Anova-type consistent estimators of variance components in unbalanced multi-way error components models

Giovanni S. F. Bruno
Department of Economics,
Bocconi University
giovanni.bruno@unibocconi.it

VII Convegno Italiano degli Utenti di Stata, Bologna 11-12 Novembre 2010

Structure of the presentation

1. Motivations
2. Anova-type estimators
3. My contribution
4. Empirical application to US state level production data
5. Conclusions

Motivations

Most economic data are multi-dimensional and with an unbalanced structure. Just a few examples:

- Employers/employees/years (Abowd et al. (1999));
- States/regions/years [Baltagi et al.(2001)Baltagi, Song, and Jung];
- sectors/firms/years [Arellano and Bond(1991)];
- Exporters/products/years [Boumahdi et al.(2006)Boumahdi, Chaaban, and Thomas];
- movies/theaters/weeks [Davis(2002)]

Motivations

Failing to accommodate the multilevel structure may lead to inaccurate and imprecise inference (biased standard errors [Moulton(1990)], biased/inefficient coefficient estimates).

Feasible GLS (FGLS) estimators of multiway error components models (ECM's) ensure efficiency, robust standard errors and computational simplicity ([Moulton(1990)] [Wooldridge(2003)] [Donald and Lang(2007)])

Motivations

The model:

$$y = X\beta + \Delta_1 u_1 + \ldots + \Delta_m u_m + u_0. \tag{1}$$

The covariance matrix of the composite error $\Delta_1 u_1 + \ldots + \Delta_m u_m + u_0$:

$$\Sigma = \sigma_0^2 I_n + \sigma_1^2 \Delta_1 \Delta_1' + ... + \sigma_m^2 \Delta_m \Delta_m'. \tag{2}$$

An estimate of Σ is required for implementing the FGLS estimator

$$\hat{\beta} = \left(X' \hat{\Sigma}^{-1} X \right)^{-1} X' \hat{\Sigma}^{-1} y.$$

Motivations

- Stata `xtreg` implements FGLS through Anova estimators only for the one-way random effect panel data model (see [XT] **xtreg**)
- Stata `xtmixed` is specifically designed for multilevel models, but restricts to maximum likelihood methods and does not support endogenous regressors (see [XT] **xtmixed**)

Motivations

Few results accomodating unbalancedness, non-normality and endogeneity:

- [Wansbeek and Kaptein(1989)] derive Anova-type unbiased estimators of variance components, but focus on the two-way unbalanced ECM with strictly exogenous regressors. No proof of consistency.
- [Baltagi and Chang(2000)] derive Anova-type estimators restricting to the one-way model with endogenous regressors. Consistency is only conjectured but not proved.
- [Davis(2002)] derives Anova-type unbiased estimators for the multiway model with endogenous regressors, but unbiasedness depends on an arbitrary conditional homoskedasticity assumption and, however, the formulas are not complete. No proof of consistency.

Anova-type estimators of variance components

The standard procedure to obtain unbiased Anova estimators equates quadratic forms in residuals estimates to their conditional expectations. It places strong restrictions on conditional moments ([Searle(1971)], [Swamy and Arora(1972)], [Westfall(1986)], [Wansbeek and Kaptein(1989)], [Baltagi and Chang(1994)] and [Davis(2002)]).

An alternative procedure ([Wallace and Hussain(1969)]; [Amemiya(1971)]; [Baltagi and Chang(2000)]) leaves the conditional expectation and the conditional covariance matrix unrestricted. Estimators are computationally simpler and can be used in IV estimation

My contribution

I derive three new Anova-type consistent estimators (ACE's) following Procedure II. This generalizes the existing estimators to allow for

- generic multi-dimensional and unbalanced data structure,
- non-normal error components
- endogeneity of regressors.

All estimators are proved to be consistent as the number of groups in each dimension tends to infinity: the number of columns in each Δ_i is large. Mata is required for computation.

My contribution

The first estimator (ACE1) is based upon (two stages) within residuals. It extends the estimator derived by [Amemiya(1971)]. In its use of within residuals, it is closely related to that by [Wansbeek and Kaptein(1989)].

The second estimator (ACE2) uses (two stages) least squares residuals. It extends [Wallace and Hussain(1969)] (WH) and [Baltagi and Chang(2000)].

The third estimator (ACE3) uses (two stages) within and between residuals. It is an adaptation of [Swamy and Arora(1972)] (SA).

ACE1 is derived as follows. The WTSLS residuals are

$$\widehat{\epsilon}_w = \left\{ I - X \left[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \right]^{-1} X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} \right\} \epsilon$$

Then, **ACE1** of σ_0^2 is

$$\widehat{\sigma}^2_{w,0} = \frac{\widehat{\epsilon}'_w Q_{[\Delta]} \widehat{\epsilon}_w}{n - r(\Delta)} \tag{3}$$

and **ACE1 of** $\sigma^2 \equiv \left(\sigma_1^2, ..., \sigma_m^2 \right)$ is

$$\widehat{\sigma}^2_w = A^{-1} \left(\widehat{B}_w - \widehat{\sigma}^2_{w,0} C \right), \tag{4}$$

where

$$A = \frac{1}{n} \begin{pmatrix} n & tr\ \Delta'_i P_{[\Delta_{\mathbf{1}}]} \Delta_i & ...tr\ \Delta'_m P_{[\Delta_{\mathbf{1}}]} \Delta_m \\ \vdots & \vdots & \vdots \\ tr\ \Delta'_1 P_{[\Delta_{\mathbf{i}}]} \Delta_1 & n & tr\ \Delta'_m P_{[\Delta_{\mathbf{i}}]} \Delta_m \\ \vdots & \vdots & \vdots \\ tr\ \Delta'_1 P_{[\Delta_{\mathbf{m}}]} \Delta_1 ... & tr\ \Delta'_i P_{[\Delta_{\mathbf{m}}]} \Delta_i & n \end{pmatrix},$$

$$\widehat{B}_w = \frac{1}{n}\left(\widehat{\epsilon}'_w P_{[\Delta_1]}\widehat{\epsilon}_w \quad \ldots \; \widehat{\epsilon}'_w P_{[\Delta_m]}\widehat{\epsilon}_w\right)' \text{ and } C = \frac{1}{n}\left(\; trP_{[\Delta_1]} \quad \ldots \quad trP_{[\Delta_m]} \; \right)'.$$

```
* This do file yields the Cobb-Douglas estimates in Baltagi et al. (2001)
* The 2-way nested FGLS estimator is implemented by estimating variance
* components through the Wansbeek and Kaptein (1989) method

use productivity.dta                                              /// Open Munnel (1990) data

sort region state year

mata                                                              /// call mata

st_view(y=.,.,"gsp")                                              /// y

st_view(Xs=.,.,("private","emp","hwy", "water", "other", "unemp"))      /// X

st_view(s1=.,.,"state")                                           /// state dimension

st_view(s2=.,.,"region")                                          /// region dimension



info1 =panelsetup(s1,1)

D1=J(rows(y),rows(info1),0)                                       /// Dummy variables matrix
                                                                  /// for s1 (state)
for (n=1; n<=rows(info1); n++)
A=J(info1[n,2]-info1[n,1]+1,1,1)
D1[|info1[n,1],n\info1[n,2],n|]=A


P1=D1*invsym(cross(D1,D1))*D1´                                    /// Projection matrix


   A=(k11, k1221, k22)                                           // set up the linear
   B=(B1 - s2_0*k12 - s2_0*k2)                                   // system

   C=svsolve(A,B)

   s2_1=C[1]                                                      // variance-comp. est.
   s2_2=C[2]

V=invsym(s2_0*I(rows(y))+s2_1*D1*D1´+s2_2*D2*D2´)               // inv. cov. matrix
```

Table: **Cobb-Douglas production function estimates with state and region effects**

	OLS	FGLS ACE1	FGLS WK	FGLS ACE2	FGLS WH	FGLS ACE3	FGLS SA
Const.	1.926 (0.053)	2.133 (0.162)	2.131 (0.160)	2.076 (0.150)	2.082 (0.152)	2.093 (0.143)	2.089 (0.144)
K	0.312 (0.011)	0.264 (0.022)	0.264 (0.027)	0.276 (0.021)	0.273 (0.021)	0.274 (0.020)	0.274 (0.020)
L	0.550 (0.016)	0.760 (0.027)	0.758 (0.027)	0.735 (0.027)	0.742 (0.026)	0.740 (0.025)	0.740 (0.025)
KH	0.059 (0.015)	0.072 (0.024)	0.072 (0.024)	0.073 (0.023)	0.075 (0.023)	0.072 (0.022)	0.073 (0.022)
KW	0.119 (0.012)	0.076 (0.014)	0.076 (0.014)	0.077 (0.014)	0.076 (0.014)	0.076 (0.014)	0.076 (0.014)
KO	0.009 (0.012)	-0.102 (0.017)	-0.102 (0.017)	-0.092 (0.018)	-0.095 (0.017)	-0.095 (0.017)	-0.094 (0.017)
Un	-0.007 (0.001)	-0.006 (0.001)	-0.006 (0.001)	-0.006 (0.001)	-0.006 (0.001)	-0.006 (0.001)	-0.006 (0.001)
σ_ε^2	0.0073	0.0014	0.0014	0.0015	0.0014	0.0014	0.0014
σ_u^2	-	0.0024	0.0022	0.0017	0.0027	0.0013	0.0015
σ_v^2	-	0.0072	0.0069	0.0043	0.0045	0.0044	0.0043

Table: **Cobb-Douglas production function estimates with state, region and interacted time-region effects**

	ACE1	ACE1	WK	ACE2	WH	ACE3	SA
Const.	-	2.297	2.286	2.154	2.159	2.201	2.198
		(0.181)	(0.177)	(0.151)	(0.154)	(0.146)	(0.146)
K	0.158	0.198	0.201	0.236	0.233	0.223	0.223
	(0.028)	(0.023)	(0.023)	(0.021)	(0.021)	(0.021)	(0.021)
L	0.814	0.798	0.794	0.749	0.756	0.758	0.758
	(0.031)	(0.028)	(0.028)	(0.027)	(0.027)	(0.026)	(0.026)
KH	0.080	0.071	0.071	0.078	0.079	0.078	0.079
	(0.027)	(0.025)	(0.024)	(0.023)	(0.023)	(0.022)	(0.022)
KW	0.032	0.047	0.048	0.052	0.053	0.046	0.046
	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
KO	-0.023	-0.048	-0.049	-0.050	-0.053	-0.042	-0.041
	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Un.	-0.001	-0.003	-0.003	-0.004	-0.004	-0.003	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
σ_ε^2	0.0009	0.0009	0.0009	0.0011	0.0010	0.0009	0.0009
σ_u^2	-	0.0048	0.0041	0.0016	0.0027	0.0013	0.0014
σ_v^2	0.0099	0.0099	0.0090	0.0044	0.0045	0.0044	0.0043
σ_z^2	-	0.0006	0.0006	0.0004	0.0004	0.0007	0.0007

Conclusions

- Three new ACE's of variance components are derived for general unbalanced multi-way error components models with possibly non-normal disturbances and endogenous regressors. Stata Do files are provided, which can be easily adpated to the specific users needs.
- They are easy to compute and are proved to be consistent under mild regularity conditions on the data generating process. The empirical application (along with Montecarlo experiments) show that the new ACE's perform well in comparison to the unbiased methods incorporating finite sample corrections.
- Future research: A general Stata code for multiway FGLS; efficient algorithms for the computation of the inverse; specification tests.

T. Amemiya.
The estimation of the variances in a variance-components model.
International economic review, 12(1):1–13, 1971.

M. Arellano and S. Bond.
Some tests of specification for panel data: Monte carlo evidence and an application to employment equations.
Review of Economic Studies, 58:277–297, 1991.

B. H. Baltagi and Y.-J. Chang.
Incomplete panels.
Journal of Econometrics, 62:67–89, 1994.

B. H. Baltagi and Y.-J. Chang.
Simultaneous equations with incomplete panels.
Econometric Theory, 16:269–279, 2000.

Badi H. Baltagi, Seuck H. Song, and Byoung C. Jung.
The unbalanced nested error component regression model.
Journal of Econometrics, 101:357–381, 2001.

Rachid Boumahdi, Jad Chaaban, and Alban Thomas.

Import demand estimation with country and product effects: Application of multi-way unbalanced panel data models to lebanese imports.
In Badi H. Baltagi, editor, *Panel Data Econometrics*. Elsevier, 2006.

Peter Davis.
Estimating multi-way error components models with unbalanced data structures.
Journal of Econometrics, 106:67–95, 2002.

Stephen G. Donald and Kevin Lang.
Inference with difference-in-differences and other panel data.
The Review of Economics and Statistics, 89(2), 2007.

Brent R. Moulton.
An illustration of a pitfall in estimating the effects of aggregate variables on micro units.
The Review of Economics and Statistics, 72(2):334–38, 1990.

A. Munnell.
Why has productivity growth declined? productivity and public investment.
New England Economic Review, pages 3–22, 1990.

S. R. Searle.

Linear Models.
New York: Wiley, 1971.

P. A.V. B. Swamy and S. S. Arora.
The exact finite sample properties of the estimators of coefficients in the error components regression models.
Econometrica, 40(2):261–275, 1972.

T. D. Wallace and Ashiq Hussain.
The use of error components models in combining cross section with time series data.
Econometrica, 37(1):55–68, 1969.

Tom Wansbeek and Arie Kaptein.
Estimation of the error components model with incomplete panels.
Journal of Econometrics, 41:341–361, 1989.

Peter H. Westfall.
Asymptotic normality of the anova estimates of components of variance in the nonnormal, unbalanced hierarchal mixed model.
The Annals of Statistics, 14:1572–1582, 1986.

Jeffrey M. Wooldridge.
Cluster-sample methods in applied econometrics.
American Economic Review - Papers and Proceedings, 93(2), 2003.