | | (*0.249*) | (*0.010*) | (*0.001*) | |
| | | [0.080] | [0.330] | [0.315] | |
| Levels | Simple | 1.848 | 4.813 | 1.838 | 359 |
| | | (*0.012*) | (*0.011*) | (*0.001*) | |
| | | [0.472] | [0.445] | [0.345] | |
| | Policy | 2.958 | 3.267 | 2.073 | 359 |
| | | (*0.001*) | (*0.001*) | (*0.001*) | |
| | | [0.346] | [0.155] | [0.180] | |

[a]Numbers in parentheses are *p*-values for the coefficient and numbers in brackets are *p*-values for the Sargan-test.

# S-GMM results for LLB (2000)

Table 5
Financial intermediation and growth: dynamic panel regressions, system estimator

| Regressors | (1) | (2) | (3) |
|---|---|---|---|
| Constant | 0.06 | − 5.677 | 4.239 |
| | *(0.954)* | *(0.001)* | *(0.001)* |
| Logarithm of initial income per capita | − 0.742 | − 0.117 | − 0.364 |
| | *(0.001)* | *(0.223)* | *(0.001)* |
| Government size[a] | − 1.341 | − 1.13 | − 1.987 |
| | *(0.001)* | *(0.001)* | *(0.001)* |
| Openness to trade[a] | 0.325 | 0.497 | 0.442 |
| | *(0.169)* | *(0.002)* | *(0.010)* |
| Inflation[b] | 1.748 | − 1.772 | − 0.178 |
| | *(0.001)* | *(0.001)* | *(0.543)* |
| Average years of secondary schooling | 0.78 | 0.638 | 0.639 |
| | *(0.001)* | *(0.001)* | *(0.001)* |
| Black market premium[b] | − 2.076 | − 1.044 | − 1.027 |
| | *(0.001)* | *(0.001)* | *(0.001)* |
| Liquid Liabilities[a] | 2.952 | | |
| | *(0.001)* | | |
| Comm. vs. Central Bank[a] | | 2.437 | |
| | | *(0.001)* | |
| Private Credit[a] | | | 1.522 |
| | | | *(0.001)* |
| Dummy 71–75 | − 1.074 | − 0.792 | − 0.959 |
| | *(0.001)* | *(0.001)* | *(0.001)* |
| Dummy 76–80 | − 1.298 | − 0.825 | − 1.177 |
| | *(0.001)* | *(0.001)* | *(0.001)* |
| Dummy 81–85 | − 3.328 | − 2.616 | − 3.179 |
| | *(0.001)* | *(0.001)* | *(0.001)* |
| Dummy 86–90 | − 2.614 | − 1.894 | − 2.434 |
| | *(0.001)* | *(0.001)* | *(0.001)* |
| Dummy 91–95 | − 3.631 | − 2.77 | − 3.308 |
| | *(0.001)* | *(0.001)* | *(0.001)* |
| Sargan test[c] (*p*-value) | 0.713 | 0.626 | 0.581 |
| Serial correlation test[d] (*p*-value) | 0.588 | 0.957 | 0.764 |

*p*-values in parentheses

[a]In the regression, this variable is included as log(variable).
[b]In the regression, this variable is included as log(1 + variable).
[c]The null hypothesis is that the instruments used are not correlated with the residuals.
[d]The null hypothesis is that the errors in the first-difference regression exhibit no second-order serial correlation.

# Roodman (2009) critique

- Roodman (2009) in an OBES article presents evidence that the LLB (2000) results from the S-GMM estimator are not statistically robust, aside from any deeper identification problems.
  - They suffer from overfitting bias, due to too many instruments.
  - Their results hinge on the level equation information, rather than the differenced equation information alone.
- He does only focus on one of the three financial development measures used by LLB (2000). The others appear to be somewhat better behaved (recall PS2).
- However, it is suggestive that there may be many results in the literature using D/S-GMM which are reliant on questionable assumptions.

# Roodman (2009) critique

TABLE 3

*Tests of Levine et al. (2000) system GMM regressions of GDP per capita growth on private credit per GDP*

| | *Original* | *Reproduction* | *Reproduction, collapsed instruments* | *Reproduction, difference GMM* |
|---|---|---|---|---|
| Log private credit/GDP | 1.52 | 1.41 | 2.34 | 0.49 |
| | (0.001) | (2.04)** | (2.21)** | (0.33) |
| Log initial GDP per capita (PPP) | −0.36 | −0.13 | −0.37 | −8.95 |
| | (0.001) | (0.20) | (0.30) | (1.84)* |
| Mean years of secondary schooling | 0.64 | 0.09 | −0.31 | 0.47 |
| | (0.001) | (0.14) | (0.25) | (0.26) |
| Log government spending/GDP | −1.34 | −0.79 | −1.22 | 0.30 |
| | (0.001) | (0.62) | (0.54) | (0.12) |
| Log(1 + black market premium) | −2.08 | −1.35 | −1.17 | −2.40 |
| | (0.001) | (2.92)*** | (0.99) | (1.52) |
| Log(1 + inflation) | 1.75 | −0.17 | 3.29 | 0.85 |
| | (0.001) | (0.10) | (1.08) | (0.17) |
| Log(Imports + exports)/GDP | 0.33 | −0.19 | −0.64 | 1.89 |
| | (0.169) | (0.28) | (0.32) | (0.68) |
| Observations | 359 | 353 | 353 | 328 |
| Instruments | | 75 | 19 | 40 |
| Arellano–Bond test for AR(2) in differences ($p$-value) | 0.76 | 0.78 | 0.93 | 0.79 |
| Hansen test of joint validity of instruments ($p$-value) | 0.58 | 0.58 | 0.03 | 0.16 |
| Difference Sargan tests ($p$-values) | | | | |
| All system GMM instruments | | 0.75 | | |
| Those based on lagged growth only | | 0.97 | 0.001 | |

*Notes*: All regressions are two-step system GMM. Period dummies not reported, $p$-values clustered by country in parentheses in first column; in remaining columns, $t$ statistics clustered by country, and incorporating the Windmeijer (2005) correction, in parenthesis.
Significant at *10%, **5%, ***1%.

# Summary

- Linear dynamic panel models (with lagged dependent variables as explanatory variables) with unobserved, individual-specific effects pose special identification problems.
  - Strict exogeneity of the regressors fails.
- Under a weaker sequential exogeneity assumption, sufficiently deep lags of the dependent variable can allow us to construct valid instruments.
- Difference and system GMM estimators are the most common and readily available frameworks within which to address the issue.
  - Both are formulated on a large $N$, small $T$ assumption. For $T$ large, it may be better to use pooled OLS and treat each panel as a separate time series.
  - We should be vigilant for many and weak instrument problems which arise easily in this framework.