Dynamic panel data methods

for cross-section panels

Franz Eigner
University Vienna

Prepared for
UK Econometric Methods of Panel Data
with Prof. Robert Kunst

27th May 2009

Structure

1. Preliminary considerations
   - Dynamic modelling
   - Bias of the LSDV estimator
2. Consistent Estimation
   - GMM estimators
   - Bias corrected LSDV estimators
3. Application - Winter tourism demand model

Motivation
for linear dynamic panel models

Before: linear fixed effects model

$$y_{it} = \beta x_{it} + \mu_i + \varepsilon_{it}, \; u_{it} = \mu_i + \varepsilon_{it}$$

Now: e.g. including AR(1)

$$y_{it} = \rho y_{i,t-1} + \beta x_{it} + \mu_i + \varepsilon_{it}, \; u_{it} = \mu_i + \varepsilon_{it} \tag{1}$$

$\longrightarrow$ Allowing feedback from current or past shocks

Dynamic modelling adequate when

1. Temporal autocorrelation in the residuals ε_{it}
2. High persistency in the dependent variable y_{it}

Estimation methods

Dealing with temporal autocorrelation

- here: Inclusion of a dynamic component.
  Find consistent estimator for $N \longrightarrow \infty$ and T fixed
  (cross-section panel)

Alternative methods:

1. Parks method (FGLS)
2. Panel corrected standard errors (PCSE)
   Prais Winston Transformation

Inconsistency of the LSDV estimator
in dynamic panel models

LSDV estimator requires strict exogeneity assumption:

$$E(\varepsilon_{i,t} \mid x_i, \mu_i) = 0, \quad t = 1, \ldots, T;\ i = 1, \ldots, N$$

Violated by inclusion of $y_{i,t-1}$.

One can show:
$\widetilde{y}_{i,t-1}$ is negatively correlated with $\widetilde{\varepsilon}_{it}$, due to
$cor(y_{i,t-1}, -\frac{1}{T-1}\varepsilon_{i,t-1}) < 0$ and $cor(-\frac{1}{T-1}y_{it}, \varepsilon_{it}) < 0$, where
$\tilde{y}_{i,t-1} = y_{i,t-1} - \frac{1}{T-1}(y_{i2} + \ldots + y_{iT})$ and
$\tilde{\varepsilon}_{it} = \varepsilon_{it} - \frac{1}{T-1}(\varepsilon_{i2} + \ldots + \varepsilon_{iT})$

$\rightarrow$ LSDV estimator is inconsistent and biased in dynamic models (for $N \rightarrow \infty$ and fixed T)

Bias of the LSDV estimator

Nickell (1981) and Hsiao (2001)

$$\rho^* = \operatorname*{plim}_{N\to\infty}(\hat{\rho}_{lsdv} - \rho) = \frac{-\sigma^2_{\tilde{\varepsilon}} h(\rho, T)}{(1-\rho^2_{\tilde{x}\tilde{y}_{-1}})\sigma^2_{\tilde{y}_{-1}}}$$

$\beta^* = -\zeta\rho^*$, where $\zeta = \sigma_{\tilde{x}\tilde{y}_{-1}}/\sigma^2_{\tilde{x}}$

$h(\rho, T) = \frac{(T-1)-T\rho+\rho^T}{T(T-1)(1-\rho)^2}$ and $\rho_{\tilde{x}\tilde{y}_{-1}} = \sigma_{\tilde{x}\tilde{y}_{-1}}/\sigma_{\tilde{x}}\sigma_{\tilde{y}_{-1}}$

annot.: variables denoted as $\tilde{x}$ and $\tilde{y}$ are within-transformed

$h(\rho, T)$ is always positive $\rightarrow$ LSDV estimate is downward biased

Bias of the LSDV estimator

Bias is especially severe, when

1. the autoregressive coefficient ρ is high
2. the number of time periods T is low
3. the ratio of $\sigma^2_{\tilde{\varepsilon}}/\sigma^2_{\tilde{y}_{-1}}$ is high

First Difference IV
Anderson/Hsiao (1981)

- Eliminating μ_i by differencing (instead of within-transformation)

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

- In matrix notation:

$$F_y = F_{y-1}\rho + FX\beta + F\varepsilon$$

where $F = I_N \bigotimes F_T$ and $F_T = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$

First Difference IV
Andersion/Hsiao (1981)

However $\triangle y_{i,t-1}$ is now correlated with the error term $\triangle \varepsilon_{i,t-1}$
$\rightarrow$ using IV method with $y_{i,t-2}$ as instrument for $\triangle y_{i,t-1}$, because

$$E(y_{i,t-2} \triangle \varepsilon_{it}) = 0$$

$\rightarrow$ inefficient, because not all information, e.g. $\triangle \varepsilon_{it} \sim MA(1)$, is used

Difference-GMM
Arellano and Bond (1991)

Efficient estimates are obtained using a GMM framework.
Following moments are exploited:

$$E[y_{i,t-s}\triangle\varepsilon_{it}] = 0 \text{ and } E[X_{i,t-s}\triangle\varepsilon_{it}] = 0 \text{ for } s \geq 2;\, t = 3, \dots T$$

$$X_i = \begin{bmatrix} y_{i2} - y_{i1} & x'_{i3} - x'_{i2} \\ y_{i3} - y_{i2} & x'_{i4} - x'_{i3} \\ \vdots & \vdots \\ y_{i,T-1} - y_{i,T-2} & x'_{iT} - x'_{i,T-1} \end{bmatrix}$$

$$Z_i = \begin{bmatrix} [y_{i1}, x'_{i1}, x'_{i2}] & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [y_{i1}, \dots, y_{i,T-2}, x'_{i1}, \dots, x'_{i,T-1}] \end{bmatrix}$$

$X = (y_{-1}, X)$, $Z = (Z'_1, Z'_2, \dots, Z'_N)'$, $\gamma' = (\rho, \beta')$

Diff-GMM
Idea of the GMM-Framework

L instruments imply a set of L moments, i.a. $g_i(\hat{\beta}) = y_{i,t-s} \triangle \varepsilon_{it}$, where exogeneity holds when $E(g_i(\beta)) = 0$. Each of the L moment equations corresponds to a sample moment $\bar{g}(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta})$.
Estimator is typically obtained by solving $\bar{g}(\hat{\beta}) = 0$.
but here: model overidentification $\longrightarrow$ minimizing the criterion J_N:

$$J_N(\hat{\beta}) = \bar{g}(\hat{\beta})' \hat{W}_N \bar{g}(\hat{\beta})$$

or in our notation:

$$J_N = \left(\frac{1}{N} \sum_{i=1}^{N} \triangle \varepsilon_i' Z_i \right) \hat{W}_N \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \triangle \varepsilon_i \right)$$

where $\hat{W}_N$ is the estimated weighting matrix.

Diff-GMM

Estimation of the weighting matrix

→ Optimal weighting matrix is the inverse of the moment covariance matrix:

$$W_N = Var(Z'\triangle\varepsilon)^{-1} = (Z'\Omega Z)^{-1}$$

Unless Ω is known, efficient GMM is not feasible → two-step procedure
First replace Ω with some simple G (here: assuming ε_{it} i.i.d.)

$$\hat{W}_{1N} = \left(\sum_{i=1}^{N} Z_i' G_T Z_i\right)^{-1} = (Z'GZ)^{-1}$$

where $G = (I_N \otimes G_T')$ and $G_T = F_T F_T' = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$

Diff-GMM
Estimation

... delivers (consistent) first-step estimates. Its residuals $\triangle\hat{\varepsilon}_{1i}$ are used for the two-step estimation of $\hat{W}$.

$$\hat{W} = \left(\sum_{i=1}^{N} Z_i' \triangle\hat{\varepsilon}_{1i} \triangle\hat{\varepsilon}_{1i}' Z_i \right)^{-1}$$

Efficient estimates for the Diff-GMM are then obtained with:

$$\hat{\gamma}^{EGMM} = \left(X'Z\hat{W}Z'X \right)^{-1} X'Z\hat{W}Z'y$$

One can show: under homoskedasticity, one-step estimates are asymptotically equivalent to the two-step estimates.

However:
If y_{it} is highly persistent, instruments are weak. (Blundell and Bond, 1998, Kitazawa, 2001)

Diff-GMM
Estimating variances

- one-step

replacing Ω with a sandwich-type proxy $\hat{\Omega}_{\beta_1}$ delivers consistent and robust variances.

$$\widehat{Var}[\hat{\beta}_1] = \left(X'Z\hat{W}_1Z'X\right)^{-1} X'Z\hat{W}_1Z'\hat{\Omega}_{\beta_1}Z\hat{W}_1Z'X\left(X'Z\hat{W}_1Z'X\right)^{-1}$$

- two-step

using the optimal weighting matrix $W = (Z'\Omega Z)^{-1}$, above formula reduces to

$$\widehat{Var}[\hat{\beta}_2] = \left(X'Z(Z'\hat{\Omega}_{\beta_1}Z)^{-1}Z'X\right)^{-1}$$

however: $\widehat{Var}[\hat{\beta}_2]$ can be heavily downward biased $\rightarrow$ Windmeyer's correction (2005)

System-GMM
Blundell and Bond (1998)

System of two equations
»level equation« and »difference equation«

Additional moments are explored:
$E[\triangle y_{i,t-1}(\mu_i + \varepsilon_{it})] = 0$ and
$E[\triangle X_{i,t-1}(\mu_i + \varepsilon_{it})] = 0, \text{ for } t = 3, \ldots, T$

> *»where Arellano-Bond instruments differences [. . .] with levels, Blundell-Bond instruments levels with differences. [...] For random walk–like variables, past changes may indeed be more predictive of current levels than past levels are of current changes« (Roodman, 2006)*

→ reduction of weak instrument problem

System-GMM

Matrix notation

$$X_i = \begin{bmatrix} y_{i2} - y_{i1} & x'_{i3} - x'_{i2} \\ y_{i3} - y_{i2} & x'_{i4} - x'_{i3} \\ \vdots & \vdots \\ y_{i,T-1} - y_{i,T-2} & x'_{iT} - x'_{i,T-1} \\ y_{i2} & x'_{i2} \\ \vdots & \vdots \\ y_{i,T-1} & x'_{iT} \end{bmatrix} \quad Z_i = \begin{bmatrix} Z_i^D & 0 \\ 0 & Z_i^L \end{bmatrix}$$

$$Z_i^L = \begin{bmatrix} [\triangle y_{i2}, \triangle x'_{i2}, \triangle x'_{i3}] & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & [\triangle y_{i2}, \ldots, \triangle y_{i,T-2}, \triangle x'_{i,2}, \ldots, \triangle x'_{i,T}] \end{bmatrix}$$

Validity of the instruments
Assumptions and tests

Performance depends strongly on the validity of the instruments.

A valid instrumental variable z requires

1) $E[\varepsilon \mid z] = 0$ (exogeneity)
2) $cov(z, x) \neq 0$ (relevance)

- Overidentifying restrictions test (Sargan/Hansen Test)
  Hansen: $J(\hat{\beta}_{EGMM}) \sim \chi^2_{L-K}$
  Sargan: as Hansen but under conditional homoskedasticity
  Difference-in-Sargan: testing subset of instruments
  $DS = S_u - S_r \sim \chi^2$
- Arellano and Bond - Autocorrelation test

More instruments increase finite sample bias (Bun/Kiviet, 2002)
$\rightarrow$ trade-off between small sample bias and efficiency

Bias corrected LSDV estimators

Consistent estimation by additive bias correction
Estimating the extent of the bias

- by using a preliminary consistent estimator
  Kiviet (1995)
  Hansen (2001)
  Hahn and Kuersteiner (2002) - not for short T
  Bruno (2005) - for unbalanced panels and short T

- without using a preliminary consistent estimator
  Bun/Carree (2005)

Bruno (2005)

Bias approximations emerge with an increasing level of accuracy. $B_1 = c_1(\bar{T}^{-1})$, $B_2 = B_1 + c_2(N^{-1}\bar{T}^{-1})$ and $B_3 = B_2 + c_3(N^{-1}\bar{T}^{-2})$ where c_1, c_2 and c_3 depend i.a. on σ_ε^2 and γ.

$\rightarrow$ they are not yet feasible. σ_ε^2 and γ have to obtained from a consistent estimator. (AH, AB, BB)

$$LSDVC_i = LSDV - \hat{B}_i, \; i = 1,2 \text{ and } 3$$

Summary of the models

Model	Transformation	Regressors	Consistency
LSDV/FE	Within	$y_{i,t-1}, x_{it}$	no
Bias corrected LSDV	Within	$y_{i,t-1}, x_{it}$	yes
First-difference IV	$\triangle$	$\triangle y_{i,t-1}, \triangle x_{it}$	yes
First-difference GMM	$\triangle$	$\triangle y_{i,t-1}, \triangle x_{it}$	yes
System-GMM	$\triangle$	$\triangle y_{i,t-1}, \triangle x_{it}, y_{i,t-1}, x_{it}$	yes

Comparison of performance according to Monte Carlo Simulations:

1. GMM more adequate for large N
2. Bias corrected LSDV performs better for small data sets

Modelling winter tourism demand for Austrian ski destinations from 1973 to 2006

Cross section panel data with N=185[1] and T=34

Nights	number of overnight stays in winter season
Snow[2]	snow cover
GDP[3]	income variable
Beds[4]	infrastructure variable
PP[5]	relative purchasing power

[1] Austrian ski resort database. JOANNEUM Research (2008)
[2] ZAMG (2009)
[3] OECD (2008)
[4] Statistik Austria (2008)
[5] OECD (2008)

Commands in statistical software packages

With STATA

- GMMs: `xtabond/xtdpdsys` or `xtabond2` (Roodman, 2006)
- Bias corrected LSDV: `xtlsdvc` (Bruno, 2005)
- Cross section autocorrelation test (Pesaran, 2004): `xtcsd`

With R

- Package `plm` (Yves/Giovanni, 2008) contains i.a. function
- `pgmm` for GMMs

other: SAS, LIMDEP, ...

Estimation table

Winter tourism demand for Austrian ski destinations from 1973 to 2006

	pool	fe	fe_tw	fe_tw_bc	diffgmm2	sysgmm2	sysgmm_v	sysgmm_g
L.NIGHTS	0.716***	0.609***	0.596***	0.637***	0.475***	0.634***	0.600***	0.632***
	(11.97)	(10.54)	(10.62)	(67.01)	(6.35)	(11.63)	(4.64)	(11.78)
L2.NIGHTS	0.215***	0.174***	0.187***	0.161***	0.166***	0.163***	0.268***	0.179***
	(3.92)	(3.82)	(4.36)	(16.72)	(5.44)	(3.93)	(3.69)	(3.78)
SNOW / 100	0.067***	0.076***	0.070***	0.071***	0.071***	0.097***	0.153***	0.095***
	(5.70)	(6.49)	(4.04)	(4.44)	(3.65)	(4.41)	(2.70)	(5.35)
log(BEDS)	0.086***	0.113***	0.132***	0.119***	0.202***	0.202***	0.134	0.222***
	(7.40)	(5.53)	(5.80)	(10.34)	(4.41)	(6.19)	(1.38)	(7.87)
log(GDP)	0.039	0.013	0.407***	0.436***	0.977	0.700	0.361	0.663***
	(0.73)	(1.42)	(3.29)	(5.85)	(1.45)	(1.49)	(0.97)	(2.72)
log(PP)	-0.041***	-0.035***	-0.030**	-0.028**	-0.024	-0.010	-0.056	0.010
	(-3.35)	(-2.76)	(-2.31)	(-2.36)	(-0.38)	(-0.16)	(-1.17)	(0.30)
R2_within		0.776	0.785					
corr(x_i,mu_i)		0.954	0.924					
sigma_u		0.193	0.174					
sigma_e		0.147	0.145					
rho		0.632	0.592					
Pesaran AR		74.3	2.8					
Pesaran p_value		0.000	0.005					
t-statistics	Robust	Robust	Robust		Corrected	Corrected	Corrected	Corrected
F	15133.0	1129.5	355.7		102.7	37285.5	274410.0	211497.0
diff AR(2)					0.621	0.901	0.345	0.926
Sargan test					0.000	0.000	0.000	0.000
Hansen test					1.000	1.000	0.202	1.000
Diff_Sarg IV						1.000	0.752	1.000
Diff_Sarg GMM						1.000	0.510	1.000
No. of instruments					562	594	147	893
No. of groups	185	185	185		185	185	185	185
No. of observations	5920	5920	5920	5920	5735	5920	5920	5920

* p<0.1, ** p<0.05, *** p<0.01

Instrument list for the GMM estimators

Equation	Type	DIFF_GMM	SYS_GMM	SYS_GMM_valid	SYS_GMM_gdp
First difference equation	IV	Diff. (SNOW log_PP log_GDP log_BEDS) time_dummies	Diff. (SNOW log_PP log_GDP log_BEDS) time_dummies	-	Diff. (log_BEDS SNOW log_PP) time dummies
	GMM	Lag(2-.). log_NIGHTS	Lag(2-.). log_NIGHTS	Lag(4-7). log_NIGHTS	Lag(2-.). log_NIGHTS log_GDP
Level equation	IV	-		log_GDP SNOW log_PP	log_GDP SNOW log_PP
	GMM	-	Diff.Lag. log_NIGHTS	Diff.(Lag(3). log_NIGHTS)	Diff.Lag.(log_NIGHTS log_GDP)

Final considerations

concerning the tourism demand model estimates

$\longrightarrow$ Biases in the estimates seem to follow the theory

however some open questions concerning the

1. consequence of cross section dependence
   Potential loss in efficiency (Phillips and Sul, 2003)
2. validity of the instruments
3. choice of the best estimator