CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Anova-Type Consistent Estimators of Variance Components in Unbalanced Multi-Way Error Components Models

Giovanni Bruno

CEA@Cass Working Paper Series

WP–CEA–02-2010

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

Giovanni S. F. Bruno
Department of Economics,
Bocconi University,
Via Roentgen 1, 20136 Milano
giovanni.bruno@unibocconi.it

October 3, 2010

Abstract

This paper introduces three new Anova-type consistent estimators of variance components for use in multi-way unbalanced error components models, with possibly non-normal errors and endogenous regressors. They are easy to compute and are proved to be consistent under mild regularity conditions. For the first time proofs of consistency for Anova estimators are offered under such a general class of models, providing novel insights into the impact of unbalancedness on the large-sample properties of the estimators. A battery of Monte Carlo experiments and an empirical application to US production data show that the estimators perform reasonably well, in comparison to unbiased methods incorporating finite-sample corrections.

JEL Code: C23.

Keywords: variance components; Anova-type estimators; multi-way error components models; unbalancedness; endogenous regressors.

1 Introduction

This paper derives new Analysis of variance (Anova) estimators of variance components for use in multi-way unbalanced error components models (ECM's), with possibly non-normal errors and endogenous regressors.

Anova-type estimators are widely used in ECM's, despite their non-optimality in unbalanced designs, because of their computational simplicity and reasonable accuracy, both when estimates of the variance components are of direct interest and as ingredients of feasible GLS (FGLS) estimators for the model coefficients. Importantly, they are amenable to use in multi-way ECM's with endogenous regressors, as I show in this paper. FGLS estimation of multi-way ECM's, on its part, ensures efficiency gains over standard least squares techniques, along with more robust standard errors and t statistics (Wooldridge (2003)).

Two closely related procedures are typically applied in the ECM literature for obtaining Anova-type estimators of variance components. The first, which I call Procedure I, gives Anova-type unbiased estimators as the final output of the following three steps (see Searle (1971), Swamy and Arora (1972), Westfall (1986), Wansbeek and Kaptein (1989), Baltagi and Chang (1994) and Davis (2002), among others).

1. Start from theoretical quadratic forms in the unobserved composite error and obtain their empirical analogs by replacing the composite error with the residuals from some regression of choice.
2. Work out the conditional expectations of the empirical quadratic forms as

linear functions of the unknown variance components.

3. Equate the values of the empirical quadratic forms to their conditional expectations and obtain the Anova-type estimator as the system solution.

The resulting Anova-type estimator is unbiased by construction, inheriting the finite sample corrections incorporated into the conditional expectations formulas. It must be observed, though, that step 2 is successfully accomplished only under suitable conditional homoskedasticity restrictions, which may be difficult to justify on economic grounds, especially when the conditioning set includes endogenous regressors.

Anova-type consistent estimators (ACE's) can be obtained as the final outcome of the following scheme, referred to throughout as Procedure II (Wallace and Hussain (1969), Amemiya (1971), Baltagi and Chang (2000)).

1. Start from theoretical quadratic forms in the unobserved composite error and work out their unconditional expectations as linear functions of the unknown variance components.
2. Equate the theoretical quadratic forms to their unconditional expectations and obtain the unfeasible Anova estimator as the solution of the system.
3. Obtain an ACE by replacing the composite error in the formula of the unfeasible estimator with the residuals from some regression of choice.

Clearly, neither procedure is per se a consistency proof. Consistency of the resulting estimators must be suitably proved, and the regularity conditions under which it holds made explicit, in either scheme. Procedure II, however, has two virtues over Procedure I. First, it leaves the conditional expectation and conditional covariance of the composite error unrestricted, and so can be applied to a broader range of economic data. Second, since the formula of the unconditional expectation is less intricate than that of the conditional expectation, ACE's obtained through Procedure II are less computationally demanding. Procedure II has been introduced in balanced two-way ECM's with exogenous regressors and normal disturbances by Wallace and Hussain (1969) (WH), with the last step implemented through pooled ordinary least squares (OLS) residuals. Later, Amemiya (1971) (AM) has applied it to the same context as WH, in a version using within residuals. More recently, Baltagi and Chang (2000) have adapted the WH approach to the unbalanced, one-way, ECM with endogenous regressors, suggesting an estimator based on the pooled two stages least squares (TSLS) residuals, but without offering a proof of consistency.[1]

I apply Procedure II to derive three new ACE's for use in unbalanced multi-way ECM's with possibly non-normal error components and, as recommended by Baltagi et al. (2002), endogenous regressors. Consistency is proved, and regularity conditions made explicit, for each new ACE presented. It is an important methodological contribution of this paper that for the first time consistency is proved for Anova-type estimators within such a general class of ECM's. This is no trivial task, at this level of generality, so that a regularity condition is required to prevent abnormal patterns of unbalancedness. More specifically, a mild restriction is maintained preventing that for all dimensions in the sample, a dimension specific measure of unbalancedness, based on the Pearson's coefficient of variation across group sizes, may grow at the same speed as the number of groups in the dimension.[2] This provides novel theoretical insight into the impact of unbalancedness on accuracy and precision of Anova estimates, as well

---

[1] The distinction between the two procedures is recognized in Maddala and Mount (1973). Notice that they do not refer to the WH and AM estimators as Anova-type estimators, keeping this name only for the unbiased estimators derived under Procedure I.

[2] In more specific treatments this is accomplished by placing restrictions directly on the group sizes. For example Westfall (1986), proving asymptotic normality of the Anova estimator in the non-normal case for an unbalanced, nested ECM (with nonstochastic regressors), maintain that group sizes are uniformly bounded, which is far more restrictive than my condition.

as a new theoretically grounded measure to quantify such an impact.

The three ACE's differ only in the residuals used in the last step of Procedure II. The first, henceforth ACE1, is based on within two stages least squares (WTSLS) residuals and as such extends the ACE derived by AM. The second, henceforth ACE2, uses TSLS residuals, extending the ACE's introduced by WH and Baltagi and Chang (2000). The last, henceforth ACE3, uses WTSLS and between TSLS (BTSLS) residuals. It is related to the unbiased estimator derived by Swamy and Arora (1972) (SA) for the balanced non-nested two-way ECM with nonstochastic regressors, using within and between residuals in step 1 of Procedure I.

An estimator closely related to ACE1 is that derived by Wansbeek and Kaptein (1989) (WK) for the unbalanced two-way non-nested ECM, based on within residuals applied in step 1 of Procedure I. In WK, Step 2 is carried out under the assumption that regressors and the composite error are independent. The resulting estimator is clearly unbiased, but no proof of its consistency is given. In this respect, I show that under a stronger version of the regularity conditions of ACE1, requiring that the regressors and the composite error be not related asymptotically, the multi-way extension of the WK estimator, derived in this paper, is indeed consistent. I also prove that it is still unbiased under the assumption of strictly exogenous regressors and conditional homoskedasticity, which is clearly a weaker restriction than independence.

Another estimator closely related to ACE1 is that by Davis (2002), who extends the WK estimator to multi-way ECM's with endogenous regressors. While neither unbiasedness, nor consistency is therein proved, I show that the conditions needed to make Davis' estimator unbiased are by necessity of an arbitrary nature, given the presence of endogenous regressors. I prove also that the finite-sample correction in the formula of Davis' estimator is incomplete.

Finally, with the asymptotic analysis of ACE1 in hand, it is easy to prove that Davis' estimators is consistent under the regularity conditions of ACE1.

Along with ACE3, I also derive an unbiased estimator, based on Procedure I, that extends the SA estimator to multi-way, unbalanced ECM's with strictly exogenous regressors and conditional homoskedasticity. The only existing extensions of the SA estimator, to date, are confined to the unbalanced one-way model (Baltagi and Chang (1994)) and to the unbalanced two-way nested model (Baltagi et al. (2001)), (with regressors in both contributions taken, implicitly or explicitly, as nonstochastic).

For all estimators a number of computational issues are taken up, including the treatment of nested multi-way ECM's (see Westfall 1986 and Baltagi et al. (2001)) and that of regressors and instruments with slow sample variation, as the intercept or dummy variables.

The finite-sample performance of the new estimators is examined in a battery of Monte Carlo experiments, especially focused on the interplay between unbalancedness, number of groups and sample size. The new estimators are also seen at work in an empirical application based on Baltagi et al. (2001), estimating US state level production functions on the data by Munnell (1990).

The following notation and conventions are used throughout. Given an arbitrary real matrix A, the collection of all vectors that are linear combinations of the columns of A is referred to as the range of A and is denoted by $\mathcal{R}(A)$. The rank of A is denoted by $r(A)$, the determinant by $\det(A)$ and the trace by $tr\ A$. A^{-} denotes a generalized inverse of A. $P_{[A]} \equiv A(A'A)^{-}A'$ stands for the projection operator onto $\mathcal{R}(A)$ with the inner product $(a,b) = a'b$. I_s, s any integer, indicates the identity matrix of order s and I indicates a conformable identity matrix when the order can be easily inferred by the context. $Q_{[A]} \equiv I - P_{[A]}$ indicates the projection matrix onto the subspace of all vectors that are orthogonal to A. The symbol $*$ denotes the Hadamard product. Convergence in probability is denoted by $\underset{p}{\to}$.

The structure of the rest of the paper is as follows. Section 2 sets up the model. The ACE's are introduced, described and proved to be consistent in Section 3, which is the most important part of the paper. Section 4 focuses on the unbiased estimators. The Monte Carlo experiments are described in Section 5. Section 6 contains the empirical application. Section 7 concludes the paper. All lemmata and the proofs of all theorems are relegated to the appendix of the paper.

2 The model

I focus on the general multi-way ECM with $m+1$ dimensions

$$y = X\beta + \epsilon \tag{1}$$

where

$$\begin{aligned}
\epsilon &\equiv \Gamma u \\
\Gamma &\equiv \begin{pmatrix} I_n & \Delta \end{pmatrix}, \Delta \equiv \begin{pmatrix} \Delta_1 & \Delta_2 & \cdots & \Delta_m \end{pmatrix} \\
u &\equiv \begin{pmatrix} u_0' & u_1' & \cdots & u_m' \end{pmatrix}',
\end{aligned}$$

Δ_i denotes the $(n \times N_i)$ matrix of dummy variables indicating the groups at the level $i = 1, ..., m$ and u_i denotes the correspondent error component $(N_i \times 1)$ vector; u_0 stands for the idiosyncratic (observation specific) error component $(n \times 1)$ vector.

The following assumptions hold throughout.

A.1 For $i = 0, 1, .., m$, u_i is a vector of i.i.d. random variables such that

$E(u_{ij}) = 0$, $E\left(u_{ij}^2\right) = \sigma_i^2 < \infty$, $E\left(u_{ij}^4\right) = \kappa_{i,4} + 3\sigma_i^4 < \infty$, $j = 1, ..., N_i$, with $N_0 \equiv n$, $\kappa_{i,4}$ and σ_i^2 fixed parameters; u_0, u_1, ..., u_m are independent random vectors.

In what follows a sequence of models (1) is defined for n tending to infinity, therefore all variables should be thought of as indexed by n. For the sake of notational simplicity, though, I take the liberty of omitting such dependence.

A.2 a) For $i = 1, .., m$, Δ_i is a matrix having exactly 1 unity element and $N_i - 1$ zero elements in every row and such that none of its columns has all zero elements, $i = 1, ..., m$. b) $r(\Delta)/n$ is bounded away from 1, i.e. there exist a constant $0 < \rho < 1$ and an integer n_0 such that

$$\frac{r(\Delta)}{n} < \rho$$

for all $n > n_0$.

Assumption A.1 implies that the composite error ϵ has the multi-way ECM variance-covariance matrix (Wansbeek and Kaptein (1989) and Davis (2002)), that is $Var(\epsilon) = \Sigma$, where

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

Assumption A.2 is fairly general and met by most real world data-sets and applications. Part a) requires that the effects in any dimension be properly specified without redundant dummies, implying that Δ_i has full column rank (f.c.r.). It also implies that the columns of Δ_i are mutually exclusive, that is any observation in the sample belongs to only one group in a given dimension i, and since this is true for all $i = 1, ..., m$, any observation in the sample can be classified according to all dimensions.

Sets of mutually exclusive dummies are the norm in ECM's. They always occur when the data points in the idiosyncratic dimension are specified in terms of the categories of all the other dimensions. Nested classifications always meet this requirement (see Westfall (1986) and Baltagi et al. (2001), among others), and also properly constructed non-nested classifications. Think, for example, of a data-set matching films and theaters over time (weeks) with observations specified by the triple index film/theater/week, (i, j, t), as in Davis (2002). All sets of dummies are mutually exclusive, regardless of the fact that in the same week the same film may be shown in different theaters and the same theater may show more than one film: observation (i, j, t) peculiar to film i shown by theater j in week t belongs only to group i in the film dimension, group j in the theater dimension and group t in the time dimension.

Not always is it necessary to index data points by all of the non idiosyncratic dimensions to get sets of mutually exclusive dummies in non-nested classifications. These would still be assured with less indexes, provided that the categories in the omitted dimensions are uniquely identified by those in the included dimensions. In a matched-film-theater data-set where theaters show only one film at the time, knowing the theater and the week identifies the film, so that the data can be expanded over theaters and weeks alone. This occurrence is beneficial not only because it implies mutually exclusive dummies along all dimensions, but also because it avoids relatively large sample sizes. In matched-employer-employee panels with many firms and workers, for example, observations are typically indexed only by workers and years and some criterion is applied to assign each worker/year observation to a single firm, for example the firm at which the worker has worked the largest number of days during the year (as in Abowd et al. (1999)). In this way, mutually exclusive dummies are preserved for the three dimensions and the data set is expanded only over employees and years: observation (i,t), peculiar to employee i working in year t, belongs only to group i in the employee dimension, group t in the time dimension and the corresponding group $j(i,t)$ in the firm dimension, uniquely determined by i and t by construction.

Importantly, Part a) of A.2, does not imply that dummies indicating groups in different dimensions are mutually exclusive, nor that Δ has f.c.r. It does imply, though, that

$$r(\Delta_i) = N_i \leq r(\Delta) \leq \sum_{i=1}^{m} N_i, \tag{3}$$

$i = 1, \ldots, m$.

Part b) is a mild regularity condition, fulfilled by any conceivable ECM. Given that $n - r(\Delta) = tr\, Q_{[\Delta]} \geq 0$, it has

$$0 \leq \frac{n - r(\Delta)}{n} < 1$$

always. Therefore, rewritten as

$$0 < 1 - \rho < \frac{n - r(\Delta)}{n} < 1,$$

for all $n \geq n_0$, part b) of A.2 simply prevents $[n - r(\Delta)]/n$ to get too close to zero as the sample size gets larger, and so implies that n and $n - r(\Delta)$ are of the same order of magnitude. For example, it is satisfied by balanced two-way panel data models in which $n = N_1 N_2$, $r(\Delta) = N_1 + N_2 - 1$ and both N_1 and N_2 tend to infinity as n tends to infinity (as in Wallace and Hussain (1969) and Amemiya (1971)) or in unbalanced multidimensional ECM's where $N_i/n \to \rho_i < 1$, $i = 1, \ldots, m$, with restrictions on the ρ's that are specific to the model considered (as in Westfall (1986) for the nested case).[3]

---

[3] In Wansbeek and Kaptein (1989) there is a finite-sample counterpart to b) of A.2, requiring that 1) each individual in the sample is observed over at least two periods and 2) $N_1/N_2 \geq 1$. This is enough to ensure $n \geq 2N_1 \geq 2N_2$ and, in turn, $n - (N_1 + N_2 - 1) > 0$.

A.1 and A.2 are compatible with both nested and non-nested ECM's. I follow Davis (2002) and keep a general treatment throughout, so that I will cover the nested case and specific non-nested models as way of examples or in specific subsections. A.1 and A.2 are only a part of the set of assumptions used in the paper, they are enough, though, to ensure existence, unbiasedness and consistency for the unfeasible Anova estimator of σ_0^2. The other assumptions will be given in due course.

3 Anova-type consistent estimators - ACE's

ACE's are constructed following the approach referred to as Procedure II in the Introduction (Wallace and Hussain (1969), Amemiya (1971), Baltagi and Chang (2000)). For ease of exposition, the steps of Procedure II are reproduced next.

1. Start from theoretical quadratic forms in the unobserved ϵ and work out their unconditional expectations as linear functions of the unknown variance components.
2. Equate the theoretical quadratic forms to their unconditional expectations and obtain the unfeasible Anova estimator as the solution of the system.
3. Obtain the ACE by replacing the unobserved ϵ in the formula of the unfeasible estimator with the residuals from some regression of choice.

To prove that the resulting ACE is actually consistent boils down to proving, first, that the unfeasible estimator is consistent and, second, that the resulting feasible estimator is equal to the unfeasible estimator plus an $o_p(1)$ remainder. Section 3.1 covers steps 1 and 2 to derive the unfeasible Anova estimator and presents the theorems assuring its existence, unbiasedness and consistency under given regularity conditions. Each subsequent section, from 3.2 to 3.4, goes through step 3 to derive ACE1-ACE3 and presents the theorem establishing existence and consistency, under given regularity conditions, for each ACE.

3.1 The unfeasible Anova estimator

The unfeasible Anova estimator for σ_0^2 is obtained, according to step 1 of Procedure II, by establishing that

$$E\left(\epsilon' Q_{[\Delta]}\epsilon\right) = \sigma_0^2 tr\ Q_{[\Delta]} = \sigma_0^2\left[n - r\left(\Delta\right)\right]$$

(see Lemma 1 in appendix for the first equality and Lemma 3 for the second) and then, turning to step 2, solving $\epsilon' Q_{[\Delta]}\epsilon = \widetilde{\sigma}_0^2 tr\ Q_{[\Delta]}$ for $\widetilde{\sigma}_0^2$:

$$\begin{aligned} \widetilde{\sigma}_0^2 &= \frac{\epsilon' Q_{[\Delta]}\epsilon}{n - r\left(\Delta\right)}, \\ &= \frac{u_0' Q_{[\Delta]} u_0}{n - r\left(\Delta\right)}. \end{aligned} \tag{4}$$

The resulting $\widetilde{\sigma}_0^2$ is clearly unbiased under A.1.

The derivation of the unfeasible Anova estimator for the variance components

$$\sigma^2 = \left(\begin{array}{cccc} \sigma_1^2 & \sigma_2^2 & \ldots & \sigma_m^2 \end{array} \right)',$$

goes along the same lines as above. It is based upon the quadratic forms $\epsilon' P_{[\Delta_i]}\epsilon$, $i = 1, ..., m$, whose expectations are readily worked out under A.1. So,

$$E\left(\epsilon' P_{[\Delta_i]}\epsilon\right) = E\left(tr\ P_{[\Delta_i]}\epsilon\epsilon'\right) = tr\ P_{[\Delta_i]}\Sigma$$

and then

$$E\left(\epsilon' P_{[\Delta_i]}\epsilon\right) = \sigma_0^2 N_i + \sigma_1^2 tr\ \Delta_1' P_{[\Delta_i]}\Delta_1 + ... \tag{5}$$
$$+\sigma_i^2 n + ... + \sigma_m^2 tr\ \Delta_m' P_{[\Delta_i]}\Delta_m$$

$i = 1, ..., m$. I can now obtain the unbiased, unfeasible Anova estimator for σ^2, $\widetilde{\sigma}^2$, as the solution of the system $A\widetilde{\sigma}^2 = B - \widetilde{\sigma}_0^2 C$:

$$\widetilde{\sigma}^2 = A^{-1}\left(B - \widetilde{\sigma}_0^2 C\right), \tag{6}$$

where

$$A \equiv \frac{1}{n}\begin{pmatrix} n & \cdots & tr\ \Delta_i' P_{[\Delta_1]}\Delta_i & \cdots & tr\ \Delta_m' P_{[\Delta_1]}\Delta_m \\ \vdots & \ddots & \vdots & & \vdots \\ tr\ \Delta_1' P_{[\Delta_i]}\Delta_1 & \cdots & n & \cdots & tr\ \Delta_m' P_{[\Delta_i]}\Delta_m \\ \vdots & & \vdots & \ddots & \vdots \\ tr\ \Delta_1' P_{[\Delta_m]}\Delta_1 & \cdots & tr\ \Delta_i' P_{[\Delta_m]}\Delta_i & \cdots & n \end{pmatrix},$$

$$B \equiv \frac{1}{n}\begin{pmatrix} \epsilon' P_{[\Delta_1]}\epsilon & \epsilon' P_{[\Delta_2]}\epsilon & \ldots & \epsilon' P_{[\Delta_m]}\epsilon \end{pmatrix},'$$

$$C \equiv \frac{1}{n}\begin{pmatrix} trP_{[\Delta_1]} & trP_{[\Delta_2]} & \ldots & trP_{[\Delta_m]} \end{pmatrix}'$$

and $\widetilde{\sigma}_0^2$ is specified in equation (4). Notice that, given a) of A.2,

$$C = \frac{1}{n}\begin{pmatrix} N_1 & N_2 & \ldots & N_m \end{pmatrix}'.$$

Matrix A is a positive matrix that is invertible in most empirical cases. One case is the nested ECM (see Section 3.1.1 below). For completely non-nested models (in which no dimension is nested into another) there is the sufficient condition that A has a strictly dominant diagonal,[4] expressed as

$$\max\left\{\sum_{i=2}^{m} tr\ \Delta_1' P_{[\Delta_i]}\Delta_1, ..., \sum_{i\neq j}^{m} tr\ \Delta_j' P_{[\Delta_i]}\Delta_j, ..., \sum_{i=1}^{m-1} tr\ \Delta_m' P_{[\Delta_i]}\Delta_m\right\} < n.$$

It is easy to prove that the foregoing condition is satisfied, and so A is non singular, in the two-way model considered by Wansbeek and Kaptein (1989).[5]

The consistency analysis is based on the asymptotic behaviour of the quadratic forms in B, $\epsilon' P_{[\Delta_i]}\epsilon/n$, $i = 1, ..., m$. Although the composite error ϵ is not i.i.d., expanding $\epsilon' P_{[\Delta_i]}\epsilon$ as

$$\begin{aligned} \epsilon' P_{[\Delta_i]}\epsilon &= \left(u_0' + \sum_{h=1}^{m} u_h'\Delta_h'\right) P_{[\Delta_i]} \left(u_0 + \sum_{h=1}^{m} \Delta_h u_h\right) \\ &= u_0' P_{[\Delta_i]} u_0 + 2\sum_{h=1}^{m} u_0' P_{[\Delta_i]}\Delta_h u_h + \sum_{h=1}^{m}\sum_{j=1}^{m} u_h'\Delta_h' P_{[\Delta_i]}\Delta_j u_j \quad (7) \end{aligned}$$

shows that $\epsilon' P_{[\Delta_i]}\epsilon$ is the sum of a fixed number of quadratic forms and polynomials in random variables that satisfy A.1. Therefore the limiting behaviour of $\epsilon' P_{[\Delta_i]}\epsilon/n$ can be worked out by focusing on that of its components:

$$\frac{u_0' P_{[\Delta_i]} u_0}{n}$$

$i = 1, ..., m$;

$$\frac{u_0' P_{[\Delta_i]}\Delta_j u_j}{n}$$

---

[4] If the model is partially nested (some dimension is nested into others), then a modified condition applies (for more on this and other algebraic aspects of ECM's see Bruno (2009)).

[5] We have seen in footnote 3 that in the unbalanced two-way panel data model considered by Wansbeek and Kaptein (1989) $n \geq 2N_1 \geq 2N_2$ and since it turns out that in this case

$$tr\ \Delta_2' P_{[\Delta_1]}\Delta_2 = N_1,\ tr\ \Delta_1' P_{[\Delta_2]}\Delta_1 = N_2,$$

the condition of a dominant diagonal is met, and hence A turns out to be non-singular. Indeed, this can be easily verified by direct inspection of $\det(A)$:

$$\det(A) = 1 - \frac{N_1 N_2}{n^2} > 0.$$

$i, j = 1, ..., m$;

$$\frac{u_j' \Delta_j' P_{[\Delta_i]} \Delta_j u_j}{n},$$

$i, j = 1, ..., m$; and

$$\frac{u_h' \Delta_h' P_{[\Delta_i]} \Delta_j u_j}{n}$$

$h \neq j = 1, ..., m$ and $i = 1, ..., m$.

Theorem 1 *Assume A.1 and A.2. Then, for all $i, j = 1, \ldots, m$,*

$$\begin{aligned} &a)\ \widetilde{\sigma}_0^2 \underset{p}{\rightarrow} \sigma_0^2 \\ &b)\ \frac{u_0' P_{[\Delta_i]} u_0}{n} - \frac{N_i}{n} \sigma_0^2 \underset{p}{\rightarrow} 0 \\ &c)\ \frac{u_0' P_{[\Delta_i]} \Delta_j u_j}{n} \underset{p}{\rightarrow} 0 \end{aligned}$$

as $n \to \infty$.

Theorem 1 provides the asymptotic results for all terms in equation (7) that involve the idiosyncratic error component, u_0. The probability limits hold under any form of unbalancedness and irrespective of how n tends to infinity. I now establish the convergence results for the remaining terms of $\epsilon' P_{[\Delta_i]} \epsilon / n$.

Let $g_i(j)$ denote the size of group $j = 1, ..., N_i$ in dimension $i = 1, ..., m$. Assumption a) of A.2 assures that the $(N_i \times N_i)$ matrix $\Delta_i' \Delta_i$ is a diagonal matrix with $g_i(j)$ as its (j, j).th element. Let

$$\overline{g}_i \equiv \frac{1}{N_i} \sum_{j=1}^{N_i} g_i(j) \tag{8}$$

denote the average group size of dimension i and $g_i^{(2)} \equiv \frac{1}{N_i} \sum_{j=1}^{N_i} g_i^2(j)$ the second moment around zero of the group sizes in the same dimension, $i = 1, ..., m$.

A one-way unit-free measure of unbalancedness that is relevant in this paper is the squared Pearson's coefficient of variation between group sizes, $\overline{v}_i^2$, defined as

$$\overline{v}_i^2 \equiv \frac{g_i^{(2)}}{\overline{g}_i^2} - 1. \tag{9}$$

$i = 1, ..., m$. By the Jensen's inequality, $\overline{v}_i^2 \geq 0$, with equality if dimension i is balanced (or if all but one clusters are of zero size, an empty case, though, which is ruled out by A.2). If the dimension is unbalanced, since

$$\textstyle\sum_{j=1}^{N_i} g_i^2\left(j\right) < \left(\sum_{j=1}^{N_i} g_i\left(j\right)\right)^2,$$

it also has

$$\overline{v}_i^2 + 1 < N_i, \tag{10}$$

$i = 1, ..., m$.[6] Another measure, closely related to $\overline{v}_i^2$, is

$$\psi_i \equiv \frac{\overline{v}_i^2 + 1}{N_i}, \tag{11}$$

with $0 < \psi_i < 1$, $i = 1, \ldots, m$. A convenient interpretation of ψ_i involves its reciprocal, ψ_i^{-1}, which can be thought of as an unbalancedness-corrected size for dimension i.

That ψ_i is key to the asymptotic analysis of the variance component estimators emerges from the following important relationships (obtained as inequalities (40) and (41) in appendix)

$$Var\left(\frac{u_i'\Delta_i' P_{[\Delta_j]}\Delta_i u_i}{n}\right) \leq Var\left(\frac{u_i'\Delta_i'\Delta_i u_i}{n}\right) = \left(\kappa_{i,4} + 2\sigma_i^4\right)\psi_i$$

---

[6] Again, equality in (10) would hold only in the moot case of all but one clusters of zero size, excluded by A.2. For a general analysis of the Pearson's coefficient of variation see Katnelson and Kotz (1957) and Kendall and Stuart (1969), p. 54.

for all $i, j = 1, \ldots, m$, and

$$Var\left(\frac{u_i'\Delta_i'P_{[\Delta_k]}\Delta_j u_j}{n}\right) < \sigma_i^2\sigma_j^2 \min\left(\psi_i, \psi_j\right),$$

for all $i \neq j = 1, \ldots, m$ and for all $k = 1, \ldots, m$.

Hence, given the other regularity conditions, ψ_i being $o(1)$ for all $i = 1, ..., m$ is necessary and sufficient for all terms

$$\frac{u_i'\Delta_i'P_{[\Delta_k]}\Delta_j u_j}{n} - E\left(\frac{u_i'\Delta_i'P_{[\Delta_k]}\Delta_j u_j}{n}\right),$$

$i, j, k = 1, ..., m$, to converge jointly to 0 in square mean, and consequently sufficient for $\epsilon' P_{[\Delta_i]}\epsilon/n - E\left(\epsilon' P_{[\Delta_i]}\epsilon/n\right) \underset{p}{\longrightarrow} 0$. This condition is explicitly stated as follows.

A.3 $\psi_i \to 0$ as $n \to \infty$, $i = 1, ..., m$.

Since the numerator in the formula of ψ_i, in equation (11), is no smaller than one, A.3 implies that $N_i \to \infty$ as $n \to \infty$, $i = 1, ..., m$, which is the norm in all ECM's dealing with asymptotic results (Wallace and Hussain (1969), Amemiya (1971), Westfall (1986), among others). In addition, it implies a mild regularity condition for the sequence of unbalanced designs. Indeed, given that $\overline{v}_i^2$ is already $O(N_i)$ by (10), A.3 boils down to ruling out all patterns where the degree of unbalancedness in a given dimension i, measured by $\overline{v}_i^2$, grows at the same speed as N_i. Therefore, given $N_i \to \infty$, A.3 is implied by the following unbalancedness restrictions, listed in an increasing level of generality. 1) All group sizes are uniformly bounded, as maintained in the nested ECM by Westfall (1986) (an assumption, though, that is not applicable in non-nested models); 2) all group sizes are of the same order of magnitude; 3) $\overline{v}_i^2$ is uniformly bounded (that 1 implies 2 and that 3, along with $N_i \to \infty$, implies A.3 is obvious; that 2 implies 3 is proved in appendix). Moreover, it turns out that $\overline{v}_i^2$ may be uniformly bounded even if group sizes are heterogeneous in the order of magnitude. An example in Appendix proves this last claim. Importantly, it is evident that A.3 is consistent with $\overline{v}_i^2$ unbounded.

Theorem 2 *Assume A.1-A.3. Then,*

$$a) \quad \frac{u_j' \Delta_j' P_{[\Delta_i]} \Delta_j u_j}{n} - \frac{tr \Delta_j' P_{[\Delta_i]} \Delta_j}{n} \sigma_j^2 \underset{p}{\rightarrow} 0$$

as $n \rightarrow \infty$ *for all* $i, j = 1, \ldots, m$, *and*

$$b) \quad \frac{u_i' \Delta_i' P_{[\Delta_k]} \Delta_j u_j}{n} \quad \underset{p}{\rightarrow} 0$$

as $n \rightarrow \infty$ *for all* $i \neq j = 1, \ldots, m$ *and* $k = 1, \ldots, m$.

Theorem 2 completes Theorem 1, providing the probability limits for the terms in $\epsilon' P_{[\Delta_i]} \epsilon / n$ not involving the idiosyncratic component.

Notice that the quadratic forms and polynomials considered in Theorems 1 and 2 are basic ingredients of all existing Anova estimators. It is therefore evident that the interest of the theorems goes beyond the ACE's considered in this paper.

The following definition, taken from White (1982), is given here to state concisely the additional assumptions needed for the consistency analysis of $\widetilde{\sigma}^2$ and the other estimators of this paper.

Definition 1 *A sequence of* $(k \times k)$ *matrices* S *is said uniformly non-singular if there exist a constant* $\delta > 0$ *and an integer* n_0 *such that* $|\det(S)| > \delta$ *for all* $n > n_0$.

A.4 The sequence of matrices A in (6) is uniformly non-singular.

Assumption A.4 prevents that A gets too close to singularity as the sample size expands.[7]

With assumption A.4 and Theorems 1 and 2 in hand, I can establish existence, unbiasedness and consistency for $\widetilde{\sigma}^2$.

Theorem 3 *Let A1-A4 hold. Then, $\widetilde{\sigma}^2$ exists and is unbiased for n sufficiently large, and*

$$\widetilde{\sigma}^2 \underset{p}{\rightarrow} \sigma^2$$

as $n \rightarrow \infty$.

Finally, notice that if $\sigma^2 = 0$, then all terms in B involving u_i, $i = 1, \ldots, m$, degenerate to 0 with probability 1 and $\widetilde{\sigma}^2$ is therefore consistent without the large sample restrictions on unbalancedness placed by A.3.

Corollary 1 *Let $\sigma^2 = 0$ and A.1, A.2 and A.4 hold. Then*

$$\widetilde{\sigma}^2 \underset{p}{\rightarrow} \sigma^2$$

as $n \rightarrow \infty$.

3.1.1 The nested ECM

If the ECM is nested, that is if $\mathcal{R}(\Delta_m) \subset \mathcal{R}(\Delta_{m-1}) \subset \ldots \subset \mathcal{R}(\Delta_1)$, it turns out that $\widetilde{\sigma}^2$ in (6) can be equivalently expressed as

$$\widetilde{\sigma}^2 = A^{*^{-1}} \left(B^* - \widetilde{\sigma}_0^2 C^* \right), \tag{12}$$

---

[7] In the unbalanced two-way panel data model considered by Wansbeek and Kaptein (1989), it has $r(\Delta) = N_1 + N_2 - 1$ and

$$\det(A) = 1 - \frac{N_1 N_2}{n^2}$$

(see footnote 3), and hence A.4 is always met when part b) of A.2 holds.

where $\widetilde{\sigma}_0^2$ is the same as in equation (4) and A^* is a lower triangular matrix with the elements onto, or below, the main diagonal equal to

$$a_{ij}^* \equiv tr \ \frac{\Delta_j' \left(P_{[\Delta_i]} - P_{[\Delta_{i+1}]}\right) \Delta_j}{n}$$

for all $j \leq i = 1, \ldots, m-1$ and $a_{mj}^* \equiv tr \ \Delta_j' P_{[\Delta_m]} \Delta_j / n$ for all $j = 1, \ldots m$ and the elements above the main diagonal all equal to 0,

$$B^* \equiv \frac{1}{n} \left[\begin{array}{cccc} \epsilon' \left(P_{[\Delta_1]} - P_{[\Delta_2]}\right) \epsilon & \epsilon' \left(P_{[\Delta_2]} - P_{[\Delta_3]}\right) \epsilon & \ldots & \epsilon' P_{[\Delta_m]} \epsilon \end{array} \right]'$$

and

$$C^* \equiv \frac{1}{n} \left[\begin{array}{cccc} tr \ \left(P_{[\Delta_1]} - P_{[\Delta_2]}\right) & tr \ \left(P_{[\Delta_2]} - P_{[\Delta_3]}\right) & \ldots & tr \ P_{[\Delta_m]} \end{array} \right]'.$$

A^* being a lower triangular matrix implies a manageable formula for its determinant, $\det A^* = \prod_{i=1}^m a_{ii}^*$, which permits to see at once that A^* is non-singular, given that

$$a_{ii}^* = \frac{n - tr \ \Delta_i' P_{[\Delta_{i+1}]} \Delta_i}{n} > 0,$$

$i = 1, ..., m$. The latter is readily proved upon noticing that, in general, $tr \ \Delta_i' P_{[\Delta_j]} \Delta_i \leq n$ (see inequality (44) in appendix) and, if the model is nested, $tr \ \Delta_i' P_{[\Delta_{i+1}]} \Delta_i < n$.

That $A^{^{-1}} \left(B - \widetilde{\sigma}_0^2 C\right) = A^{*^{-1}} \left(B^* - \widetilde{\sigma}_0^2 C^*\right)$ is also easy to prove. If $\mathcal{R}\left(\Delta_m\right) \subset \mathcal{R}\left(\Delta_{m-1}\right) \subset \ldots \subset \mathcal{R}\left(\Delta_1\right)$, then all terms onto and above the main diagonal of

A reduces to unity:

$$A=\begin{pmatrix} 1 & \cdots & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & & \vdots \\ tr\ \Delta_1' P_{[\Delta_i]} \Delta_1/n & \cdots & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ tr\ \Delta_1' P_{[\Delta_m]} \Delta_1/n & \cdots & tr\ \Delta_i' P_{[\Delta_m]} \Delta_i/n & \cdots & 1 \end{pmatrix}.$$

Transform the system originating $\widetilde{\sigma}^2$, $A\widetilde{\sigma}^2 = B - \widetilde{\sigma}_0^2 C$, by subtracting the second equation from the first, the third from the second and so on until the m.th equation, which is left unchanged. The outcome of the transformation is exactly $A^*\widetilde{\sigma}^2 = B^* - \widetilde{\sigma}_0^2 C^*$. Then, notice that $A^* = PA$, $B^* = PB$ and $C^* = PC$, where P is the $(m \times m)$ identity matrix with all elements onto the first upper sub-diagonal equal to -1. Hence, P is an elementary row operator matrix and, as such, is invertible with $\det(P) = 1$. Thus, the two systems have exactly the same solution $\widetilde{\sigma}^2$, proving the following corollary to Theorem 3.

Corollary 2 *Let A.1-A.4 hold and* $\mathcal{R}(\Delta_m) \subset \mathcal{R}(\Delta_{m-1}) \subset \ldots \subset \mathcal{R}(\Delta_1)$. *Then,* $\widetilde{\sigma}^2$ *defined in equation (6) can be equivalently expressed as in equation (12), exists, is unbiased and*

$$\widetilde{\sigma}^2 \underset{p}{\to} \sigma^2$$

as $n \to \infty$.

3.2 A consistent estimator using WTSLS residuals - ACE1

In this and the next two subsections I go through the last step of Procedure II to obtain the three new ACE's for use in Instrumental Variables ECM regressions.

The following assumption is common to all ACE's.

A.5 Let the sequences of $(k \times k)$ matrices, Q_{XX} and $Q_{XX,i}$, and $(k \times 1)$ vectors, $Q_{X\epsilon}$ and $Q_{X\epsilon,i}$, be $O(1)$, $i = 1, ..., m$, then

$$\text{a)} \qquad \frac{X'Q_{[\Delta]}X}{n} - Q_{XX} \underset{p}{\rightarrow} 0$$
$$\frac{X'Q_{[\Delta]}\epsilon}{n} - Q_{X\epsilon} \underset{p}{\rightarrow} 0$$

as $n \to \infty$, and

$$\text{b)} \qquad \frac{X'P_{[\Delta_i]}X}{n} - Q_{XX,i} \underset{p}{\rightarrow} 0$$
$$\frac{X'P_{[\Delta_i]}\epsilon}{n} - Q_{X\epsilon,i} \underset{p}{\rightarrow} 0,$$

as $n \to \infty$, $i = 1, ..., m$.

A.5 is a general assumption. It does not require convergence in probability to fixed limiting values and permits that X and ϵ may be related asymptotically.

Before introducing the matrix of available instruments, I give the following definitions, taken from White (1982), to state concisely the additional assumptions needed to characterize the asymptotic behaviour of the instrument matrix in this and the next two subsections.

Definition 2 *A sequence of* $(k \times k)$ *semi-positive definite matrices* S *is said uniformly positive definite if it is uniformly non-singular.*

Definition 3 *A sequence of* $(n \times k)$ *matrices* S *is said to have uniformly full-column-rank if there exists a sequence of* $(k \times k)$ *submatrices* S^* *which is uniformly non-singular.*

Let Z denote the $(n \times p)$ matrix of available instruments with $p \geq k$. The following assumption is specific to ACE1 and is key to both existence and consistency of the estimator.

A.6.WTSLS Let the sequence of $(p \times k)$ matrices Q_{ZX} be $O(1)$ and uniformly of f.c.r. and the sequence of $(p \times p)$ matrices Q_{ZZ} be $O(1)$ and uniformly positive definite, then

$$\begin{aligned} \frac{Z'Q_{[\Delta]}X}{n} - Q_{ZX} &\underset{p}{\rightarrow} 0 \\ \frac{Z'Q_{[\Delta]}Z}{n} - Q_{ZZ} &\underset{p}{\rightarrow} 0 \\ \frac{Z'Q_{[\Delta]}\epsilon}{n} &\underset{p}{\rightarrow} 0 \end{aligned}$$

as $n \rightarrow \infty$.

With this in hand, I go through step 3 of Procedure II to derive ACE1. The WTSLS residual vector

$$\begin{aligned} \widehat{\epsilon}_w &= \Big\{ I - X \left[X'Q_{[\Delta]}Z \left(Z'Q_{[\Delta]}Z \right)^{-1} Z'Q_{[\Delta]}X \right]^{-1} \\ & X'Q_{[\Delta]}Z \left(Z'Q_{[\Delta]}Z \right)^{-1} Z'Q_{[\Delta]} \Big\} \epsilon \\ &= M_w \epsilon \end{aligned} \tag{13}$$

replaces ϵ into the unfeasible estimator formulas, (4) and (6) , to give

$$\widehat{\sigma}^2_{w,0} = \frac{\widehat{\epsilon}'_w Q_{[\Delta]} \widehat{\epsilon}_w}{n - r(\Delta)} \tag{14}$$

and

$$\widehat{\sigma}^2_w = A^{-1} \left(\widehat{B}_w - \widehat{\sigma}^2_{w,0} C \right), \tag{15}$$

where

$$\widehat{B}_w \equiv \frac{1}{n} \left(\widehat{\epsilon}'_w P_{[\Delta_1]} \widehat{\epsilon}_w \ \widehat{\epsilon}'_w P_{[\Delta_2]} \widehat{\epsilon}_w \ ... \ \widehat{\epsilon}'_w P_{[\Delta_m]} \widehat{\epsilon}_w \right)'.$$

ACE1 presents itself as a direct extension of the Anova-type estimator by Amemiya (1971) in its use of within residuals into the formula of the unfeasible, unbiased Anova estimator.

Theorem 4 *Assume A.1, A.2 and A.4-A.6.WTSLS. Then,* $\widehat{\sigma}_{w,0}^2$ *and* $\widehat{\sigma}_w^2$ *exist with probability approaching 1; and*

$$\widehat{\sigma}_{w,0}^2 \underset{p}{\rightarrow} \sigma_0^2$$

as $n \to \infty$*; assuming also A.3,*

$$\widehat{\sigma}_w^2 \underset{p}{\rightarrow} \sigma^2$$

as $n \to \infty$.

Establishing the existence of ACE1, in both its part, boils down to establishing that the matrix in brackets in equation (13) may be inverted. The answer of Theorem 4 is in the affirmative, in that a non-null determinant for that matrix becomes more and more likely as the sample size gets larger. Technically speaking, this does not exclude that X and Z may not be of f.c.r. or that some regressors and instruments may lie onto $\mathcal{R}(\Delta)$ in finite samples, only these events become increasingly unlikely for n sufficiently large. Therefore, specific variables like the unity vector, which by construction lie onto $\mathcal{R}(\Delta)$ with probability one, and as such would prevent non-singularity with probability one, cannot find a place into X and Z. This raises the issue of accommodating the intercept in model (1). Solutions to this problem are specific to the particular estimation method considered, therefore I postpone them until Section 3.6.

Along the same lines as Section 3.1.1, it turns out that if the model is nested, then a numerically equivalent expression for $\widehat{\sigma}_w^2$ is given by $\widehat{\sigma}_w^2 = {A^*}^{-1}\left(\widehat{B}_w^* - \widehat{\sigma}_{w,0}^2 C^*\right)$, with $\widehat{\sigma}_{w,0}^2$ unchanged, A^* and C^* defined as in Section

3.1.1 and

$$\widehat{B}_w^* \equiv \frac{1}{n}\left[\begin{array}{cccc} \widehat{\epsilon}_w'\left(P_{[\Delta_1]}-P_{[\Delta_2]}\right)\widehat{\epsilon}_w & \widehat{\epsilon}_w'\left(P_{[\Delta_2]}-P_{[\Delta_3]}\right)\widehat{\epsilon}_w & \ldots & \widehat{\epsilon}_w' P_{[\Delta_m]}\widehat{\epsilon}_w \end{array}\right]'.$$

The derivations of Section 3.1.1 carry over into ACE1 due to the fact that in either case a common residual vector is used for all variance components.

3.3 A consistent estimator using TSLS residuals - ACE2

ACE2 is based upon residuals from the pooled TSLS regression

$$\begin{aligned}\widehat{\epsilon}_{ls} &= \left\{I - X\left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1}X'Z\left(Z'Z\right)^{-1}Z'\right\}\epsilon \\ &= M_{ls}\epsilon\end{aligned}$$

and is obtained by replacing ϵ with $\widehat{\epsilon}_{ls}$ into the Anova unfeasible estimator formulas, (4) and (6), to have

$$\widehat{\sigma}_{ls,0}^2 = \frac{\widehat{\epsilon}_{ls}' Q_{[\Delta]}\widehat{\epsilon}_{ls}}{n - r\left(\Delta\right)}, \tag{16}$$

and

$$\widehat{\sigma}_{ls}^2 = A^{-1}\left(\widehat{B}_{ls} - \widehat{\sigma}_{ls,0}^2 C\right), \tag{17}$$

where

$$\widehat{B}_{ls} \equiv \frac{1}{n}\left(\widehat{\epsilon}_{ls}' P_{[\Delta_1]}\widehat{\epsilon}_{ls}\ \widehat{\epsilon}_{ls}' P_{[\Delta_2]}\widehat{\epsilon}_{ls}\ \ldots\ \widehat{\epsilon}_{ls}' P_{[\Delta_m]}\widehat{\epsilon}_{ls}\right)'.$$

ACE2 can be thought of as a direct extension of the Anova-type estimators by Wallace and Hussain (1969) and Baltagi and Chang (2000) in its use of TSLS residuals into the formula of the unfeasible, unbiased Anova estimator.

With the following assumption added to A1-A5, ACE2 is proved to exist and to be consistent.

A.6.TSLS Let the sequence of $(p \times k)$ matrices Q_{ZX} be $O(1)$ and uniformly of f.c.r. and the sequence of $(p \times p)$ matrices Q_{ZZ} be $O(1)$ and uniformly positive definite. Then,

$$\begin{aligned} &\frac{Z'X}{n} - Q_{ZX} \underset{p}{\to} 0 \\ &\frac{Z'Z}{n} - Q_{ZZ} \underset{p}{\to} 0 \\ &\frac{Z'\epsilon}{n} \underset{p}{\to} 0 \end{aligned}$$

as $n \to \infty$.

Theorem 5 *Assume A.1, A.2 and A.4-A.6.TSLS. Then,* $\widehat{\sigma}^2_{ls,0}$ *and* $\widehat{\sigma}^2_{ls}$ *exist with probability approaching* 1; *and*

$$\widehat{\sigma}^2_{ls,0} \underset{p}{\to} \sigma^2_0$$

as $n \to \infty$; *assuming also A.3,*

$$\widehat{\sigma}^2_{ls} \underset{p}{\to} \sigma^2$$

as $n \to \infty$.

As opposed to ACE1, ACE2 allows that X and Z contain variables that lie onto $\mathcal{R}(\Delta)$ for all sample sizes, since non-singularity of $X'Z(Z'Z)^{-1}Z'X$ would not necessarily break down in this case.

If the model is nested, what established for ACE1 is valid also for ACE2, with the only difference that $\widehat{\sigma}^2_{ls,0}$ and TSLS residuals are used.

3.4 A consistent estimator using WTSLS and BTSLS residuals - ACE3

ACE3 uses $\widehat{\sigma}_{w,0}^2$ of ACE1 to estimate σ_0^2 and the BTSLS residuals

$$\begin{aligned}\widehat{\epsilon}_{b,i} &= \Big\{ I - X \left[X' P_{[\Delta_i]} Z \left(Z' P_{[\Delta_i]} Z \right)^{-1} Z' P_{[\Delta_i]} X \right]^{-1} \\ &\quad X' P_{[\Delta_i]} Z \left(Z' P_{[\Delta_i]} Z \right)^{-1} Z' P_{[\Delta_i]} \Big\} \epsilon \\ &= M_{b,i} \epsilon\end{aligned}$$

$i = 1, ..., m$, to estimate B in (6). So, ACE3 for σ^2 is obtained as

$$\widehat{\sigma}_b^2 = A^{-1} \left(\widehat{B}_b - \widehat{\sigma}_{w,0}^2 C \right), \tag{18}$$

where

$$\widehat{B}_b \equiv \frac{1}{n} \left(\widehat{\epsilon}_{b,1}' P_{[\Delta_1]} \widehat{\epsilon}_{b,1} \ \widehat{\epsilon}_{b,2}' P_{[\Delta_2]} \widehat{\epsilon}_{b,2} \ ... \ \widehat{\epsilon}_{b,m}' P_{[\Delta_m]} \widehat{\epsilon}_{b,m} \right)'.$$

In its use of WTSLS and BTSLS residuals, ACE3 is an adaptation of the unbiased estimator derived by Swamy and Arora (1972), through Procedure I, in the context of a two-way ECM with nonstochastic regressors.

Consistency of $\widehat{\sigma}_b^2$ is proved under A.6.WTSLS and the following

A.6.BTSLS Let the sequence of $(p \times k)$ matrices $Q_{ZX,i}$ be $O(1)$ and uniformly of f.c.r. and the sequence of $(p \times p)$ matrices $Q_{ZZ,i}$ be $O(1)$ and uniformly positive definite, $i = 1, ..., m$. Then

$$\begin{aligned}\frac{Z' P_{[\Delta_i]} X}{n} - Q_{ZX,i} &\underset{p}{\rightarrow} 0 \\ \frac{Z' P_{[\Delta_i]} Z}{n} - Q_{ZZ,i} &\underset{p}{\rightarrow} 0 \\ \frac{Z' P_{[\Delta_i]} \epsilon}{n} &\underset{p}{\rightarrow} 0\end{aligned}$$

as $n \to \infty$.

Theorem 6 *Assume A.1, A.2, A.4-A.6.WTSLS. and A.6.BTSLS. Then,* $\widehat{\sigma}_b^2$ *exists with probability approaching* 1*; assuming also A.3*

$$\widehat{\sigma}_b^2 \underset{p}{\to} \sigma^2$$

as $n \to \infty$.

In the nested case, as opposed to the unfeasible Anova estimator, ACE1 and ACE2, $\widehat{\sigma}_b^2$ as obtained in equation (18) is not numerically equivalent to that that would be obtained by using the analog of formula (12). This is brought about by the use of a specific between residual vector for each of the m system equations originating $\widehat{\sigma}_b^2$.

3.4.1 A simpler ACE3

It is not hard to prove that under assumptions A.1-A.6.WTSLS and A.6.BTSLS a computationally simpler version of ACE3 can be obtained by using the same residual vector from a given between regression, say $\widehat{\epsilon}_{b,j}$, $j = 1, ..., m$, in the construction of $\widehat{B}_b$:

$$\widehat{B}_b \equiv \frac{1}{n} \left(\widehat{\epsilon}_{b,j}' P_{[\Delta_1]} \widehat{\epsilon}_{b,j} \; \widehat{\epsilon}_{b,j}' P_{[\Delta_2]} \widehat{\epsilon}_{b,j} \; ... \; \widehat{\epsilon}_{b,j}' P_{[\Delta_m]} \widehat{\epsilon}_{b,j} \right)'.$$

Indeed, for consistency it is enough that the set of assumptions A.6.BTSLS hold only for the chosen level j, along with A.1-A.6.WTSLS. Notice that using the same residual vector for estimating all terms in B is a common feature of ACE1 and ACE2 (and the unfeasible estimator, of course).

As opposed to ACE3 defined in (18), the simplified ACE3 lends itself to be easily implemented in multi-way nested models, since its using a common residual vector makes it immaterial to adjust the formulas to the nested case, exactly as it happens in the case of the other estimators. In addition, a natural candidate emerges for $\widehat{\epsilon}_{b,j}$ in the nested model: it is the residual vector from the between regression at the innermost level, which ensures the largest degrees of freedom.

3.5 A special case: $\mathcal{R}(X) \subseteq \mathcal{R}(Z)$

In the case regressors are part of the available instrument set or, more generally, can be expressed as linear combinations of the instruments, that is $\mathcal{R}(X) \subseteq \mathcal{R}(Z)$, the formulas of ACE1, ACE2 and ACE3 simplify as follows.

In ACE1 $\widehat{\epsilon}_w$ collapses to the within residuals

$$\widehat{\epsilon}_w = \left\{ I - X \left[X' Q_{[\Delta]} X \right]^{-1} X' Q_{[\Delta]} \right\} \epsilon \tag{19}$$

in equations (14) and (15).

In ACE2 the TSLS estimator for β collapses to the OLS estimator and so $\widehat{\epsilon}_{ls}$ collapses to the OLS residuals

$$\begin{aligned} \widehat{\epsilon}_{ls,x} &= \left\{ I - X \left(X'X \right)^{-1} X' \right\} \epsilon \\ &= Q_{[X]} \epsilon, \end{aligned}$$

in equations (16) and (17).

Finally, the formula of $\widehat{\sigma}_b^2$ in ACE3 simplifies similarly, with

$$\widehat{\epsilon}_{b,i} = \left\{ I - X \left[X' P_{[\Delta_i]} X \right]^{-1} X' P_{[\Delta_i]} \right\} \epsilon \tag{20}$$

replacing $\widehat{\epsilon}_{b,i}$ in equations (18), $i = 1, ..., m$, and $\widehat{\sigma}_{w,0}^2$ taken exactly as in ACE1.

All ACE's remain consistent under the usual assumptions and the further condition that $X = ZK$, where K is a $O(1)$ sequence of $(p \times k)$ matrices with uniformly f.c.r.

3.6 Allowing the intercept

In Section 3.2 we have seen that including the intercept into X and Z is not consistent with assumption A6.WTSLS. Therefore, in this case the intercept may find a place into equation (1) only as an explicit component of the model, that is $y = \underline{1}\beta_0 + X\beta + \epsilon$, where $\underline{1}$ is the vector of all unity elements. This, however, breaks down consistency of $\widehat{\sigma}_w^2$ in ACE1 since $\widehat{\epsilon}_w = M_w\epsilon + \underline{1}\beta_0$ (this problem does not affect the formula of $\widehat{\sigma}_{w,0}^2$ in (14) as the $\underline{1}\beta_0$ term is wiped out by $Q_{[\Delta]}$). As a solution, I just modify the unfeasible Anova estimator, $\widetilde{\sigma}^2$, so that $P_{[\Delta_i]} - P_{[\underline{1}]}$ replaces $P_{[\Delta_i]}$ in A, B and C and $\left(n - tr\ \Delta_i' P_{[\underline{1}]}\Delta_i\right)/n$ replaces the (i,i) .th terms of A. Then, upon obtaining ACE1 by plugging $\widehat{\epsilon}_w$ into the modified $\widetilde{\sigma}^2$, the term $\underline{1}\beta_0$ gets wiped out everywhere. The resulting ACE1 is consistent since Theorem 3 applies also to the modified $\widetilde{\sigma}^2$. The latter is proved easily on noticing that $\underline{1}$ is a dummy matrix respecting assumption a) of A.2 (which is all is required to the matrices entering the projection operators), so that Theorems 1 and 2 yield at once all the probability limits ensuring that the modified $\widetilde{\sigma}^2$ is indeed consistent:

$$\frac{u_0' P_{[\underline{1}]} u_0}{n} - \frac{\sigma_0^2}{n} \underset{p}{\rightarrow} 0,\ \frac{u_0' P_{[\underline{1}]} \Delta_i u_i}{n} \underset{p}{\rightarrow} 0,\ \frac{u_i' \Delta_i' P_{[\underline{1}]} \Delta_i u_i}{n} - \frac{tr \Delta_i' P_{[\underline{1}]} \Delta_i}{n} \sigma_i^2 \underset{p}{\rightarrow} 0$$

$i = 1, ..., m$ and

$$\frac{u_i' \Delta_i' P_{[\underline{1}]} \Delta_j u_j}{n} \underset{p}{\rightarrow} 0,$$

$i \neq j = 1, ..., m$.

While ACE2 can accommodate the unity vector into X and Z, so that the analysis of Sections 3.3 can go through with no modifications required for $\widetilde{\sigma}^2$, plugging TSLS residuals into $\widetilde{\sigma}^2$ modified as in the previous paragraph still produces consistent estimators, given that the modified $\widetilde{\sigma}^2$ is consistent. Indeed, Wallace and Hussain (1969) use an unfeasible estimator modified as in the previous paragraph in the first place.

As for ACE2, X and Z containing variables that lie onto $\mathcal{R}(\Delta)$ for all sample sizes does not necessarily bring about singularities into the BTSLS residuals. Therefore, when it comes to estimating σ^2, ACE3 can be implemented using regressor and instrument matrices that include such variables, even if these are excluded from the implementation of $\widehat{\sigma}^2_{w,0}$. This is perfectly legitimate, even if $\widehat{\sigma}^2_{w,0}$ is part of ACE3, since the formula of $\widehat{\sigma}^2_{w,0}$ in equation (14) is invariant to the inclusion of variables that lie onto $\mathcal{R}(\Delta)$, and so no incidental parameter problem would be at work here. Finally, the remark of the previous paragraph, on using the modified $\widetilde{\sigma}^2$ to construct ACE2, carries over into ACE3.

4 Anova-type unbiased estimators

The Anova-type estimators of Section 3 are consistent under fairly general conditions and are also computationally simple. Their finite sample properties, however, are unknown. Below, I consider specific regularity conditions under which it is possible to derive unbiased versions of $\widehat{\sigma}^2_{w,0}$, $\widehat{\sigma}^2_w$ and $\widehat{\sigma}^2_b$.

4.1 Anova-type unbiased estimators using WTSLS residuals

4.1.1 Exogenous regressors

The unbalanced two-way ECM is the framework within which Wansbeek and Kaptein (1989) derive what they call the quadratic unbiased estimator (QUE) of variance components. They use Procedure I and accomplish step 2 under the assumption that the composite error is independent of the regressors. A weaker assumption that assures the validity of the computations in step 2 is $E\left(\epsilon\epsilon'|X\right)=\Sigma$. It can be justified as an implication of conditional homoskedasticity, $Var\left(\epsilon|X\right)=\Sigma$, and exogenous regressors, $E\left(\epsilon|X\right)=0$.

The unbiased versions of $\widehat{\sigma}^2_{w,0}$ and $\widehat{\sigma}^2_w$ derived below, maintaining $\mathcal{R}\left(X\right)\subseteq\mathcal{R}\left(Z\right)$ as in Section 3.5 and $E\left(\epsilon\epsilon'|X\right)=\Sigma$, are exactly the multidimensional extension of the Wansbeek and Kaptein's QUE and for this reason the resulting estimator is referred to as WK throughout. It is obtained as the final outcome of the three steps of Procedure I, described in the Introduction and implemented here as follows.

1. Obtain the following $m+1$ empirical quadratic forms based on within residuals

$$\begin{aligned} q_{w,0} &= \widehat{\epsilon}'_w Q_{[\Delta]}\widehat{\epsilon}_w \\ q_{w,i} &= \widehat{\epsilon}'_w P_{[\Delta_i]}\widehat{\epsilon}_w \end{aligned} \tag{21}$$

   $i=1,...,m$, where $\widehat{\epsilon}_w$ is the same as in (19).

2. Work out the conditional expectation of $q_{w,0}$ and $q_{w,i}$ under $E\left(\epsilon\epsilon'|X\right)=\Sigma$ to have (details are in appendix)

$$\begin{aligned} E\left(q_{w,0}|X\right) &= \sigma_0^2\left[n-r\left(\Delta\right)-k\right] \\ E\left(q_{w,i}|X\right) &= \sigma_0^2\left[N_i+\kappa_i\right]+\sum_{s\neq i}^{m}\sigma_s^2 tr\ \Delta'_s P_{[\Delta_i]}\Delta_s+\sigma_i^2 n \end{aligned} \tag{22}$$

   where

$$\kappa_i = tr\left(X'Q_{[\Delta]}X\right)^{-1}X'P_{[\Delta_i]}X,$$

   $i=1,...,m$. [8]

3. The unbiased WK estimator for $\left(\sigma_0^2,\ \sigma^2\right)$, say $\left(\tilde{\sigma}^2_{w,0},\ \tilde{\sigma}^2_w\right)$, is eventually

---

[8] Boumahdi et al. (2006) report equations (22) for the three-way model. There is an obvious typo in their first equation (corresponding to my equation for $E\left(q_{w,0}|X\right)$), since the number of regressors, k, has therein the positive sign, while it should have the negative sign.

obtained as the solution of the following system of $m+1$ linear equations:

$$\begin{aligned} q_{w,0} &= \tilde{\sigma}^2_{w,0}\left[n - r(\Delta) - k\right] \\ q_{w,i} &= \tilde{\sigma}^2_{w,0}\left[N_i + \kappa_i\right] + \sum_{s \neq i}^{m} \tilde{\sigma}^2_{w,s} tr\ \Delta'_s P_{[\Delta_i]} \Delta_s + \tilde{\sigma}^2_{w,i} n \end{aligned} \tag{23}$$

$i = 1, ..., m$. Therefore,

$$\tilde{\sigma}^2_{w,0} = \frac{\widehat{\epsilon}_{w,x} Q_{[\Delta]} \widehat{\epsilon}_{w,x}}{n - r(\Delta) - k}, \tag{24}$$

and

$$\tilde{\sigma}^2_w = A^{-1}\left[\widehat{B}_w - \tilde{\sigma}^2_{w,0}\left(C + \frac{1}{n}\kappa\right)\right], \tag{25}$$

where $\kappa = (\kappa_1 \dots \kappa_m)'$ and $\widehat{B}_w$ is obtained upon making the substitutions required in Section 3.5 for the case $\mathcal{R}(X) \subseteq \mathcal{R}(Z)$.

Evidently, the only difference between ACE1 with strictly exogenous regressors and the WK estimator is found in the finite sample corrections k and κ adopted by the latter. Since k is finite and κ, under A.5, is $O(1)$, then $\tilde{\sigma}^2_{w,0}$ and $\tilde{\sigma}^2_{w,0}$ are consistent under A.1-A.6WTSLS and the further condition that $X = ZK$, where K is a $O(1)$ sequence of $(p \times k)$ matrices with uniformly f.c.r.

If the regression model includes the intercept, one proceeds as in Section 3.6, with the caution of modifying also the term κ: the formula for $\tilde{\sigma}^2_{w,0}$ is left unchanged and the formula for $\tilde{\sigma}^2_w$ modifies with $P_{[\Delta_i]} - P_{[\underline{1}]}$ replacing $P_{[\Delta_i]}$ in A, $\widehat{B}_w$, C and κ of equation (25) and $\left(n - tr\ \Delta'_i P_{[\underline{1}]} \Delta_i\right)/n$ replacing the terms of the main diagonal of A.[9]

---

[9] It is easy to prove that this approach is identical to the solution suggested by Wansbeek and Kaptein (1989) of using centered residuals, $Q_{[\underline{1}]}\widehat{\epsilon}_w$, into the empirical quadratic forms in (21).

4.1.2 The nested ECM

If the model is nested a numerically equivalent representation for $\tilde{\sigma}_w^2$ in (25) can be obtained by using the same modifications applied for ACE1, described at the end of Section 3.2, transforming κ_i as

$$\kappa_i^* = tr\left(X'Q_{[\Delta]}X\right)^{-1} X'\left(P_{[\Delta_i]} - P_{[\Delta_{i+1}]}\right)X$$

for all $i = 1, ..., m-1$, and leaving κ_m unchanged. While this nested representation is numerically equivalent to that in (25), it may be easier to implement as the modified matrix A^* is lower triangular (see Baltagi et al. (2001) for the two-way case).

4.1.3 Endogenous regressors

Davis (2002) applies Procedure I to provide a multidimensional extension of the estimator by Wansbeek and Kaptein (1989) in the general context of Section 3.2, that is of a multi-way ECM with endogenous regressors. He does not mention the moment restrictions under which completing step 2. In analogy with the previous section, the following conditional moment restriction

$$E\left(\varepsilon\varepsilon'|X, Z\right) = \Sigma \tag{26}$$

assures that the computations of step 2 in Davis (2002) are legitimate, although it must be observed that a degrees-of-freedom correction term is missing in the formula of Davis' estimator, as I show in appendix. Condition (26), however, is equivalent to

$$Var\left(\epsilon|X, Z\right) = \Sigma - E\left(\epsilon|X, Z\right)E\left(\epsilon'|X, Z\right), \tag{27}$$

which is bound to leading to undesired restrictions in a context where the researcher would rather leave the conditional moments of the composite error unrestricted, and may be even incompatible with endogenous regressors whatsoever.[10] For example, it is possible to prove that if $Var\left(\epsilon|X,Z\right)$ is a matrix of constants, as would happen if ϵ, X and Z were jointly normal, then condition (26) necessarily implies $E\left(\epsilon|X\right)=0$. In fact, by condition (27) it follows that also $E\left(\epsilon|X,Z\right)E\left(\epsilon'|X,Z\right)$ is a matrix of constants and hence so is $E\left(\epsilon|X,Z\right)$. Therefore, given $E\left(\epsilon\right)=0$, it has $E\left(\epsilon|X,Z\right)=0$ and eventually $E\left(\epsilon|X\right)=0$ by repeated application of the law of iterated expectations.[11]

With the consistency analysis of Section 3 in hand, it is easy to prove that Davis' estimator, rectified or not with the missing term and whether or not condition (26) holds, is consistent under assumptions A.1-A.6.WTSLS.

4.2 Anova-type unbiased estimators using BTSLS residuals

The Anova-type unbiased estimator derived below is the multidimensional extension of the estimator derived by Swamy and Arora (1972), through Procedure I and based on within and between residuals, in the context of a two-way ECM with nonstochastic regressors. Therefore, it is referred to throughout as the SA estimator. To work it out I maintain the same assumptions, and follow the same steps, as in Section 4.1.1.

1. Obtain the following m empirical quadratic forms based on between resid-

---

[10] Wooldridge (2002), p. 95, warns against assumptions on the conditional covariance matrix of the disturbances in the presence of endogenous variables.

[11] The computations of step 2 in Davis (2002) can be accomplished also under the weaker condition that $E\left[\varepsilon\varepsilon'|XK\right]=\Sigma$, with $K\equiv\left(X'P_{\left[Q_{[\Delta]}Z\right]}X\right)^{-1}X'P_{\left[Q_{[\Delta]}Z\right]}$. This does not change the nature of my remarks since restrictions on $E\left[\varepsilon|XK\right]$ are also difficult to accept in models with endogenous regressors.

uals

$$q_{b,i} \quad = \quad \widehat{\epsilon}'_{b,i} P_{[\Delta_i]} \widehat{\epsilon}_{b,i}$$

$i = 1, \ldots, m$, where $\widehat{\epsilon}'_{b,i}$ is the same as in (20).

2. Work out the conditional expectation of $q_{b,i}$ under $E(\epsilon\epsilon'|X) = \Sigma$ to have (details are in appendix)

$$\begin{aligned} E(q_{b,i}|X) &= (N_i - k)\,\sigma_0^2 + [n - \zeta_i]\,\sigma_i^2 \\ &+ \sum_{j \neq i}^{m} \left[tr\ \Delta'_j P_{[\Delta_i]} \Delta_j - \zeta_{ij}\right] \sigma_j^2, \end{aligned} \qquad (28)$$

   where $\zeta_i \equiv tr\ \left(X' P_{[\Delta_i]} X\right)^{-1} X' \Delta_i \Delta'_i X$ and

$$\zeta_{ij} \equiv tr\ \left(X' P_{[\Delta_i]} X\right)^{-1} X' P_{[\Delta_i]} \Delta_j \Delta'_j P_{[\Delta_i]} X,$$

   $i = 1, ..., m.$

3. The unbiased SA estimator for σ^2 , say $\tilde{\sigma}_b^2$ is eventually obtained as the solution of the system of m linear equations, equating $q_{b,i}$ to the expression for $E(q_{b,i}|X)$ obtained in step 2, with σ_0^2 replaced by $\tilde{\sigma}_{w,0}^2$:

$$\tilde{\sigma}_b^2 \quad = \quad (A - D)^{-1} \left[\widehat{B}_b - \tilde{\sigma}_{w,0}^2 \left(C - \frac{k}{n}\right)\right] \qquad (29)$$

   where

$$D = \frac{1}{n} \begin{pmatrix} \zeta_1 & \cdots & \zeta_{1i} & \cdots & \zeta_{1m} \\ \vdots & \ddots & \vdots & & \vdots \\ \zeta_{i1} & \cdots & \zeta_i & \cdots & \zeta_{im} \\ \vdots & & \vdots & \ddots & \vdots \\ \zeta_{m1} & \cdots & \zeta_{mi} & \cdots & \zeta_m \end{pmatrix}.$$

   and $\widehat{B}_b$ is obtained upon making the substitutions required in Section 3.5

for the case $\mathcal{R}(X) \subseteq \mathcal{R}(Z)$. The SA estimator $\tilde{\sigma}_b^2$ easily compares to that obtained by Baltagi and Chang (1994) for the unbalanced one-way model.

If the ECM is nested, then the quadratic forms in step 1 can be modified with $P_i \equiv P_{[\Delta_i]} - P_{[\Delta_{i+1}]}$ replacing $P_{[\Delta_i]}$ for all $i = 1, \ldots, m-1$, $P_m \equiv P_{[\Delta_m]}$, and $\widehat{\epsilon}_{b,i}$ redefined accordingly as

$$\widehat{\epsilon}_{b,i} = \left\{ I - X \left[X' P_i X \right]^{-1} X' P_i \right\} \epsilon.$$

Then, one proceeds exactly as in steps 2 and 3 to obtain the multi-way extension to the SA estimator derived in Baltagi et al. (2001) for the two-way nested model. Notice that, as for ACE3, the estimator resulting from these modifications is in general different from the one in equation (29).

5 Monte Carlo experiments

The Monte Carlo experiments of this paper study the finite-sample performance of the estimators analysed in Sections 3 and 4, with special attention to the role of unbalancedness and its measures.

5.1 The Designs

The basic framework is the same as in Davis (2002) with three novel elements. As in Davis (2002), I consider the following three-way model

$$y_{ijt} = x_{ijt}\beta + u_{1,i} + u_{2,j} + u_{3,t} + u_{0,ijt}$$

where $\beta = 1$, $x_{ijt} \sim IIN(0,1)$, $u_{0,ijt} \sim IIN\left(0, \sigma_0^2\right)$, $u_{1,i} \sim IIN\left(0, \sigma_1^2\right)$, $u_{2,j} \sim IIN\left(0, \sigma_2^2\right)$, $u_{3,t} \sim IIN\left(0, \sigma_3^2\right)$, and x_{ijt}, $u_{0,ijt}$, $u_{1,i}$, $u_{2,j}$, $u_{3,t}$ are independent random variables, $i = 1, ..., N_1$, $j = 1, ..., N_2$, $t = 1, ..., N_3$. Ten different parametrizations are considered for σ_0^2 and $\sigma^2 \equiv \left(\sigma_1^2\,\sigma_2^2\,\sigma_3^2\right)$: $\sigma_0^2 = 1 - \sum_{i=1}^{3} \sigma_i^2$ and σ^2 alternates between $(0\,0\,0)$, $(0\,0\,0.2)$, $(0\,0\,0.4)$, $(0\,0\,0.6)$, $(0\,0\,0.8)$, $(0\,0.2\,0.2)$, $(0\,0.2\,0.4)$, $(0\,0.2\,0.6)$, $(0\,0.4\,0.4)$ and $(0.2\,0.2\,0.2)$. As in Davis (2002), I consider symmetrically unbalanced structures that are uniquely identified by the triple of integers (a, b, c). This implies that along any dimension there are a subjects with group size $b \times c$, b subjects with group size $a \times c$ and c subjects with group size $a \times b$, thereby $N_1 = N_2 = N_3 = a + b + c$ and $n = 3 \times a \times b \times c$. Hence, the unbalanced structure is the same for dimensions 1, 2 and 3. For example, the commonly used Ahrens and Pincus (AP) index of unbalancedness (see Baltagi and Chang (1994), Davis (2002) and Bruno (2005), among others), defined as

$$\omega_h \equiv \frac{N_h}{\overline{g}_h \sum_{s=1}^{N_h} g_h^{-1}(s)},$$

with $\overline{g}_h$ denoting the average group size as in (8), turns out to be

$$\omega_h = \frac{(a+b+c)^2}{3\left(a^2+b^2+c^2\right)} \equiv \omega$$

$h = 1, 2, 3$.

From the theorems of Section 3 it emerges that, differently from the estimators of σ_0^2, proved to be consistent under no restrictions on unbalancedness and on the way n tends to infinity, all estimators for σ^2 are proved to be consistent under assumption A.3, which boils down to taking large unbalancedness-corrected dimension sizes ψ_h^{-1} for all $h = 1, ..., m$. For this reason, as a first novel element of the present analysis, the role played by finite values of ψ_h^{-1} is investigated in the Monte Carlo experiments. Notice that here

$$\psi_h^{-1} = \frac{9abc}{ab+ac+bc} \equiv \psi^{-1}$$

$h = 1,\ 2,\ 3$.

A broad variety of patterns $(a\,b\,c)$ must be considered to ensure enough variability for ω and ψ^{-1}. Davis (2002) focuses only on ω and considers three patterns for $n = 192$, $(4\,4\,4)$, $(2\,8\,4)$ and $(2\,16\,2)$. It turns out, though, that ω and ψ^{-1} decrease together across these patterns, so that it is hard to assess the impact of each parameter within those boundaries. Therefore, as a second novel feature of my Monte Carlo experiments, I consider four additional patterns that are peculiar to $n = 648$, as indicated in Table 1.

Table 1: Unbalanced designs

n	patterns $(a\,b\,c)$	$\psi^{-1} = \frac{N}{\bar{v}^2+1}$	ω	$\bar{v}^2$	$\det A$	$\kappa(A)$
192	4 4 4	12	1	0	0.945	1.500
	2 8 4	10.286	0.778	0.361	0.925	1.611
	2 16 2	8.471	0.505	1.361	0.879	1.850
648	6 6 6	18	1	0	0.981	1.274
	3 12 6	15.429	0.778	0.361	0.975	1.320
	3 24 3	12.706	0.505	1.361	0.959	1.423
	2 18 6	12.462	0.619	1.086	0.962	1.397

The first three patterns for $n = 648$ replicate the same group size ratios - and so the same values for both ω and $\bar{v}^2$, given that both measures are unit-free- as the corresponding patterns for $n = 192$. The fourth pattern, $(2\,18\,6)$, is suitably included to try to disentangle the impact of unbalancedness, however measured, from that of ψ^{-1} in comparison with pattern $(3\,24\,3)$. From pattern $(2\,18\,6)$ to $(3\,24\,3)$, in fact, unbalancedness increases by 22.6% as measured by ω and by 25.3% as measured by $\bar{v}^2$, in the face of a negligible, less than 2%, increase in ψ^{-1}.

The last distinctive feature of the present analysis recognizes that another way through which the unbalanced patterns may determine the finite-sample performance of the estimators of σ^2 is given by how far matrix A is from singularity. To this purpose table 1 also reports the determinant of A. A related measure, which is more informative on the identification content in the data, is given by the condition number of A, $\kappa(A)$, defined as

$$\kappa(A) \equiv \frac{\lambda_{max}(A)}{\lambda_{min}(A)},$$

where $\lambda_{max}(A)$ and $\lambda_{min}(A)$ denote the maximal and minimal singular values of A, respectively. The condition number here measures the maximum relative error in the estimators of σ^2 induced by an additive perturbation in the quadratic forms, with sizes gauged in terms of the Euclidean norm; it is reported by the last column of Table 1[12]. For example, $\kappa(A) = 1.500$ of pattern $(4\,4\,4)$ indicates that a 2% perturbation in the quadratic forms brings about at most a 3% perturbation in the estimators. The identification content as measured by $\kappa(A)$, quite satisfactory overall, tends to worsen when unbalancedness gets more severe.

5.2 Results

Table 2 reports the root mean square errors (RMSEs) of the three consistent estimators of σ_0^2, the unfeasible $\widetilde{\sigma}_0^2$, $\widehat{\sigma}_{w,0}^2$ of ACE1 and $\widehat{\sigma}_{ls,0}^2$ of ACE2, and the four consistent estimators for the remaining components, $\sigma^2 = \left(\sigma_1^2\,\sigma_2^2\,\sigma_3^2\right)$, the unfeasible $\widetilde{\sigma}^2$, $\widehat{\sigma}_w^2$ of ACE1, $\widehat{\sigma}_{ls}^2$ of ACE2 and $\widehat{\sigma}_b^2$ of ACE3. I start commenting results for the estimators of σ_0^2, and then turn to the remaining variance components.

---

[12] A is symmetric only in the two balanced patterns, and so only in those cases the singular values coincide with the eigenvalues.

5.2.1 Estimators of σ_0^2

The RMSEs of the three estimators, $\widetilde{\sigma}_0^2$, $\widehat{\sigma}_{w,0}^2$ and $\widehat{\sigma}_{ls,0}^2$ are always very close in size, virtually identical in about 50% of the cases when $n = 192$ and 77% of the cases when $n = 648$. Interestingly, by comparing $\sigma^2 = (0\,0\,0.4)$ with $\sigma^2 = (0\,0.2\,0.2)$, then $\sigma^2 = (0\,0\,0.8)$ with $\sigma^2 = (0\,0.4\,0.4)$ and finally $\sigma^2 = (0\,0\,0.6)$ with both $\sigma^2 = (0\,0.2\,0.4)$ and $\sigma^2 = (0.2\,0.2\,0.2)$ uncovers that, as far as the RMSEs of $\widetilde{\sigma}_0^2$, $\widehat{\sigma}_{w,0}^2$ and $\widehat{\sigma}_{ls,0}^2$ are concerned, it is the true value of σ_0^2 that matters and not whether the data generating process is truthfully, one-way, two-way or three-way. The theoretical finding that unbalancedness does not have a first-order impact on the large-sample performance of $\widetilde{\sigma}_0^2$, $\widehat{\sigma}_{w,0}^2$ and $\widehat{\sigma}_{ls,0}^2$ carries over into finite-samples since the RMSEs are stable across all patterns, for given sample size and parametrization. To the opposite, sample size clearly matters, with RMSEs for all estimators improving markedly from $n = 192$ to $n = 648$.

5.2.2 Estimators of σ^2

In analyzing the RMSE performance of $\widetilde{\sigma}^2$, $\widehat{\sigma}_w^2$, $\widehat{\sigma}_{ls}^2$ and $\widehat{\sigma}_b^2$, attention should center on the parameter space. When $\sigma^2 = (0\,0\,0)$ we know from corollary 1 that unbalancedness does not have a first-order impact on the quadratic forms, since the terms for which it is crucial all degenerate to zero with probability 1, and hence its role is relegated to the identification content of A. This limited impact emerges also in finite samples, evidenced by the absence of common trending patterns in the RMSEs of the three components of all estimators. In addition, the RMSEs of all estimators are overall very close to each other, with virtually zero differences in the last three patterns of $n = 648$.

Focusing on the role of unbalancedness in the estimation of the non-zero variance components, I notice that the RMSE performance of all estimators deteriorates when ω and ψ^{-1} decrease jointly for given n. To the opposite, from pattern (2 18 6) to (3 24 3) where in the face of a 22.6% reduction in ω there is only a negligible increase in ψ_h^{-1}, the RMSEs of all estimators for the non-zero variance components remain relatively stable, not worsening in more than 50% of the cases. This fact is supportive of the central role plaid by ψ_h also in finite samples, and deserves being investigated further in future Monte Carlo experiments.

I turn now to the relative performance of $\widetilde{\sigma}^2$, $\widehat{\sigma}^2_w$, $\widehat{\sigma}^2_{ls}$ and $\widehat{\sigma}^2_b$. When $n = 192$, there are the following results.

1. The RMSEs of the unfeasible $\widetilde{\sigma}^2$ and $\widehat{\sigma}^2_w$ are always very close to each other, virtually equal in most cases.

2. Strikingly, for all parametrizations and patterns, no components of $\widehat{\sigma}^2_{ls}$ have higher RMSEs than the corresponding ones in $\widetilde{\sigma}^2$ and $\widehat{\sigma}^2_w$; at least one component of $\widehat{\sigma}^2_{ls}$ has lower RMSE than the corresponding one in $\widetilde{\sigma}^2$ and $\widehat{\sigma}^2_w$; and when the model is truthfully three-dimensional $\widehat{\sigma}^2_{ls}$ strictly dominates both $\widetilde{\sigma}^2$ and $\widehat{\sigma}^2_w$ over all components in patterns (4 4 4) and (2 16 2) and over two components in pattern (2 8 4).

3. The comparison between $\widehat{\sigma}^2_{ls}$ and $\widehat{\sigma}^2_b$ is less clear-cut as only in three cases one predominates over the other. Specifically, $\widehat{\sigma}^2_{ls}$ has lower RMSEs than $\widehat{\sigma}^2_b$ over all components in two parametrizations of the two-way model in the balanced pattern, the opposite occurs in the highly unbalanced pattern when the model is three-dimensional; in all other cases the two estimators cannot be ranked with each other, although the RMSEs of the dominated components are relatively higher in $\widehat{\sigma}^2_b$ than $\widehat{\sigma}^2_{ls}$, especially when the data generating process is one or two-way.

When sample size goes up to $n = 648$, I observe the following.

1. The RMSEs of all estimators shrink and so do the differences in RMSEs

among $\widetilde{\sigma}^2$, $\widehat{\sigma}^2_w$ and $\widehat{\sigma}^2_{ls}$, so that in most cases their RMSEs are virtually equal.

2. As it happens for $n = 192$, $\widehat{\sigma}^2_b$ keeps scoring well in the unbalanced patterns when the model is truthfully three-way, resulting the best estimator in pattern (2 18 6), with about 2% and 1% lower RMSEs in the second and third components, respectively, and a stable RMSE in the first one, and almost the best in pattern (3 24 3) with about 4% and 1% lower RMSEs in the second and third components, respectively, and less than 1% higher in the first one.

Table 3 reports the RMSE results for the unbiased estimators of Section 4, $\widetilde{\sigma}^2_{w,0}$ shared by WK and SA, $\widetilde{\sigma}^2_w$ of WK and $\widetilde{\sigma}^2_b$ of SA. It turns out that the RMSEs are overall very close to, and often slightly higher than, the corresponding $\widehat{\sigma}^2_{w,0}$, $\widehat{\sigma}^2_w$ and $\widehat{\sigma}^2_b$ for either sample size, all unbalanced patterns and parametrizations, which indicates that the finite-sample corrections are immaterial for the finite sample RMSE performances of estimators.

Finally, confirming results from Baltagi et al. (2002) and Baltagi and Chang (1994), I find that when any element of σ^2 is zero, negative estimates occur in about 50% of the replications for all estimators, but as as soon as a variance component raises to a non-zero value, the frequency of negative estimates becomes negligible. When $n = 638$, negative estimates are almost never observed.

6 Empirical Application

As an empirical demonstration of the Anova-type variance components estimators, I reexamine the state level Cobb-Douglas production function estimated by Baltagi et al. (2001) on the US production data provided by Munnell (1990). The data set is a panel for 48 U.S. states grouped into nine regions over the period 1970-1986. Unbalancedness occurs only along the region dimension. In what follows $i = 1, ..., 9$ indexes the regions and g_i refers to the number of states in region i, so that the group size of region i is $g(i) = 17 \times g_i$.

Baltagi et al. (2001) estimated the following equation

$$y_{ijt} = \beta_0 + \beta_1 K_{ijt} + \beta_2 KH_{ijt} + \beta_3 KW_{ijt} + \beta_4 KO_{ijt} + \beta_5 L_{ijt} + \beta_6 U_{ijt} + \epsilon_{ijt}^{2-way}, \tag{30}$$

$i = 1, ..., 9$, $j = 1, ..., g_i$ and $t = 1, ..., 17$, where $\epsilon_{ijt}^{2-way} = u_i + v_{ij} + \varepsilon_{ijt}$ is a two-way composite error, with the state component v_{ij} nested into the region component u_i; y is the dependent variable measured as the gross state product; K is the private capital stock; L is employment in nonagricultural payrolls; the three components of the public capital stock are given by KH, highway and streets, KW, water and sewer facilities and KO, other public buildings and structures; finally U denotes the state unemployment rate and is included into the specification to capture the business cycle at the state level. All variables except U are in natural logarithms.

The production function (30) is estimated by FGLS using the Anova-type estimators described in Sections 3 and 4. For ease of comparison, I maintain strictly exogeneity of regressors and all estimators are adjusted for the presence of the intercept, as explained in Sections 3.6 and 4.1. Results are shown in Table 4. I implement ACE1, ACE2 and the simplest version of ACE3, described in Section 3.4.1, which uses the state level between residuals for both σ_u^2 and σ_v^2. The WK estimator is computed through the formulas in (24) and (25) not explicitly accounting for the nested structure; results are numerically identical to the nested WK estimates in Baltagi et al. (2001), unsurprisingly given what established in Section 4.1.2. The SA estimator is implemented as in Baltagi et al. (2001), by adjusting formula (29) for the nested structure, as explained at the end of Section 4.2, obtaining an exact replication of their estimates[13]. Table 4 also includes the estimators implemented in Baltagi et al. (2001) that have some bearing on the present analysis: the unbiased, nested adaptation of the estimator by Wallace and Hussain (1969) (WH), the Maximum likelihood estimator (ML) (both derived in Baltagi et al. (2001)) OLS and LSDV with state fixed effects.

The Anova-F test of joint significance for the state effects rejects the hypothesis of no state effects, which calls into question the validity of the OLS coefficients and standard errors estimates (see Moulton (1990)). All methods show little variation as far as estimation of the idiosyncratic error variance, σ_ε^2, is concerned (ranging from 0.0013 to 0.0015). More marked differences exist across estimation methods for the components, σ_u^2 and σ_v^2, confirming what already found in Baltagi et al. (2001) and explainable with the lack of precision that may arise from the joint effect of unbalancedness and small sample size, as established in the theoretical part of this paper. Importantly, differences tend to vanish when the comparison is restricted to estimators using the same residual vector, especially so for the pairs ACE1/WK and ACE3/SA. The ACE2 and WH estimates of σ_u^2 are instead quite far apart (0.0017 and 0.0027, respectively). The ACE3/SA comparison is particularly striking, if we think that the simplest version of ACE3 gives estimates that are almost equal to the more computationally demanding SA, incorporating finite sample corrections and using residuals from two different between regressions. Finally, to confirm another empirical finding by Baltagi et al. (2001), the FGLS coefficients and standard errors estimates are largely unaffected by differences in the variance components estimates.

The two way nested structure for the composite error forces the intraclass

---

[13] The formulas in Baltagi et al. (2001) are slightly simpler than the ones showed in Section 4 since they take into account the fact that the panel is unbalanced only along the region dimension.

correlation between states in the same region to be the same regardless of the time period. This may be detrimental for the precision of the coefficient estimates and the accuracy of standard errors and variance components estimates. I, therefore, extend the composite error in equation (30) to have the three-way (partially nested) structure

$$\epsilon_{ijt}^{3-way} = \epsilon_{ijt}^{2-way} + z_{it},$$

$i = 1, ..., 9$, $j = 1, ..., g_i$ and $t = 1, ..., 17$, where z_{it} is an error component that varies only across region and over time. This permits that the intraclass correlation between states of the same region in the same year be different from that between states of the same region in different years. Notice that z_{it} is nested into u_i and that the groups along this dimension are given by all combinations of regions and years, with their number given by $9 \times 17 = 153$ and their size varying across regions. The degree of unbalancedness along the region-year dimension, therefore, equals that along the region dimension. Results for the three-way model are reported in Table 5. In addition to the LSDV estimator treating $u_i + v_{ij}$ and z_{it} as fixed and the three-way FGLS estimators based on ACE1, ACE2, ACE3, WK, SA and WH[14], a mixed estimator of the type studied in Hussain (1969), Kang (1985) and Bruno (2009), treating the z_{it} component as fixed and the state component as random, is also considered (the region component is absorbed by the fixed interaction z_{it}).

The Anova F-test of joint significance of the z_{it} effects carried out for the LSDV model rejects the hypothesis of no interaction effects, given the state effects, at any conventional level of significance. So, on this basis, the specifications including the latent component z_{it} are to be preferred to the ones of Table 4. The regularities observed in the two-way estimates are confirmed here. Little

---

[14] I computed the unbiased, three-way WH estimator, incorporating the finite sample corrections, by adapting the formulas in Davis (2002) to the case $\mathcal{R}(X) \subseteq \mathcal{R}(Z)$.

variation is observed in the estimates of σ_ε^2 across the different estimators. It is also confirmed that the finite sample corrections have a negligible impact on the variance components estimates. Notably, ACE3 in the simplest version and SA produce virtually identical estimates. While the variance components estimates from the pair ACE1/WK stand quite far apart from those of the other Anova-type estimators, all of the implied intraclass correlations estimates, reported in Table 6, are of comparable size. In fact, there is no marked variation in the coefficients and standard errors estimates across the different FGLS estimators. Finally it is worth noting that the inclusion of the interacted time-region effects strongly reduces the estimated magnitude of the KW, KO and U coefficients for all estimators of Table 5.

7 Conclusions

Three new ACE's of variance components are derived for general unbalanced multi-way error components models with non-normal disturbances and endogenous regressors. They are easy to compute and are proved to be consistent under mild regularity conditions on the data generating process. The Monte Carlo analysis and the empirical application to US production data show that the new ACE's perform well in comparison to the unbiased methods incorporating finite sample corrections, even when the size of some dimension is very small, as in the case of the region dimension (9 groups) in the empirical application.

The consistency analysis provides new general results on the convergence in probability of quadratic forms and polynomials, whose interest, I believe, goes beyond the techniques considered in this paper. In particular, it offers novel insights into the impact of unbalancedness on the large sample properties of the variance components estimators. A suitably unbalancedness-adjusted dimension size, based on the Pearson's coefficient of variation across group sizes, turns out to be relevant both for the convergence properties and, as found in my Monte Carlo evidence, the finite-sample performances of the variance components estimators, and as such is worth controlling in empirical applications. More extensive Monte Carlo experiments, therefore, seem worth pursuing on this issue and are in the author's agenda.

I have also derived new Anova-type unbiased estimators, along with a number of new results on the existing Anova-type unbiased estimators.

Further interesting directions of future research regard the explicit treatment of spatial and time dimensions in the ECM, with a coverage of the specific inference issues raised by the presence of a spatially and serially correlated idiosyncratic error component and by dynamic specifications.

References

Abowd, J. M., Kramarz, F., Margolis, D. N., 1999. High wage workers and high wage firms. Econometrica 67, 251–333.

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

Anderson, T. W., 1971. The Statistical Analysis of Time Series. New York: Wiley.

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

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

Baltagi, B. H., Song, S. H., Jung, B. C., 2001. The unbalanced nested error component regression model. Journal of Econometrics 101, 357–381.

Baltagi, B. H., Song, S. H., Jung, B. C., 2002. A comparative study of alternative estimators for the unbalanced two-way error component regression model. Econometrics Journal.

Boumahdi, R., Chaaban, J., Thomas, A., 2006. Import demand estimation with country and product effects: Application of multi-way unbalanced panel data models to lebanese imports. In: Baltagi, B. H. (Ed.), Panel Data Econometrics. Elsevier.

Bruno, G. S. F., 2005. Approximating the bias of the lsdv estimator for dynamic unbalanced panel data models. Economics Letters 87, 361–366.

Bruno, G. S. F., 2009. Estimators and tests for unbalanced multi-way error component models with correlated effects and linearly dependent regressors. Unpublished manuscript.

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

Hussain, A., 1969. A mixed model for regressions. Biometrika 56, 327–336.

Kang, S., 1985. A note on the equivalence of specification tests in the two-factor multivariate variance components model. Journal of Econometrics 28, 193–03.

Katnelson, J., Kotz, S., 1957. On the upper limits of some measure of variability. Theoretical and applied climatology 8, 103–107.

Kendall, M. G., Stuart, A., 1969. The Advanced Theory of Statistics. Vol. 1. Charles Griffin & Company.

Maddala, G. S., Mount, T. D., 1973. A comparative study of alternative estimators for variance components models used in econometric applications. Journal of the American Statistical Association 68, 324–328.

Moulton, B. R., 1990. 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.

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

Rao, C. R., 1973. Linear Statistical Inference and Its Applications. New York: Wiley.

Searle, S. R., 1971. Linear Models. New York: Wiley.

Searle, S. R., 1982. Matrix Algebra Useful for Statistics. New York: Wiley.

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

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

Wansbeek, T., Kaptein, A., 1989. Estimation of the error components model with incomplete panels. Journal of Econometrics 41, 341–361.

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

White, H., 1982. Instrumental variables regression with independent observations. Econometrica 50 (2), 483–499.

White, H., 2001. Asymptotic Theory for Econometricians, revised edition Edition. Emerald.

Wooldridge, J. M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, MA.

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

A Tables

Table 2: RMSEs of the consistent estimators

σ_1^2, σ_2^2, σ_3^2	$\widetilde{\sigma}_0^2$	$\hat{\sigma}_{w,0}^2$	$\widehat{\sigma}_{ls,0}^2$ [a]	$\widetilde{\sigma}^2$			$\hat{\sigma}_w^2$			$\hat{\sigma}_{ls}^2$			$\hat{\sigma}_b^2$		
Pattern 4-4-4: $n = 192$, $\omega = 1$, $\psi = 0.083$															
0.0 0.0 0.0	0.164	0.163	0.162	0.129	0.025	0.029	0.129	0.025	0.029	0.128	0.024	0.029	0.129	0.024	0.028
0.0 0.0 0.2	0.131	0.130	0.130	0.105	0.035	0.100	0.105	0.036	0.101	0.104	0.035	0.100	0.111	0.034	0.100
0.0 0.0 0.4	0.098	0.098	0.098	0.082	0.051	0.176	0.083	0.051	0.176	0.082	0.050	0.175	0.105	0.049	0.175
0.0 0.0 0.6	0.065	0.065	0.065	0.064	0.069	0.252	0.064	0.069	0.252	0.064	0.068	0.250	0.113	0.066	0.251
0.0 0.0 0.8	0.033	0.033	0.033	0.054	0.088	0.328	0.055	0.089	0.328	0.054	0.088	0.326	0.133	0.084	0.327
0.0 0.2 0.2	0.098	0.098	0.098	0.087	0.109	0.106	0.088	0.110	0.106	0.087	0.108	0.105	0.099	0.108	0.106
0.0 0.2 0.4	0.065	0.065	0.066	0.070	0.117	0.179	0.070	0.117	0.179	0.070	0.116	0.178	0.101	0.115	0.179
0.0 0.2 0.6	0.033	0.033	0.034	0.060	0.127	0.255	0.060	0.127	0.255	0.060	0.126	0.253	0.117	0.125	0.254
0.0 0.4 0.4	0.033	0.033	0.034	0.066	0.194	0.187	0.066	0.194	0.187	0.066	0.193	0.186	0.104	0.193	0.187
0.2 0.2 0.2	0.065	0.065	0.066	0.120	0.123	0.121	0.120	0.123	0.121	0.119	0.121	0.120	0.127	0.118	0.118
Pattern 2-8-4: $n = 192$, 0.778, $\psi = 0.097$															
0.0 0.0 0.0	0.167	0.168	0.167	0.129	0.029	0.032	0.131	0.030	0.032	0.128	0.029	0.031	0.129	0.028	0.031
0.0 0.0 0.2	0.134	0.134	0.134	0.104	0.040	0.107	0.106	0.040	0.108	0.104	0.040	0.106	0.111	0.038	0.107
0.0 0.0 0.4	0.100	0.101	0.100	0.083	0.059	0.186	0.084	0.059	0.187	0.083	0.058	0.184	0.102	0.056	0.184
0.0 0.0 0.6	0.067	0.067	0.067	0.067	0.080	0.266	0.067	0.080	0.266	0.067	0.080	0.263	0.104	0.076	0.262
0.0 0.0 0.8	0.033	0.034	0.034	0.062	0.103	0.347	0.062	0.103	0.347	0.062	0.102	0.343	0.118	0.098	0.340
0.0 0.2 0.2	0.100	0.101	0.101	0.090	0.113	0.116	0.090	0.114	0.117	0.089	0.113	0.115	0.099	0.111	0.115
0.0 0.2 0.4	0.067	0.067	0.067	0.075	0.125	0.195	0.076	0.125	0.196	0.075	0.124	0.193	0.097	0.121	0.192
0.0 0.2 0.6	0.033	0.034	0.034	0.071	0.140	0.275	0.071	0.140	0.276	0.071	0.139	0.273	0.108	0.135	0.270
0.0 0.4 0.4	0.033	0.034	0.034	0.076	0.199	0.206	0.076	0.199	0.206	0.076	0.198	0.204	0.099	0.195	0.202
0.2 0.2 0.2	0.067	0.067	0.067	0.125	0.131	0.132	0.125	0.131	0.132	0.125	0.129	0.131	0.127	0.124	0.129
Pattern 2-16-2: $n = 192$, $\omega = 0.505$, $\psi = 0.118$															
0.0 0.0 0.0	0.158	0.158	0.157	0.116	0.035	0.037	0.118	0.035	0.037	0.115	0.035	0.037	0.116	0.034	0.037
0.0 0.0 0.2	0.127	0.126	0.126	0.096	0.049	0.114	0.097	0.049	0.114	0.096	0.049	0.114	0.098	0.049	0.112
0.0 0.0 0.4	0.095	0.095	0.094	0.079	0.072	0.202	0.080	0.072	0.202	0.079	0.072	0.202	0.083	0.071	0.199
0.0 0.0 0.6	0.063	0.063	0.063	0.067	0.097	0.292	0.067	0.097	0.292	0.067	0.097	0.291	0.076	0.096	0.286
0.0 0.0 0.8	0.032	0.032	0.032	0.063	0.124	0.383	0.064	0.124	0.383	0.063	0.123	0.382	0.078	0.121	0.375
0.0 0.2 0.2	0.095	0.095	0.095	0.086	0.125	0.128	0.086	0.125	0.128	0.085	0.125	0.128	0.088	0.122	0.124
0.0 0.2 0.4	0.063	0.063	0.063	0.076	0.144	0.216	0.076	0.144	0.216	0.076	0.143	0.215	0.080	0.139	0.209
0.0 0.2 0.6	0.032	0.032	0.032	0.073	0.165	0.305	0.073	0.165	0.305	0.073	0.164	0.304	0.079	0.159	0.297
0.0 0.4 0.4	0.032	0.032	0.032	0.080	0.226	0.232	0.080	0.226	0.232	0.079	0.224	0.230	0.082	0.218	0.223
0.2 0.2 0.2	0.063	0.063	0.063	0.125	0.144	0.149	0.125	0.144	0.149	0.124	0.143	0.148	0.123	0.136	0.143

---

[a] $\sigma_0^2 = 1 - \sum_{i=1}^{3} \sigma_i^2$

Table 2 (continued)

σ_1^2, σ_2^2, σ_3^2	$\widetilde{\sigma}_0^2$	$\hat{\sigma}_{w,0}^2$	$\widehat{\sigma}_{ls,0}^2$	$\widetilde{\sigma}^2$			$\hat{\sigma}_w^2$			$\hat{\sigma}_{ls}^2$			$\hat{\sigma}_b^2$		
Pattern 6-6-6: $n = 648$, $\omega = 1$, $\psi = 0.055$															
0.0 0.0 0.0	0.078	0.078	0.078	0.052	0.009	0.009	0.052	0.009	0.010	0.052	0.009	0.010	0.052	0.009	0.009
0.0 0.0 0.2	0.063	0.063	0.063	0.042	0.018	0.074	0.042	0.018	0.074	0.042	0.018	0.074	0.047	0.018	0.073
0.0 0.0 0.4	0.047	0.047	0.047	0.034	0.031	0.139	0.034	0.031	0.139	0.034	0.031	0.139	0.053	0.031	0.138
0.0 0.0 0.6	0.031	0.031	0.031	0.028	0.044	0.205	0.028	0.045	0.205	0.028	0.044	0.205	0.067	0.044	0.203
0.0 0.0 0.8	0.016	0.016	0.016	0.027	0.058	0.270	0.027	0.058	0.270	0.027	0.058	0.270	0.085	0.058	0.267
0.0 0.2 0.2	0.047	0.047	0.047	0.038	0.076	0.078	0.038	0.076	0.078	0.038	0.076	0.078	0.046	0.075	0.077
0.0 0.2 0.4	0.031	0.031	0.031	0.032	0.083	0.143	0.032	0.083	0.143	0.033	0.083	0.143	0.055	0.082	0.141
0.0 0.2 0.6	0.016	0.016	0.017	0.031	0.092	0.208	0.031	0.092	0.208	0.031	0.092	0.208	0.072	0.091	0.205
0.0 0.4 0.4	0.016	0.016	0.017	0.034	0.143	0.148	0.034	0.143	0.148	0.034	0.143	0.147	0.060	0.142	0.146
0.2 0.2 0.2	0.031	0.031	0.031	0.082	0.083	0.085	0.082	0.083	0.085	0.082	0.083	0.085	0.085	0.081	0.084
Pattern 3-12-6: $n = 648$, $\omega = 0.778$, $\psi = 0.065$															
0.0 0.0 0.0	0.082	0.082	0.082	0.054	0.010	0.010	0.054	0.010	0.010	0.054	0.010	0.010	0.054	0.010	0.010
0.0 0.0 0.2	0.065	0.066	0.066	0.044	0.022	0.076	0.045	0.022	0.076	0.045	0.022	0.076	0.048	0.022	0.074
0.0 0.0 0.4	0.049	0.049	0.049	0.037	0.039	0.143	0.037	0.039	0.143	0.037	0.039	0.143	0.050	0.039	0.140
0.0 0.0 0.6	0.033	0.033	0.033	0.033	0.056	0.211	0.033	0.056	0.211	0.033	0.056	0.211	0.059	0.056	0.207
0.0 0.0 0.8	0.016	0.016	0.016	0.034	0.073	0.279	0.033	0.073	0.279	0.034	0.073	0.278	0.072	0.073	0.273
0.0 0.2 0.2	0.049	0.049	0.049	0.039	0.085	0.082	0.039	0.085	0.082	0.039	0.085	0.081	0.045	0.084	0.080
0.0 0.2 0.4	0.033	0.033	0.033	0.034	0.093	0.148	0.034	0.094	0.148	0.034	0.093	0.147	0.050	0.092	0.145
0.0 0.2 0.6	0.016	0.016	0.017	0.035	0.103	0.215	0.035	0.103	0.215	0.035	0.103	0.214	0.062	0.102	0.210
0.0 0.4 0.4	0.016	0.016	0.017	0.037	0.160	0.154	0.037	0.160	0.154	0.037	0.160	0.154	0.055	0.157	0.151
0.2 0.2 0.2	0.033	0.033	0.033	0.089	0.090	0.089	0.089	0.091	0.089	0.089	0.090	0.089	0.091	0.087	0.087
Pattern 3-24-3: $n = 648$, $\omega = 0.505$, $\psi = 0.079$															
0.0 0.0 0.0	0.081	0.081	0.081	0.055	0.012	0.013	0.055	0.012	0.013	0.055	0.012	0.013	0.055	0.012	0.013
0.0 0.0 0.2	0.065	0.065	0.065	0.046	0.025	0.083	0.046	0.025	0.083	0.045	0.025	0.083	0.047	0.025	0.082
0.0 0.0 0.4	0.049	0.049	0.049	0.038	0.045	0.157	0.038	0.045	0.157	0.038	0.045	0.157	0.043	0.045	0.155
0.0 0.0 0.6	0.033	0.032	0.032	0.035	0.065	0.231	0.035	0.065	0.231	0.035	0.065	0.231	0.046	0.065	0.228
0.0 0.0 0.8	0.016	0.016	0.016	0.038	0.086	0.305	0.038	0.086	0.306	0.038	0.086	0.305	0.053	0.086	0.302
0.0 0.2 0.2	0.049	0.049	0.049	0.041	0.093	0.090	0.041	0.093	0.090	0.041	0.093	0.090	0.044	0.090	0.089
0.0 0.2 0.4	0.033	0.032	0.032	0.037	0.103	0.162	0.037	0.103	0.162	0.037	0.103	0.162	0.044	0.100	0.160
0.0 0.2 0.6	0.016	0.016	0.016	0.039	0.116	0.236	0.039	0.116	0.236	0.039	0.116	0.235	0.050	0.113	0.233
0.0 0.4 0.4	0.016	0.016	0.016	0.041	0.177	0.170	0.041	0.177	0.170	0.041	0.176	0.170	0.049	0.171	0.168
0.2 0.2 0.2	0.033	0.032	0.032	0.096	0.103	0.098	0.096	0.103	0.098	0.096	0.103	0.098	0.097	0.099	0.097
Pattern 2-18-6: $n = 648$, $\omega = 0.619$, $\psi = 0.080$															
0.0 0.0 0.0	0.081	0.080	0.080	0.053	0.011	0.012	0.053	0.011	0.012	0.053	0.011	0.012	0.053	0.011	0.012
0.0 0.0 0.2	0.064	0.064	0.064	0.044	0.025	0.082	0.044	0.025	0.082	0.043	0.025	0.082	0.045	0.025	0.082
0.0 0.0 0.4	0.048	0.048	0.048	0.037	0.043	0.157	0.037	0.043	0.157	0.037	0.043	0.157	0.042	0.043	0.156
0.0 0.0 0.6	0.032	0.032	0.032	0.034	0.062	0.232	0.034	0.062	0.232	0.034	0.062	0.232	0.044	0.062	0.231
0.0 0.0 0.8	0.016	0.016	0.016	0.036	0.081	0.307	0.036	0.081	0.307	0.036	0.081	0.307	0.052	0.081	0.305
0.0 0.2 0.2	0.048	0.048	0.048	0.038	0.090	0.089	0.038	0.090	0.089	0.038	0.090	0.089	0.040	0.089	0.088
0.0 0.2 0.4	0.032	0.032	0.032	0.035	0.098	0.162	0.035	0.098	0.162	0.035	0.098	0.162	0.040	0.097	0.161
0.0 0.2 0.6	0.016	0.016	0.016	0.037	0.109	0.236	0.037	0.109	0.236	0.037	0.109	0.236	0.047	0.108	0.234
0.0 0.4 0.4	0.016	0.016	0.016	0.040	0.170	0.171	0.039	0.169	0.171	0.039	0.169	0.171	0.046	0.168	0.170
0.2 0.2 0.2	0.032	0.032	0.032	0.089	0.103	0.101	0.089	0.103	0.101	0.089	0.103	0.101	0.089	0.101	0.100

Table 3: RMSEs of the unbiased estimators

σ_1^2, σ_2^2, σ_3^2	$\widetilde{\sigma}_{w,0}^2$	$\widetilde{\sigma}_w^2$			$\widetilde{\sigma}_b^2$		
Pattern 4-4-4							
0.0 0.0 0.0	0.164	0.131	0.025	0.029	0.130	0.026	0.030
0.0 0.0 0.2	0.131	0.106	0.036	0.101	0.109	0.037	0.107
0.0 0.0 0.4	0.099	0.083	0.051	0.176	0.098	0.053	0.187
0.0 0.0 0.6	0.066	0.065	0.069	0.252	0.099	0.072	0.267
0.0 0.0 0.8	0.033	0.055	0.089	0.328	0.113	0.092	0.349
0.0 0.2 0.2	0.099	0.088	0.110	0.106	0.095	0.115	0.113
0.0 0.2 0.4	0.066	0.070	0.117	0.179	0.091	0.123	0.191
0.0 0.2 0.6	0.033	0.060	0.127	0.255	0.101	0.134	0.271
0.0 0.4 0.4	0.033	0.066	0.194	0.187	0.093	0.206	0.199
0.2 0.2 0.2	0.066	0.121	0.123	0.121	0.128	0.130	0.126
Pattern 2-8-4							
0.0 0.0 0.0	0.169	0.131	0.030	0.032	0.130	0.030	0.033
0.0 0.0 0.2	0.135	0.106	0.040	0.108	0.111	0.040	0.113
0.0 0.0 0.4	0.102	0.084	0.059	0.187	0.099	0.059	0.196
0.0 0.0 0.6	0.068	0.068	0.080	0.266	0.098	0.081	0.280
0.0 0.0 0.8	0.034	0.062	0.103	0.347	0.108	0.104	0.363
0.0 0.2 0.2	0.102	0.091	0.114	0.117	0.099	0.117	0.123
0.0 0.2 0.4	0.068	0.076	0.125	0.196	0.095	0.128	0.206
0.0 0.2 0.6	0.034	0.071	0.140	0.276	0.103	0.143	0.289
0.0 0.4 0.4	0.034	0.076	0.199	0.206	0.097	0.205	0.217
0.2 0.2 0.2	0.068	0.126	0.131	0.132	0.131	0.133	0.139
Pattern 2-16-2							
0.0 0.0 0.0	0.159	0.118	0.035	0.037	0.117	0.035	0.038
0.0 0.0 0.2	0.127	0.098	0.049	0.114	0.098	0.050	0.117
0.0 0.0 0.4	0.095	0.080	0.072	0.202	0.082	0.074	0.207
0.0 0.0 0.6	0.064	0.067	0.097	0.292	0.073	0.099	0.298
0.0 0.0 0.8	0.032	0.064	0.124	0.383	0.074	0.126	0.390
0.0 0.2 0.2	0.095	0.086	0.125	0.128	0.088	0.126	0.130
0.0 0.2 0.4	0.064	0.076	0.144	0.216	0.079	0.144	0.218
0.0 0.2 0.6	0.032	0.073	0.165	0.305	0.078	0.165	0.309
0.0 0.4 0.4	0.032	0.080	0.226	0.232	0.082	0.226	0.233
0.2 0.2 0.2	0.064	0.125	0.144	0.149	0.126	0.143	0.150

Table 3 (continued)

σ_1^2, σ_2^2, σ_3^2	$\widetilde{\sigma}_{w,0}^2$	$\widetilde{\sigma}_w^2$			$\widetilde{\sigma}_b^2$		
Pattern 6-6-6							
0.0 0.0 0.0	0.079	0.052	0.009	0.010	0.052	0.009	0.010
0.0 0.0 0.2	0.063	0.042	0.018	0.074	0.046	0.019	0.077
0.0 0.0 0.4	0.047	0.034	0.031	0.139	0.048	0.033	0.144
0.0 0.0 0.6	0.031	0.028	0.045	0.205	0.059	0.047	0.212
0.0 0.0 0.8	0.016	0.027	0.058	0.270	0.074	0.061	0.279
0.0 0.2 0.2	0.047	0.038	0.076	0.078	0.044	0.079	0.081
0.0 0.2 0.4	0.031	0.033	0.083	0.143	0.050	0.086	0.147
0.0 0.2 0.6	0.016	0.031	0.092	0.208	0.063	0.095	0.215
0.0 0.4 0.4	0.016	0.034	0.143	0.148	0.054	0.149	0.152
0.2 0.2 0.2	0.031	0.082	0.083	0.085	0.086	0.086	0.088
Pattern 3-12-6							
0.0 0.0 0.0	0.082	0.054	0.010	0.010	0.055	0.011	0.011
0.0 0.0 0.2	0.066	0.045	0.022	0.076	0.047	0.024	0.077
0.0 0.0 0.4	0.049	0.037	0.039	0.143	0.047	0.041	0.146
0.0 0.0 0.6	0.033	0.033	0.056	0.211	0.053	0.059	0.215
0.0 0.0 0.8	0.016	0.034	0.073	0.279	0.065	0.077	0.284
0.0 0.2 0.2	0.049	0.039	0.085	0.082	0.043	0.088	0.083
0.0 0.2 0.4	0.033	0.034	0.094	0.148	0.046	0.097	0.151
0.0 0.2 0.6	0.016	0.035	0.103	0.215	0.056	0.108	0.219
0.0 0.4 0.4	0.016	0.037	0.160	0.154	0.051	0.165	0.158
0.2 0.2 0.2	0.033	0.089	0.091	0.089	0.091	0.093	0.091
Pattern 3-24-3							
0.0 0.0 0.0	0.081	0.055	0.012	0.013	0.055	0.012	0.013
0.0 0.0 0.2	0.065	0.046	0.025	0.083	0.047	0.026	0.084
0.0 0.0 0.4	0.049	0.038	0.045	0.157	0.042	0.047	0.159
0.0 0.0 0.6	0.033	0.035	0.065	0.231	0.044	0.068	0.235
0.0 0.0 0.8	0.016	0.038	0.086	0.306	0.051	0.089	0.310
0.0 0.2 0.2	0.049	0.041	0.093	0.090	0.043	0.094	0.091
0.0 0.2 0.4	0.033	0.037	0.103	0.162	0.043	0.105	0.165
0.0 0.2 0.6	0.016	0.039	0.116	0.236	0.048	0.118	0.239
0.0 0.4 0.4	0.016	0.041	0.177	0.170	0.048	0.178	0.173
0.2 0.2 0.2	0.033	0.096	0.103	0.098	0.098	0.104	0.100
Pattern 2-18-6							
0.0 0.0 0.0	0.081	0.053	0.011	0.012	0.053	0.012	0.012
0.0 0.0 0.2	0.064	0.044	0.025	0.082	0.045	0.026	0.084
0.0 0.0 0.4	0.048	0.037	0.043	0.157	0.041	0.045	0.160
0.0 0.0 0.6	0.032	0.034	0.062	0.232	0.043	0.064	0.236
0.0 0.0 0.8	0.016	0.036	0.081	0.307	0.050	0.084	0.313
0.0 0.2 0.2	0.048	0.038	0.090	0.089	0.040	0.092	0.091
0.0 0.2 0.4	0.032	0.035	0.098	0.162	0.040	0.101	0.165
0.0 0.2 0.6	0.016	0.037	0.109	0.236	0.045	0.113	0.240
0.0 0.4 0.4	0.016	0.039	0.169	0.171	0.045	0.174	0.174
0.2 0.2 0.2	0.032	0.089	0.103	0.101	0.090	0.106	0.103

Table 4: **Cobb-Douglas production function estimates with state and region effects** - $N_{state} = 48$, balanced; $N_{region} = 9$, $\psi^{-1}_{region} = 8$, $AP_{region} = 0.88$. Standard errors in parentheses. Estimates for WH and ML are from Table 6 in Baltagi et al. (2001).

	OLS	LSDV	FGLS ACE1	FGLS WK	FGLS ACE2	FGLS WH	FGLS ACE3[a]	FGLS SA	ML
Constant	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)	2.129 (0.154)
K	0.312 (0.011)	0.235 (0.026)	0.264 (0.022)	0.264 (0.027)	0.276 (0.021)	0.273 (0.021)	0.274 (0.020)	0.274 (0.020)	0.267 (0.021)
L	0.550 (0.016)	0.801 (0.030)	0.760 (0.027)	0.758 (0.027)	0.735 (0.027)	0.742 (0.026)	0.740 (0.025)	0.740 (0.025)	0.754 (0.026)
KH	0.059 (0.015)	0.077 (0.031)	0.072 (0.024)	0.072 (0.024)	0.073 (0.023)	0.075 (0.023)	0.072 (0.022)	0.073 (0.022)	0.071 (0.023)
KW	0.119 (0.012)	0.079 (0.015)	0.076 (0.014)	0.076 (0.014)	0.077 (0.014)	0.076 (0.014)	0.076 (0.014)	0.076 (0.014)	0.076 (0.014)
KO	0.009 (0.012)	-0.115 (0.018)	-0.102 (0.017)	-0.102 (0.017)	-0.092 (0.018)	-0.095 (0.017)	-0.095 (0.017)	-0.094 (0.017)	-0.100 (0.017)
Unemp	-0.007 (0.001)	-0.005 (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)	-0.006 (0.001)
fixed state effects	-	YES[b]	-	-	-	-	-	-	-
σ^2_ε	0.0073	0.0014	0.0014	0.0014	0.0015	0.0014	0.0014	0.0014	0.0013
σ^2_u	-	-	0.0024	0.0022	0.0017	0.0027	0.0013	0.0015	0.0015
σ^2_v	-	-	0.0072	0.0069	0.0043	0.0045	0.0044	0.0043	0.0063

---

[a] Between residuals are from the state level, as described in Section 3.4.1

[b] F(47, 762) = 76.71, p-value=0.0000

Table 5: **Cobb-Douglas production function estimates with state, region and interacted time-region effects -** $N_{state} = 48$, balanced; $N_{region} = 9$, $\psi^{-1}_{region} = 8$, $AP_{region} = 0.88$ $N_{time*reg} = 153$, $\psi^{-1}_{time*reg} = 136$, $AP_{time*reg} = 0.88$. Standard errors in parentheses.

	LSDV	Mixed ACE1	FGLS ACE1	FGLS WK	FGLS ACE2	FGLS WH	FGLS ACE3[a]	FGLS SA
Constant	-	-	2.297 (0.181)	2.286 (0.177)	2.154 (0.151)	2.159 (0.154)	2.201 (0.146)	2.198 (0.146)
K	0.128 (0.030)	0.158 (0.028)	0.198 (0.023)	0.201 (0.023)	0.236 (0.021)	0.233 (0.021)	0.223 (0.021)	0.223 (0.021)
L	0.871 (0.035)	0.814 (0.031)	0.798 (0.028)	0.794 (0.028)	0.749 (0.027)	0.756 (0.027)	0.758 (0.026)	0.758 (0.026)
KH	0.064 (0.031)	0.080 (0.027)	0.071 (0.025)	0.071 (0.024)	0.078 (0.023)	0.079 (0.023)	0.078 (0.022)	0.079 (0.022)
KW	0.036 (0.016)	0.032 (0.015)	0.047 (0.014)	0.048 (0.014)	0.052 (0.014)	0.053 (0.014)	0.046 (0.014)	0.046 (0.014)
KO	-0.021 (0.018)	-0.023 (0.017)	-0.048 (0.016)	-0.049 (0.016)	-0.050 (0.016)	-0.053 (0.016)	-0.042 (0.016)	-0.041 (0.016)
Unemp	-0.000 (0.001)	-0.001 (0.001)	-0.003 (0.001)	-0.003 (0.001)	-0.004 (0.001)	-0.004 (0.001)	-0.003 (0.001)	-0.003 (0.001)
fixed state effects	YES	-	-	-	-	-	-	-
fixed region*time effects	YES[b]	YES	-	-	-	-	-	-
σ^2_ε	0.0009	0.0009	0.0009	0.0009	0.0011	0.0010	0.0009	0.0009
σ^2_u	-	-	0.0048	0.0041	0.0016	0.0027	0.0013	0.0014
σ^2_v	-	0.0099	0.0099	0.0090	0.0044	0.0045	0.0044	0.0043
σ^2_z	-	-	0.0006	0.0006	0.0004	0.0004	0.0007	0.0007

---

[a] Between residuals are from the state level, as described in Section 3.4.1

[b] F(144, 618) = 3.54, p-value=0.000

Table 6: **Intraclass correlation estimates for the 3-way estimators**

	ACE1	WK	ACE2	WH	ACE3	SA
$\frac{\sigma_u^2+\sigma_v^2}{\sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2+\sigma_z^2}$	0.91	0.90	0.80	0.84	0.78	0.78
$\frac{\sigma_u^2+\sigma_z^2}{\sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2+\sigma_z^2}$	0.33	0.32	0.27	0.36	0.27	0.27
$\frac{\sigma_u^2}{\sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2+\sigma_z^2}$	0.30	0.28	0.21	0.31	0.18	0.18

B Lemmata and proofs

This appendix contains the proofs of all theorems in the paper. Before that, it provides the lemmata containing the probability and algebraic results that are used in the proofs.

The following lemma provides the formulas for the expectation and the variance of a quadratic form in (possibly) non-normal random variables under more general conditions than A.1.

Lemma 1 *(Lemmata 3.4.3 and 3.4.4 in Anderson (1971)) Let* $\sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}u_iu_j = u'Su$ *where* S *is symmetric and* $E(u_i) = 0$, $E\left(u_i^2\right) = \sigma^2 < \infty$, $E(u_iu_j) = 0$, $i \neq j$, $E(u_iu_ju_ku_l) = 0$ *unless the subscripts are 1) equal in pairs, in which case* $E\left(u_i^2u_j^2\right) = \sigma^4$, $i \neq j$, *or 2) all equal, in which case* $E\left(u_i^4\right) = \kappa_4 + 3\sigma^4 < \infty$. *Then* $E(u'Su) = \sigma^2 trS$ *and*

$$Var(u'Su) = \kappa_4 tr\ S(I * S) + 2\sigma^4 trS^2$$

where $*$ *denotes the Hadamard product.*

Notice that $tr\ S(I * S) = \sum_{i=1}^{n} s_{ii}^2$. The following lemma contains results on the mean and the variance of polynomials in independent zero-mean random vectors.

Lemma 2 *Let* $u = (u_1...u_m)'$ *and* $v = (v_1...v_n)'$ *be independent zero mean random vectors, each satisfying the assumptions of Lemma 1, with* $E\left(u_i^2\right) = \sigma_u^2$ *and* $E\left(v_i^2\right) = \sigma_v^2$; *and let* $\sum_{i=1}^{m}\sum_{j=1}^{n} s_{ij}u_iv_j = u'Sv$ *where* S *is a* $(m \times n)$ *matrix. Then* $E(u'Sv) = 0$ *and*

$$Var(u'Sv) = \sigma_u^2\sigma_v^2 tr\ S'S.$$

Proof. By the independence of u and v and the fact that u and v are zero mean vectors it follows that

$$E\left(\sum_{i=1}^{m}\sum_{j=1}^{n} s_{ij}u_i v_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} s_{ij}E\left(v_j\right)E\left(u_i\right) = 0$$

and hence

$$\begin{aligned} Var\left(u'Sv\right) &= E\left(\sum_{i=1}^{m}\sum_{j=1}^{n} s_{ij}u_i v_j\right)^2 \qquad (31) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} s_{ij}^2 E\left(u_i^2 v_j^2\right) \\ &= \sigma_u^2\sigma_v^2 \sum_{i=1}^{m}\sum_{j=1}^{n} s_{ij}^2 \\ &= \sigma_u^2\sigma_v^2 tr\, S'S, \end{aligned}$$

where the second equality in (31) follows from the fact that the cross-products terms in $E\left(\sum_{i=1}^{m}\sum_{j=1}^{n} s_{ij}u_i v_j\right)^2$ are of the following three types

1. $s_{ij}s_{ik}E\left(u_i^2 v_j v_k\right),$
2. $s_{ik}s_{jk}E\left(u_i u_j v_k^2\right),$
3. $s_{ij}s_{kl}E\left(u_i u_j v_k v_l\right),$

and $E\left(u_i^2 v_j v_k\right) = E\left(u_i u_j v_k^2\right) = E\left(u_i u_j v_k v_l\right) = 0$ by the independence and zero mean assumptions; the third equality in (31) follows from the independence of u and v; and the last from the fact that the sum of squares of elements of a matrix S can be expressed as $tr\, S'S$ (see Searle (1982), p. 46). ■

The following lemma collects properties of orthogonal projector matrices that are used in the paper.

Lemma 3 *Let S be a $(n \times m)$ non-zero matrix, consider the $(n \times n)$ orthogonal projection matrix $P_{[S]} \equiv S\left(S'S\right)^{-}S'$ and let $\pi\left(i,j\right)$ denote the element of $P_{[S]}$*

in row i and column j, then a) $0 \leq \pi(i,i) \leq 1$, $i = 1,...n$; b) $|\pi(i,j)| \leq 1$, $i,j = 1,...n$ c) $tr\left(P_{[S]}\right) = r\left(P_{[S]}\right)$ d) $r\left(P_{[S]}\right) = r(S)$.

Proof. By the properties of orthogonal projectors, $P_{[S]}S = S$, so $P_{[S]}$ cannot have all zero elements. Also, $P_{[S]}$ is idempotent and symmetric, therefore

$$\pi(i,i) = \pi^2(i,i) + \sum_{j \neq i}^{n} \pi^2(i,j) \geq 0$$

$i = 1,...n$, and so

$$\pi^2(i,i) \leq \pi(i,i) \text{ and } \sum_{j \neq i}^{n} \pi^2(i,j) \leq \pi(i,i)$$

proving a) and b). To prove c) see Rao (1973) 3.(d), p. 34. d) Upon noticing that $(S'S)^{-} S'$ is a generalized inverse of S, apply result 3.(e) in Rao (1973), p. 34. ∎

The following concepts are important for the asymptotic analysis of the ACE's.

Definition 4 *Let $g_i(s)$ denote the number of observations of group s in dimension i, $s = 1,...,N_i$, $i = 1,...,m$.*

Definition 5 *Let $g_{ij}(s,t)$ denote the number of common observations between group s in dimension i and group t in dimension j, $s = 1,...,N_i$, $t = 1,...,N_j$ and $i,j = 1,...,m$*

Clearly, $g_{ij}(s,t) = g_{ji}(t,s)$. Some important properties of g_i and g_{ij} in terms of their relationship with the dummy matrices Δ_i are collected in the following Lemma .

Lemma 4 *Assume a) of A.2. Then the following hold. a) $\Delta_i'\Delta_i$ is a $(N_i \times N_i)$*

diagonal matrix and $g_i(s)$ *is the* (s,s).th *element of* $\Delta_i'\Delta_i$ *with*

$$\sum_{s=1}^{N_i} g_i(s) = n, \tag{32}$$

$i = 1, ..., m$; $g_{ij}(s,t)$ *is the* (s,t).th *element of the* $(N_i \times N_j)$ *matrix* $\Delta_i'\Delta_j$ *and*

$$\sum_{t=1}^{N_j} g_{ij}(s,t) = g_i(s) \tag{33}$$

$s = 1, ..., N_i$ *and* $i, j = 1, ..., m$;

$$\begin{aligned} &b) & tr\ \Delta_j' P_{[\Delta_i]} \Delta_j &= \sum_{t=1}^{N_j}\sum_{s=1}^{N_i} \frac{g_{ij}^2(s,t)}{g_i(s)}, \\ &c) & tr\ \Delta_j' P_{[\Delta_i]} \Delta_j \left(I * \Delta_j' P_{[\Delta_i]} \Delta_j\right) &= \sum_{t=1}^{N_j} \left[\sum_{s=1}^{N_i} \frac{g_{ij}^2(s,t)}{g_i(s)}\right]^2, \end{aligned}$$

$i, j = 1, ..., m$;

$$d) \quad tr\ \Delta_j' P_{[\Delta_i]} \Delta_k \Delta_k' P_{[\Delta_i]} \Delta_j = \sum_{t=1}^{N_k}\sum_{r=1}^{N_j} \left(\sum_{s=1}^{N_i} \frac{g_{ik}(s,t)\, g_{ij}(s,r)}{g_i(s)}\right)^2,$$

$i, j, k = 1, ..., m$.

Proof. a) Obvious. b) Since $\Delta_i'\Delta_i$ is diagonal, $(\Delta_i'\Delta_i)^{-1}$ is also diagonal with generic diagonal element given by $1/g_i(s)$, $s = 1, ..., N_i$. Hence, $g_{ij}(s,t)/g_i(s)$ is the (s,t).th element of the $(N_i \times N_j)$ matrix $F_{ij} \equiv (\Delta_i'\Delta_i)^{-1}\Delta_i'\Delta_j$, $i, j = 1, ..., m$, and expresses the number of common observations in relative terms, that is as a portion of the size of group s in dimension i. Since $P_{[\Delta_i]}\Delta_j = \Delta_i F_{ij}$, $P_{[\Delta_i]}\Delta_j$ is the $(n \times N_j)$ matrix made by the rows of F_{ij}, with each row s in F_{ij}

$$f_{ij}(s) \equiv \left[\frac{g_{ij}(s,1)}{g_i(s)}\ \frac{g_{ij}(s,2)}{g_i(s)} ... \frac{g_{ij}(s,N_j)}{g_i(s)}\right]$$

being repeated $g_i(s)$ times in $P_{[\Delta_i]}\Delta_j$. Therefore, the sum of the squared elements of $P_{[\Delta_i]}\Delta_j$, that is $tr\ \Delta_j' P_{[\Delta_i]}\Delta_j$, is given by

$$\sum_{s=1}^{N_i}\sum_{t=1}^{N_j}\frac{g_{ij}^2(s,t)}{g_i^2(s)}g_i(s)=\sum_{s=1}^{N_i}\sum_{t=1}^{N_j}\frac{g_{ij}^2(s,t)}{g_i(s)}.$$

c) Given that $P_{[\Delta_i]}\Delta_j = \Delta_i F_{ij}$, the (t,r).th element of the symmetric $(N_j \times N_j)$ matrix $\Delta_j' P_{[\Delta_i]}\Delta_j$ is worked out as

$$\sum_{s=1}^{N_i}\frac{g_{ij}(s,t)\, g_{ij}(s,r)}{g_i(s)} \tag{34}$$

$t, r = 1, ..., N_j$, with

$$\sum_{s=1}^{N_i}\frac{g_{ij}^2(s,t)}{g_i(s)}$$

being the t.th element of the main diagonal of $\Delta_j' P_{[\Delta_i]}\Delta_j$. To derive expression (34) I have simply gone through the product between the $t.th$ row of Δ_j' and the $r.th$ column of $P_{[\Delta_i]}\Delta_j$. The $t.th$ row of Δ_j' has $g_j(t) = \sum_{s=1}^{N_i} g_{ij}(s,t)$ unity elements so that, of the $g_i(1)$ elements $g_{ij}(1,r)/g_i(1)$ in the $r.th$ column of $P_{[\Delta_i]}\Delta_j$, $g_{ij}(1,t)$ are picked up and summed to get

$$\frac{g_{ij}(1,t)\, g_{ij}(1,r)}{g_i(1)},$$

this is added to

$$\frac{g_{ij}(2,t)\, g_{ij}(2,r)}{g_i(2)},$$

obtained in the same way, and so on, until the $N_i.th$ group in dimension i, to obtain expression (34).

Hence,

$$tr\ \Delta_j' P_{[\Delta_i]}\Delta_j\left(I * \Delta_j' P_{[\Delta_i]}\Delta_j\right)=\sum_{t=1}^{N_j}\left[\sum_{s=1}^{N_i}\frac{g_{ij}^2(s,t)}{g_i(s)}\right]^2.$$

d) Consider the $(N_k \times N_j)$ matrix $\Delta_k' P_{[\Delta_i]} \Delta_j$, $i, j, k = 1, ..., m$, of which $\Delta_j' P_{[\Delta_i]} \Delta_j$ is a particular case. The fact that $P_{[\Delta_i]} \Delta_j = \Delta_i F_{ij}$ is again useful to understand the expression for the the (t, r) .th element of $\Delta_k' P_{[\Delta_i]} \Delta_j$, which is given by

$$\sum_{s=1}^{N_i} \frac{g_{ik}(s,t)\, g_{ij}(s,r)}{g_i(s)}$$

(the difference with the (t, r) .th element of $\Delta_j' P_{[\Delta_i]} \Delta_j$ lies in the fact that, here, t and r do not necessarily represent clusters in the same dimension). Hence, the sum of its squared elements is given by

$$tr\ \Delta_j' P_{[\Delta_i]} \Delta_k \Delta_k' P_{[\Delta_i]} \Delta_j = \sum_{t=1}^{N_k} \sum_{r=1}^{N_j} \left(\sum_{s=1}^{N_i} \frac{g_{ik}(s,t)\, g_{ij}(s,r)}{g_i(s)} \right)^2 .$$

■

In standard panels with $m = 2$, where Δ_1 is the $(n \times N_1)$ matrix of individual dummies and Δ_2 is the $(n \times N_2)$ matrix of time dummies, $\Delta_1' \Delta_2$ reduces to a $(N_1 \times N_2)$ matrix of zeros and ones indicating the absence or presence of a given individual in a given year. Thereby, $g_{12}(s,t) = g_{12}^2(s,t)$, so that

$$tr\ \Delta_i' P_{[\Delta_j]} \Delta_i = \sum_{s=1}^{N_i} \sum_{t=1}^{N_j} \frac{g_{ij}^2(s,t)}{g_j(t)} = \sum_{t=1}^{N_j} \sum_{s=1}^{N_i} \frac{g_{ij}(s,t)}{g_j(t)} = N_j,\ i \neq j = 1, 2$$

(see Wansbeek and Kaptein (1989)).

Theorem 1 Assume A.1-A.2. Then, for all $i, j = 1, \ldots, m$,

$$\text{a) } \widetilde{\sigma}_0^2 \underset{p}{\rightarrow} \sigma_0^2$$

$$\text{b) } \frac{u_0' P_{[\Delta_i]} u_0}{n} - \frac{N_i}{n} \sigma_0^2 \underset{p}{\rightarrow} 0$$

$$\text{c) } \frac{u_0' P_{[\Delta_i]} \Delta_j u_j}{n} \underset{p}{\rightarrow} 0$$

as $n \rightarrow \infty$.

Proof. a) A.1 implies all assumptions of Lemma 1, therefore $E\left(\widetilde{\sigma}_0^2\right) = \sigma_0^2$ and the variance of $\widetilde{\sigma}_0^2$ is

$$Var\left(\frac{u_0' Q_{[\Delta]} u_0}{n - r(\Delta)}\right) = \left[\frac{1}{n - r(\Delta)}\right]^2 \kappa_{0,4} \sum_{i=1}^{n} \phi^2(i,i) + \frac{2\sigma_0^4 tr Q_{[\Delta]}}{[n - r(\Delta)]^2}$$

where $\phi(i,j)$ denote the (i,j).th element of $Q_{[\Delta]}$. Since $Q_{[\Delta]}$ is the orthogonal complement of $P_{[\Delta]}$, $r\left(Q_{[\Delta]}\right) = n - r(\Delta)$, then c) and d) of Lemma 3 implies $tr\left(Q_{[\Delta]}\right) = n - r(\Delta)$, so that

$$Var\left(\frac{u_0' Q_{[\Delta]} u_0}{n - r(\Delta)}\right) = \left[\frac{1}{n - r(\Delta)}\right]^2 \kappa_{0,4} \sum_{i=1}^{n} \phi^2(i,i) + \frac{2\sigma_0^4}{n - r(\Delta)}.$$

From a) of Lemma 3 and $tr\left(Q_{[\Delta]}\right) \equiv \sum_{i=1}^{n} \phi(i,i) = n - r(\Delta)$ it follows that $\sum_{i=1}^{n} \phi^2(i,i) \leq n - r(\Delta)$. In addition, from b) of A.2 it follows that n and $n - r(\Delta)$ are of the same order of magnitude, so that $n - r(\Delta) \to \infty$ as $n \to \infty$. Hence, given that σ_0^4 and $\kappa_{0,4}$ are finite, both terms in the foregoing equation go to 0 as $n \to \infty$ and then $Var\left(\frac{u_0' Q_{[\Delta]} u_0}{n - r(\Delta)}\right) \to 0$ as well.

Part b) is proved at once, upon obtaining from Lemma 1

$$E\left(u_0' P_{[\Delta_i]} u_0\right) = tr\, P_{[\Delta_i]} \sigma_0^2$$

and

$$Var\left(\frac{u_0' P_{[\Delta_i]} u_0}{n}\right) = \frac{1}{n^2} \kappa_4 \sum_{j=1}^{n} \pi_i^2(j,j) + \frac{2\sigma_0^4 tr\, P_{[\Delta_i]}}{n^2},$$

where $\pi_i(j,j)$ denotes the $j.th$ element of the main diagonal of $P_{[\Delta_i]}$, and noticing that $tr\, P_{[\Delta_i]} = r(\Delta_i) = N_i$ and $\sum_{j=1}^{n} \pi_i^2(j,j) \leq tr\, P_{[\Delta_i]} \leq n$ by Lemma 3 and a) of A.2.

c) By Lemma 2, $E\left(u_0' P_{[\Delta_i]}\Delta_j u_j\right) = 0$ and

$$
\begin{aligned}
Var\left(\frac{u_0' P_{[\Delta_i]}\Delta_j u_j}{n}\right) &= \frac{\sigma_0^2\sigma_j^2}{n^2} tr\ \Delta_j' P_{[\Delta_i]}\Delta_j . \\
&= \frac{\sigma_0^2\sigma_j^2}{n^2}\sum_{s=1}^{N_i}\sum_{t=1}^{N_j}\frac{g_{ij}^2(s,t)}{g_i(s)} \qquad (35)
\end{aligned}
$$

where the last equality follows from b) of Lemma 4. By equation (33) of Lemma 4

$$
\sum_{t=1}^{N_j}\frac{g_{ij}^2(s,t)}{g_i(s)} \leq g_i(s), \qquad (36)
$$

$i, j = 1, ..., m$, $s = 1, ..., N_i$, with equality holding if $i = j$, since by a) of Lemma 4, $g_{ii}(s,s) = g_i(s)$ and $g_{ii}(s,t) = 0$ for all $s \neq t$. Hence, summing both sides of (36) over $s = 1, ..., N_i$ and given (32) yields

$$
\sum_{s=1}^{N_i}\sum_{t=1}^{N_j}\frac{g_{ij}^2(s,t)}{g_i(s)} \leq n,
$$

with equality holding if $i = j$. Hence, given (35),

$$
Var\left(\frac{u_0' P_{[\Delta_i]}\Delta_j u_j}{n}\right) \leq Var\left(\frac{u_0'\Delta_j u_j}{n}\right) = \frac{\sigma_0^2\sigma_j^2}{n}
$$

for all $i, j = 1, \ldots, m$, which implies that $Var\left(\frac{u_0' P_{[\Delta_i]}\Delta_j u_j}{n}\right) \to 0$ as $n \to \infty$ for all $i, j = 1, \ldots, m$,. ■

Theorem 2 Assume A.1-A.3. Then,

$$
a) \quad \frac{u_j'\Delta_j' P_{[\Delta_i]}\Delta_j u_j}{n} - \frac{tr\Delta_j' P_{[\Delta_i]}\Delta_j}{n}\sigma_j^2 \underset{p}{\to} 0
$$

as $n \to \infty$ for all $i, j = 1, \ldots, m$, and

$$
b) \quad \frac{u_i'\Delta_i' P_{[\Delta_k]}\Delta_j u_j}{n} \underset{p}{\to} 0
$$

as $n \to \infty$ for all $i \neq j = 1, \ldots, m$ and $k = 1, \ldots, m$.

Proof. a) By Lemma 1,

$$E\left(\frac{u_j'\Delta_j' P_{[\Delta_i]}\Delta_j u_j}{n}\right) = \frac{tr\Delta_j' P_{[\Delta_i]}\Delta_j}{n}\sigma_j^2$$

and

$$\begin{aligned} Var\left(\frac{u_j'\Delta_j' P_{[\Delta_i]}\Delta_j u_j}{n}\right) &= \kappa_{j,4}\frac{tr\Delta_j' P_{[\Delta_i]}\Delta_j\left(I * \Delta_j' P_{[\Delta_i]}\Delta_j\right)}{n^2} \\ &\quad + 2\sigma_j^4\frac{tr\Delta_j' P_{[\Delta_i]}\Delta_j\Delta_j' P_{[\Delta_i]}\Delta_j}{n^2}. \end{aligned}$$

By c) and d) of Lemma 4,

$$\begin{aligned} Var\left(\frac{u_j'\Delta_j' P_{[\Delta_i]}\Delta_j u_j}{n}\right) &= \frac{\kappa_{j,4}}{n^2}\sum_{t=1}^{N_j}\left[\sum_{s=1}^{N_i}\frac{g_{ij}^2(s,t)}{g_i(s)}\right]^2 \\ &\quad + \frac{2\sigma_j^4}{n^2}\sum_{t=1}^{N_j}\sum_{r=1}^{N_j}\left(\sum_{s=1}^{N_i}\frac{g_{ij}(s,t)\,g_{ij}(s,r)}{g_i(s)}\right)^2. \end{aligned} \tag{37}$$

There are two important results for the bounds of the above traces and, consequently, the limiting behaviour of $Var\left(u_i'\Delta_i' P_{[\Delta_j]}\Delta_i u_i/n\right)$ when $n \to \infty$. By equation (33), $\sum_{s=1}^{N_i} g_{ij}(s,t) = g_j(t)$, so that multiplying all terms in the left hand side of the foregoing equation by the corresponding relative size $g_{ij}(s,t)/g_i(s) \leq 1$, $s = 1, \ldots N_i$, gives

$$\sum_{s=1}^{N_i}\frac{g_{ij}^2(s,t)}{g_i(s)} \leq g_j(t)$$

$t = 1, \ldots, N_j$, with equality if $i = j$. So, taking the squares of both sides in the foregoing inequality leaves the direction of the inequality unchanged. Then, summing the squares over $t = 1, ..., N_j$ gives the first important result

$$\sum_{t=1}^{N_j} \left[\sum_{s=1}^{N_i} \frac{g_{ij}^2 (s,t)}{g_i (s)} \right]^2 \leq \sum_{t=1}^{N_j} g_j^2 (t) , \tag{38}$$

with equality if $i = j$. The second important result is

$$\begin{aligned} \sum_{t=1}^{N_j} \sum_{r=1}^{N_j} \left(\sum_{s=1}^{N_i} \frac{g_{ij} (s,t) g_{ij} (s,r)}{g_i (s)} \right)^2 &\leq \sum_{t=1}^{N_j} \left(\sum_{r=1}^{N_j} \sum_{s=1}^{N_i} \frac{g_{ij} (s,t) g_{ij} (s,r)}{g_i (s)} \right)^2 \\ &= \sum_{t=1}^{N_j} \left[\sum_{s=1}^{N_i} \left(\frac{g_{ij} (s,t)}{g_i (s)} \sum_{r=1}^{N_j} g_{ij} (s,r) \right) \right]^2 \\ &= \sum_{t=1}^{N_j} g_j^2 (t) . \end{aligned} \tag{39}$$

The inequality simply follows from the fact that $\sum_{s=1}^{N_i} g_{ij} (s,t) g_{ij} (s,r) / g_i (s)$, $r = 1, ..., N_j$, are positive numbers. It becomes equality if $i = j$, since by a) of Lemma 4,

$$g_{jj} (s,t) g_{jj} (s,r) / g_j (s) = 0$$

for all $r, s \neq t$ and

$$g_{jj}^2 (t,t) / g_j (t) = g_j (t) .$$

The last equality in (39) follows from equation (33). So, from equations (38) and (39) and given (37),

$$\begin{aligned} Var \left(\frac{u_j' \Delta_j' P_{[\Delta_i]} \Delta_j u_j}{n} \right) &\leq Var \left(\frac{u_j' \Delta_j' \Delta_j u_j}{n} \right) = \frac{\left(\kappa_{j,4} + 2\sigma_j^4 \right)}{n^2} \sum_{t=1}^{N_j} g_j^2 (t) \\ &= \left(\kappa_{j,4} + 2\sigma_j^4 \right) \frac{\overline{v}_j^2 + 1}{N_j} \end{aligned} \tag{40}$$

for all $i, j = 1, \ldots, m$, where the last equality in (40) follows from the definition of $\overline{v}_j^2$ in (9) and the fact that, given (8) and (32), $n = N_j \overline{g}_j$. Hence, given A.1

and A.3,

$$Var\left(\frac{u_j'\Delta_j'P_{[\Delta_i]}\Delta_j u_j}{n}\right)\to 0$$

as $n\to\infty$, and so the result follows.

b) Given $i\neq j=1,...,m$ and $k=1,...,m$, Lemma 2 applies, so that $E\left(u_i'\Delta_i'P_{[\Delta_k]}\Delta_j u_j/n\right)=0$ and

$$\begin{aligned}Var\left(\frac{u_i'\Delta_i'P_{[\Delta_k]}\Delta_j u_j}{n}\right) &= \frac{\sigma_i^2\sigma_j^2}{n^2}tr\,\Delta_j'P_{[\Delta_k]}\Delta_i\Delta_i'P_{[\Delta_k]}\Delta_j\\ &= \frac{\sigma_i^2\sigma_j^2}{n^2}\sum_{r=1}^{N_i}\sum_{t=1}^{N_j}\left(\sum_{s=1}^{N_k}\frac{g_{ik}\left(r,s\right)g_{kj}\left(s,t\right)}{g_k\left(s\right)}\right)^2,\end{aligned}$$

where the last equality follows from d) of Lemma 4. Since

$$\begin{aligned}\sum_{r=1}^{N_i}\sum_{t=1}^{N_j}\left(\sum_{s=1}^{N_k}\frac{g_{ik}\left(r,s\right)g_{kj}\left(s,t\right)}{g_k\left(s\right)}\right)^2 &< \sum_{r=1}^{N_i}\left(\sum_{t=1}^{N_j}\sum_{s=1}^{N_k}\frac{g_{ik}\left(r,s\right)g_{kj}\left(s,t\right)}{g_k\left(s\right)}\right)^2\\ &= \sum_{r=1}^{N_i}\left[\sum_{s=1}^{N_k}\left(\frac{g_{ik}\left(r,s\right)}{g_k\left(s\right)}\sum_{t=1}^{N_j}g_{kj}\left(s,t\right)\right)\right]^2\\ &= \sum_{r=1}^{N_i}\left(\sum_{s=1}^{N_k}g_{ik}\left(r,s\right)\right)^2=\sum_{r=1}^{N_i}g_i^2\left(r\right),\end{aligned}$$

then

$$Var\left(\frac{u_i'\Delta_i'P_{[\Delta_k]}\Delta_j u_j}{n}\right)<\sigma_i^2\sigma_j^2\frac{\overline{v}_i^2+1}{N_i}.$$

Along the same lines, it can be proved that also

$$Var\left(\frac{u_i'\Delta_i'P_{[\Delta_k]}\Delta_j u_j}{n}\right)<\sigma_i^2\sigma_j^2\frac{\overline{v}_j^2+1}{N_j}.$$

Hence,

$$Var\left(\frac{u_i'\Delta_i'P_{[\Delta_k]}\Delta_j u_j}{n}\right)<\sigma_i^2\sigma_j^2\min\left(\frac{\overline{v}_i^2+1}{N_i},\frac{\overline{v}_j^2+1}{N_j}\right). \qquad (41)$$

■

Theorem 3 Let A1-A4 hold. Then, $\widetilde{\sigma}^2$ exists and is unbiased for n sufficiently large, and

$$\widetilde{\sigma}^2 \underset{p}{\rightarrow} \sigma^2$$

as $n \rightarrow \infty$.

Proof. By A.4, A^{-1} exists for n sufficiently large, which assures that $\widetilde{\sigma}^2$ exists for n sufficiently large. In light of equation (5) and since $\widetilde{\sigma}_0^2$ is unbiased, $\widetilde{\sigma}^2$ is unbiased either. By Lemma 3 and given a) of A.2, C is $O(1)$, hence a) of Theorem 1 and a standard asymptotic result (e.g. Proposition 2.30 in White (2001)) assures that

$$\widetilde{\sigma}_0^2 C - \sigma_0^2 C \underset{p}{\rightarrow} 0 \tag{42}$$

as $n \rightarrow \infty$. By Theorems 1 and 2, $B - \sigma_0^2 C - A\sigma^2 \underset{p}{\rightarrow} 0$ as $n \rightarrow \infty$ and so, given (42),

$$B - \widetilde{\sigma}_0^2 C - A\sigma^2 \underset{p}{\rightarrow} 0 \tag{43}$$

as $n \rightarrow \infty$.

Given equations (32) and (33),

$$n = \sum_{s=1}^{N_i} \sum_{t=1}^{N_j} g_{ij}(s,t), \; i,j = 1,...,m$$

Hence, from the trace formula in b) of Lemma 4, it is clear that

$$\frac{tr\Delta_j' P_{[\Delta_i]} \Delta_j}{n} \leq 1, \tag{44}$$

with equality if $i = j$, implying that A is $O(1)$. By A.4, $1/\det(A)$ is $O(1)$ and since A is of finite dimension with all elements $O(1)$, it follows that A^{-1} is $O(1)$

as well. Hence, given (43),

$$A^{-1}\left(B-\widetilde{\sigma}_0^2 C\right) \underset{p}{\rightarrow} \sigma^2.$$

as $n \rightarrow \infty$. ■

Theorem 4 Assume A.1, A.2 and A.4-A.6.WTSLS. Then, $\widehat{\sigma}_{w,0}^2$ and $\widehat{\sigma}_w^2$ exist with probability approaching 1 (w.p.a. 1); and

$$\widehat{\sigma}_{w,0}^2 \underset{p}{\rightarrow} \sigma_0^2$$

as $n \rightarrow \infty$; assuming also A.3,

$$\widehat{\sigma}_w^2 \underset{p}{\rightarrow} \sigma^2$$

as $n \rightarrow \infty$.

Proof. The WTSLS residuals $\widehat{\epsilon}_w$ can be equivalently written as

$$\begin{aligned}\widehat{\epsilon}_w &= \left\{ I - X\left[\frac{X'Q_{[\Delta]}Z}{n}\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1}\frac{Z'Q_{[\Delta]}X}{n}\right]^{-1}\right. \\ &\quad \left.\frac{X'Q_{[\Delta]}Z}{n}\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1}\frac{Z'Q_{[\Delta]}}{n}\right\}\epsilon.\end{aligned} \tag{45}$$

Given b) of A.2 and A.4, ACE1 exists w.p.a. 1 if and only if the inverse of the matrix in brackets in equation (45) exists w.p.a. 1.

Given A.6.WTSLS, since the inverse matrix function is continuous and Q_{ZZ} is $O(1)$ and uniformly positive definite, proposition 2.30 in White (2001) assures that

$$\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1} - Q_{ZZ}^{-1} \underset{p}{\rightarrow} 0$$

as $n \rightarrow \infty$. Since Q_{ZZ} is uniformly positive definite, $1/\det(Q_{ZZ})$ is $O(1)$ and hence, given that Q_{ZZ} is $O(1)$, Q_{ZZ}^{-1} is $O(1)$ as well. By continuity of the matrix product and the fact that both Q_{ZX} and Q_{ZZ}^{-1} are $O(1)$, then

$$\frac{X'Q_{[\Delta]}Z}{n}\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1}\frac{Z'Q_{[\Delta]}X}{n} - Q'_{ZX}Q_{ZZ}^{-1}Q_{ZX} \underset{p}{\rightarrow} 0 \tag{46}$$

as $n \rightarrow \infty$. Let $D_Q \equiv det\left(Q'_{ZX}Q_{ZZ}^{-1}Q_{ZX}\right)$. Given A.6.WTSLS, Lemma 2.19 in White (2001) assures that the sequence $\left\{Q'_{ZX}Q_{ZZ}^{-1}Q_{ZX}\right\}$ is $O(1)$ and uniformly positive definite. Hence, given that the determinant is a continuous function and $Q'_{ZX}Q_{ZZ}^{-1}Q_{ZX}$ is $O(1)$, the probability limit in (46) and proposition 2.30 in White (2001) assure that

$$det\left[\frac{X'Q_{[\Delta]}Z}{n}\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1}\frac{Z'Q_{[\Delta]}X}{n}\right] - D_Q \underset{p}{\rightarrow} 0 \tag{47}$$

as $n \rightarrow \infty$. Since the sequence $\left\{Q'_{ZX}Q_{ZZ}^{-1}Q_{ZX}\right\}$ is uniformly positive definite, $D_Q > \delta > 0$ for n sufficiently large, so that

$$\begin{aligned} prob\left\{det\left[\frac{X'Q_{[\Delta]}Z}{n}\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1}\frac{Z'Q_{[\Delta]}X}{n}\right] > \frac{\delta}{2} > 0\right\} &\geq \\ prob\left\{det\left[\frac{X'Q_{[\Delta]}Z}{n}\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1}\frac{Z'Q_{[\Delta]}X}{n}\right] > D_Q - \frac{\delta}{2}\right\} &\geq \\ prob\left\{D_Q - \frac{\delta}{2} < det\left[\frac{X'Q_{[\Delta]}Z}{n}\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1}\frac{Z'Q_{[\Delta]}X}{n}\right] < D_Q + \frac{\delta}{2}\right\} & \end{aligned}$$

for n sufficiently large. Hence by (47),

$$\lim_{n\rightarrow\infty} prob\left\{det\left[\frac{X'Q_{[\Delta]}Z}{n}\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1}\frac{Z'Q_{[\Delta]}X}{n}\right] > \frac{\delta}{2} > 0\right\} = 1$$

so that the inverse of

$$\frac{X'Q_{[\Delta]}Z}{n}\left(\frac{Z'Q_{[\Delta]}Z}{n}\right)^{-1}\frac{Z'Q_{[\Delta]}X}{n}$$

exists w.p.a. 1 as $n \to \infty$, and so does ACE1.

To prove consistency of $\widehat{\sigma}_{w,0}^2$ and $\widehat{\sigma}_w^2$, we expand the right hand sides of the following equations

$$\widehat{\epsilon}_w' Q_{[\Delta]} \widehat{\epsilon}_w = \epsilon' M_w' Q_{[\Delta]} M_w \epsilon$$

and

$$\widehat{\epsilon}_w' P_{[\Delta_i]} \widehat{\epsilon}_w = \epsilon' M_w' P_{[\Delta_i]} M_w \epsilon$$

for $i = 1, ..., m$ to have

$$\begin{aligned}
& \epsilon' M_w' Q_{[\Delta]} M_w \epsilon \\
= \;& \epsilon' \Big\{ Q_{[\Delta]} - Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \\
& \left[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \right]^{-1} X' Q_{[\Delta]} \Big\} \\
& \Big\{ I - X \left[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \right]^{-1} \\
& X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} \Big\} \epsilon \\
= \;& \epsilon' Q_{[\Delta]} \epsilon + \Gamma_{w,0},
\end{aligned}$$

where

$$\begin{aligned}
\Gamma_{w,0} \equiv \;& -2 \epsilon' Q_{[\Delta]} X \left[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \right]^{-1} \\
& X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} \epsilon + \epsilon' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \\
& \left[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \right]^{-1} X' Q_{[\Delta]} X \\
& \left[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \right]^{-1} X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} \epsilon
\end{aligned}$$

and

$$\begin{aligned}
& \epsilon' M_w' P_{[\Delta_i]} M_w \epsilon \\
= \;& \epsilon' \Big\{ P_{[\Delta_i]} - Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \\
& \Big[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \Big]^{-1} X' P_{[\Delta_i]} \Big\} \\
& \Big\{ I - X \Big[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \Big]^{-1} \\
& X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} \Big\} \epsilon \\
= \;& \epsilon' P_{[\Delta_i]} \epsilon + \Gamma_{w,i},
\end{aligned}$$

where

$$\begin{aligned}
\Gamma_{w,i} \;\equiv\; & -2\epsilon' P_{[\Delta_i]} X \Big[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \Big]^{-1} \\
& X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} \epsilon + \epsilon' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \\
& \Big[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \Big]^{-1} X' P_{[\Delta_i]} X \\
& \Big[X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} X \Big]^{-1} X' Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z \right)^{-1} Z' Q_{[\Delta]} \epsilon
\end{aligned}$$

for $i = 1, \ldots, m$.

Therefore

$$\widehat{\sigma}_{w,0}^2 = \widetilde{\sigma}_0^2 + \frac{n}{n - r(\Delta)} \frac{\Gamma_{w,0}}{n}. \tag{48}$$

As proved above, the sequence $\left\{ Q_{ZX}' Q_{ZZ}^{-1} Q_{ZX} \right\}$ is $O(1)$ and uniformly positive definite. Hence, given A.5, A.6.WTSLS and the probability limit in (46) proposition 2.30 in White (2001) applies to yield

$$\begin{aligned}
\frac{\Gamma_{w,0}}{n} \;&-\; 2Q_{X\epsilon}' \left(Q_{ZX}' Q_{ZZ}^{-1} Q_{ZX} \right)^{-1} Q_{ZX}' Q_{ZZ}^{-1} \cdot 0 \\
&+\; 0 \cdot Q_{ZZ}^{-1} Q_{ZX} \left(Q_{ZX}' Q_{ZZ}^{-1} Q_{ZX} \right)^{-1} Q_{XX} \\
& \left(Q_{ZX}' Q_{ZZ}^{-1} Q_{ZX} \right)^{-1} Q_{ZX}' Q_{ZZ}^{-1} \cdot 0 \underset{p}{\rightarrow} 0,
\end{aligned}$$

as $n \to \infty$, and since $n/\left[n - r\left(\Delta\right)\right]$ is $O\left(1\right)$,

$$\frac{n}{n - r\left(\Delta\right)} \frac{\Gamma_{w,0}}{n} \underset{p}{\to} 0 \tag{49}$$

as $n \to \infty$. Then, by Theorem 1, $\widetilde{\sigma}_0^2 \underset{p}{\to} \sigma_0^2$ as $n \to \infty$. Hence, given (48) and (49)

$$\widehat{\sigma}_{w,0}^2 \underset{p}{\to} \sigma_0^2. \tag{50}$$

By the same token, it is proved that

$$\frac{\Gamma_w}{n} \underset{p}{\to} 0 \tag{51}$$

as $n \to \infty$, where $\Gamma_w \equiv \left(\Gamma_{w,1} \ldots \Gamma_{w,m}\right)'$. Given $\hat{B}_w = B + \Gamma_w/n$ and (48), it has

$$\begin{aligned} \widehat{\sigma}_w^2 &= A^{-1}\left[B + \frac{\Gamma_w}{n} - \left(\widetilde{\sigma}_0^2 + \frac{n}{n - r\left(\Delta\right)} \frac{\Gamma_{w,0}}{n}\right) C\right], \\ &= \widetilde{\sigma}^2 + A^{-1}\left(\frac{\Gamma_w}{n} - \frac{n}{n - r\left(\Delta\right)} \frac{\Gamma_{w,0}}{n} C\right). \end{aligned}$$

Finally, given A.3, it is legitimate to apply Theorem 3, which, along with the probability limits (49) and (51) and the fact that A^{-1} and C are $O\left(1\right)$, yields the result. ■

Theorem 5 Assume A.1, A.2 and A.4-A.6.TSLS. Then, $\widehat{\sigma}_{ls,0}^2$ and $\widehat{\sigma}_{ls}^2$ exist w.p.a. 1; and

$$\widehat{\sigma}_{ls,0}^2 \underset{p}{\to} \sigma_0^2$$

as $n \to \infty$; assuming also A.3,

$$\widehat{\sigma}_{ls}^2 \underset{p}{\to} \sigma^2$$

as $n \to \infty$.

Proof. The proof follows exactly the same steps as Theorem **4**'s. Here, I only report the specific expressions for $\widehat{\epsilon}_{ls}' Q_{[\Delta]} \widehat{\epsilon}_{ls}$ and $\widehat{\epsilon}_{ls}' P_{[\Delta_i]} \widehat{\epsilon}_{ls}$.

$$\widehat{\epsilon}' Q_{[\Delta]} \widehat{\epsilon} \quad = \quad \epsilon' Q_{[\Delta]} \epsilon + \Gamma_{ls,0},$$

where

$$\begin{aligned} \Gamma_{ls,0} \quad \equiv \quad & -2\epsilon' Q_{[\Delta]} X \left[X'Z \left(Z'Z\right)^{-1} Z'X \right]^{-1} X'Z \left(Z'Z\right)^{-1} Z'\epsilon \\ & + \epsilon' Z \left(Z'Z\right)^{-1} Z'X \left[X'Z \left(Z'Z\right)^{-1} Z'X \right]^{-1} X' Q_{[\Delta]} X \\ & \left[X'Z \left(Z'Z\right)^{-1} Z'X \right]^{-1} X'Z \left(Z'Z\right)^{-1} Z'\epsilon \end{aligned}$$

and

$$\widehat{\epsilon}_{ls}' P_{[\Delta_i]} \widehat{\epsilon}_{ls} \quad = \quad \epsilon' P_{[\Delta_i]} \epsilon + \Gamma_{ls,i},$$

where

$$\begin{aligned} \Gamma_{ls,i} \quad \equiv \quad & -2\epsilon' P_{[\Delta_i]} X \left[X'Z \left(Z'Z\right)^{-1} Z'X \right]^{-1} X'Z \left(Z'Z\right)^{-1} Z'\epsilon \\ & + \epsilon' Z \left(Z'Z\right)^{-1} Z'X \left[X'Z \left(Z'Z\right)^{-1} Z'X \right]^{-1} X' P_{[\Delta_i]} X \\ & \left[X'Z \left(Z'Z\right)^{-1} Z'X \right]^{-1} X'Z \left(Z'Z\right)^{-1} Z'\epsilon, \end{aligned}$$

$i = 1, \ldots, m$. Then, the proof proceeds as in Theorem 4's, based on Theorems 1 and 3, A5 and A.6.TSLS. ∎

Theorem 6 Assume A.1, A.2, A.4-A.6.WTSLS. and A.6.BTSLS. Then, $\widehat{\sigma}_b^2$

exists w.p.a. 1; assuming also A.3

$$\widehat{\sigma}_b^2 \underset{p}{\to} \sigma^2$$

as $n \to \infty$.

Proof. The proof follows exactly the same steps as in Theorem **4**'s. Here I only report the specific expression for $\widehat{\epsilon}_b' P_{[\Delta_i]} \widehat{\epsilon}_b$:

$$\widehat{\epsilon}_{b,i}' P_{[\Delta_i]} \widehat{\epsilon}_{b,i} = \epsilon' P_{[\Delta_i]} \epsilon + \Gamma_{b,i}$$

where

$$\begin{aligned}
\Gamma_{b,i} \equiv\ & -2\epsilon' P_{[\Delta_i]} X \left[X' P_{[\Delta_i]} Z \left(Z' P_{[\Delta_i]} Z\right)^{-1} Z' P_{[\Delta_i]} X\right]^{-1} \\
& X' P_{[\Delta_i]} Z \left(Z' P_{[\Delta_i]} Z\right)^{-1} Z' P_{[\Delta_i]} \epsilon + \epsilon' P_{[\Delta_i]} Z \left(Z' P_{[\Delta_i]} Z\right)^{-1} Z' P_{[\Delta_i]} X \\
& \left[X' P_{[\Delta_i]} Z \left(Z' P_{[\Delta_i]} Z\right)^{-1} Z' P_{[\Delta_i]} X\right]^{-1} X' P_{[\Delta_i]} X \\
& \left[X' P_{[\Delta_i]} Z \left(Z' P_{[\Delta_i]} Z\right)^{-1} Z' P_{[\Delta_i]} X\right]^{-1} X' P_{[\Delta_i]} Z \left(Z' P_{[\Delta_i]} Z\right)^{-1} Z' P_{[\Delta_i]} \epsilon.
\end{aligned}$$

$i = 1, \ldots, m$. Then, the proof proceeds exactly as in Theorem 4's, based on Theorem 3, A5 and A.6.BTSLS. ■

C Results on unbiased estimators

C.1 Derivation of equations (22)

By assumption, $E(\epsilon\epsilon' | X) = \Sigma$. Hence,

$$\begin{aligned}
E(q_{w,0} | X) &= E\left(tr\ \epsilon' M_w' Q_{[\Delta]} M_w \epsilon\right) \\
&= tr\ Q_{[\Delta]} M_w \Sigma M_w'
\end{aligned}$$

and

$$\begin{aligned} E\left(q_{w,i}|X\right) &= E\left(tr\ \epsilon' M_w' P_{[\Delta_i]} M_w \epsilon\right) \\ &= tr\ P_{[\Delta_i]} M_w \Sigma M_w', \end{aligned}$$

where

$$M_w = I - X\left[X' Q_{[\Delta]} X\right]^{-1} X' Q_{[\Delta]}.$$

Elaborating $M_w \Sigma M_w'$ gives

$$\begin{aligned} M_w \Sigma M_w' &= \left[I - X\left(X' Q_{[\Delta]} X\right)^{-1} X' Q_{[\Delta]}\right] \\ &\quad \left(\sigma_0^2 I_n + \sigma_1^2 \Delta_1 \Delta_1' + \ldots + \sigma_m^2 \Delta_m \Delta_m'\right) \\ &\quad \left[I - Q_{[\Delta]} X\left(X' Q_{[\Delta]} X\right)^{-1} X'\right] \\ &= \left(\Sigma - \sigma_0^2 X\left(X' Q_{[\Delta]} X\right)^{-1} X' Q_{[\Delta]}\right) \\ &\quad \left[I - Q_{[\Delta]} X\left(X' Q_{[\Delta]} X\right)^{-1} X'\right] \\ &= \Sigma - \sigma_0^2 X\left(X' Q_{[\Delta]} X\right)^{-1} X' Q_{[\Delta]} - \\ &\quad \sigma_0^2 Q_{[\Delta]} X\left(X' Q_{[\Delta]} X\right)^{-1} X' + \sigma_0^2 X\left(X' Q_{[\Delta]} X\right)^{-1} X', \end{aligned} \tag{52}$$

so that

$$Q_{[\Delta]} M_w \Sigma M_w' = \sigma_0^2\left[Q_{[\Delta]} - Q_{[\Delta]} X\left(X' Q_{[\Delta]} X\right)^{-1} X' Q_{[\Delta]}\right],$$

since $Q_{[\Delta]}\Sigma = \sigma_0^2 Q_{[\Delta]}$, and

$$\begin{aligned} P_{[\Delta_i]} M_w \Sigma M_w' &= P_{[\Delta_i]}\Sigma - \sigma_0^2 P_{[\Delta_i]} X\left(X' Q_{[\Delta]} X\right)^{-1} X' Q_{[\Delta]} \\ &\quad + \sigma_0^2 P_{[\Delta_i]} X\left(X' Q_{[\Delta]} X\right)^{-1} X'. \end{aligned}$$

Hence,

$$tr\, Q_{[\Delta]} M_w \Sigma M_w' = \sigma_0^2 \left(n - r\left(\Delta\right) - k\right)$$

proving the first equation of (22), and

$$tr\ P_{[\Delta_i]} M_w \Sigma M_w' = tr\ P_{[\Delta_i]} \Sigma + \sigma_0^2 tr\ \left(X' Q_{[\Delta]} X\right)^{-1} X' P_{[\Delta_i]} X,$$

proving the block of m equations in (22), given that

$$P_{[\Delta_i]} \Sigma = \sigma_0^2 P_{[\Delta_i]} + \sigma_1^2 P_{[\Delta_i]} \Delta_1 \Delta_1' + ... + \sigma_i^2 \Delta_i \Delta_i' + ... + \sigma_m^2 P_{[\Delta_i]} \Delta_m \Delta_m',$$

and so

$$\begin{aligned} tr\ P_{[\Delta_i]} \Sigma &= \sigma_0^2 r\left(\Delta_i\right) + \sigma_1^2 tr \Delta_1' P_{[\Delta_i]} \Delta_1 + ... \\ &\quad + \sigma_i^2 n + ... + \sigma_m^2 tr \Delta_m' P_{[\Delta_i]} \Delta_m. \end{aligned} \tag{53}$$

C.1.1 The missing term in Davis (2002)

The formula of the extended WK estimator of σ_0^2 reported in Davis' Lemma 3 turns out to miss a term in the degrees-of-freedom correction. While Davis does not make explicit the regularity conditions underlying Lemma 3, the derivations in the proof of the Lemma are legitimate under $E\left(\epsilon\epsilon'|X,Z\right) = \Sigma$ and indeed an expression analogous to that for $M_w \Sigma M_w'$ in equation (52), say $M_w^* \Sigma M_w^{*\prime}$, is therein correctly elaborated as follows (in my notation). Let

$$P_{\left[Q_{[\Delta]} Z\right]} = Q_{[\Delta]} Z \left(Z' Q_{[\Delta]} Z\right)^{-1} Z' Q_{[\Delta]},$$

then

$$\begin{aligned} M_w^*\Sigma M_w^{*\prime} &= \Sigma - \sigma_0^2 X\left(X'P_{\left[Q_{[\Delta]}Z\right]}X\right)^{-1} X'P_{\left[Q_{[\Delta]}Z\right]} \\ &\quad -\sigma_0^2 P_{\left[Q_{[\Delta]}Z\right]}X\left(X'P_{\left[Q_{[\Delta]}Z\right]}X\right)^{-1} X' \\ &\quad +\sigma_0^2 X\left(X'P_{\left[Q_{[\Delta]}Z\right]}X\right)^{-1} X'. \end{aligned} \tag{54}$$

Incidentally, the expression for $M_w^*\Sigma M_w^{*\prime}$ reduces to that of $M_w\Sigma M_w'$ in (52) when $P_{\left[Q_{[\Delta]}Z\right]}$ is replaced by $Q_{[\Delta]}$, as appropriate when X lies onto $R(Z)$. What is wrong is the expression for the trace of $Q_{[\Delta]}M_w^*\Sigma M_w^{*\prime}$ as reported in Davis' proof:

$$tr\, Q_{[\Delta]}M_w^*\Sigma M_w^{*\prime} = \sigma_0^2\left[tr\ Q_{[\Delta]} + tr\ \left(X'P_{\left[Q_{[\Delta]}Z\right]}X\right)^{-1} X'Q_{[\Delta]}X\right]. \tag{55}$$

In order to obtain the correct expression, I elaborate $Q_{[\Delta]}M_w^*\Sigma M_w^{*\prime}$ as follows

$$\begin{aligned} Q_{[\Delta]}M_w^*\Sigma M_w^{*\prime} &= \sigma_0^2 Q_{[\Delta]} - \sigma_0^2 Q_{[\Delta]} X\left(X'P_{\left[Q_{[\Delta]}Z\right]}X\right)^{-1} X'P_{\left[Q_{[\Delta]}Z\right]} \\ &\quad -\sigma_0^2 P_{\left[Q_{[\Delta]}Z\right]}X\left(X'P_{\left[Q_{[\Delta]}Z\right]}X\right)^{-1} X' \\ &\quad +\sigma_0^2 Q_{[\Delta]} X\left(X'P_{\left[Q_{[\Delta]}Z\right]}X\right)^{-1} X'. \end{aligned}$$

Hence, the correct expression for the trace of $Q_{[\Delta]}M_w^*\Sigma M_w^{*\prime}$ is

$$tr\ Q_{[\Delta]}M_w^*\Sigma M_w^{*\prime} = \sigma_0^2\left[n - r(\Delta) - 2k + tr\ \left(X'P_{\left[Q_{[\Delta]}Z\right]}X\right)^{-1} X'Q_{[\Delta]}X\right],$$

which demonstrates that Davis' formula (55) misses the term $-2k$ inside brackets. Accordingly, the correct estimator for σ_0^2 is

$$\tilde{\sigma}_{w,0}^{*2} = \frac{\epsilon' M_w^{*\prime} Q_{[\Delta]} M_w^* \epsilon}{n - r(\Delta) - k + \kappa}, \tag{56}$$

where

$$\kappa \equiv tr \, \left(X' P_{\left[Q_{[\Delta]} Z \right]} X \right)^{-1} X' Q_{[\Delta]} X - k$$

is the adjustment in the degrees-of-freedom correction due to the presence of endogenous regressors.

Notice that the expression for κ can be rearranged as

$$\kappa = tr \, \left(X' P_{\left[Q_{[\Delta]} Z \right]} X \right)^{-1} X' \left(Q_{[\Delta]} - P_{\left[Q_{[\Delta]} Z \right]} \right) X$$

and since $Q_{[\Delta]} - P_{\left[Q_{[\Delta]} Z \right]} = Q_{[\Delta, Z]}$ is an idempotent and symmetric matrix, then

$$\kappa = tr \, Q_{[\Delta, Z]} X \left(X' P_{\left[Q_{[\Delta]} Z \right]} X \right)^{-1} X' Q_{[\Delta, Z]} \geq 0.$$

Notice also how the rectified estimator (56) now easily compares to the corresponding WK estimator (24) where no adjustment is needed. Interestingly, it turns out that a degrees-of-freedom correction that did not include κ would overestimate σ_0^2. While crucial for unbiasedness, the typo is clearly immaterial as far as consistency is concerned and indeed it is not hard to verify that Davis' estimator is consistent under the same regularity conditions valid for ACE1.

C.2 Derivation of equations (28)

Let

$$\begin{aligned} q_{b,i} &= \widehat{\epsilon}'_{b,i} P_{[\Delta_i]} \widehat{\epsilon}_{b,i} \\ &= \epsilon' M'_{b,i} P_{[\Delta_i]} M_{b,i} \epsilon, \end{aligned}$$

where

$$M_{b,i} = I - X \left[X' P_{[\Delta_i]} X \right]^{-1} X' P_{[\Delta_i]},$$

$i = 1, \ldots, m$. Taking expectations, under $E\left(\epsilon\epsilon'|X\right) = \Sigma$, yields

$$E\left(q_{b,i}|X\right) \quad = \quad tr\, M_{b,i}' P_{[\Delta_i]} M_{b,i} \Sigma. \tag{57}$$

Since $P_{[\Delta_i]} M_{b,i}$ is idempotent and symmetric,

$$M_{b,i}' P_{[\Delta_i]} M_{b,i} = P_{[\Delta_i]} - P_{[\Delta_i]} X \left(X' P_{[\Delta_i]} X\right)^{-1} X' P_{[\Delta_i]}.$$

So,

$$M_{b,i}' P_{[\Delta_i]} M_{b,i} \Sigma = P_{[\Delta_i]} \Sigma - P_{[\Delta_i]} X \left(X' P_{[\Delta_i]} X\right)^{-1} X' P_{[\Delta_i]} \Sigma$$

and

$$\begin{aligned} tr\, M_{b,i}' P_{[\Delta_i]} M_{b,i} \Sigma &= tr\ P_{[\Delta_i]} \Sigma - \\ &\quad tr\ P_{[\Delta_i]} X \left(X' P_{[\Delta_i]} X\right)^{-1} X' P_{[\Delta_i]} \Sigma \\ &= tr\ P_{[\Delta_i]} \Sigma - \\ &\quad tr\ \left(X' P_{[\Delta_i]} X\right)^{-1} X' P_{[\Delta_i]} \Sigma P_{[\Delta_i]} X. \end{aligned} \tag{58}$$

Expanding $P_{[\Delta_i]} \Sigma$ gives

$$\begin{aligned} P_{[\Delta_i]} \Sigma &= \sigma_0^2 P_{[\Delta_i]} + \sigma_1^2 P_{[\Delta_i]} \Delta_1 \Delta_1' + \ldots \\ &\quad + \sigma_i^2 \Delta_i \Delta_i' + \ldots + \sigma_m^2 P_{[\Delta_i]} \Delta_m \Delta_m', \end{aligned}$$

and hence

$$\begin{aligned} tr\ P_{[\Delta_i]} \Sigma &= N_i \sigma_0^2 + \sigma_1^2 tr\ \Delta_1' P_{[\Delta_i]} \Delta_1 + \ldots \\ &\quad + \sigma_i^2 n + \ldots + \sigma_m^2 tr\ \Delta_m' P_{[\Delta_i]} \Delta_m. \end{aligned} \tag{59}$$

Furthermore,

$$P_{[\Delta_i]}\Sigma P_{[\Delta_i]} = \sigma_0^2 P_{[\Delta_i]} + \sum_{j\neq i}^{m} \sigma_j^2 P_{[\Delta_i]}\Delta_j\Delta_j' P_{[\Delta_i]} + \sigma_i^2\Delta_i\Delta_i'. \tag{60}$$

Substituting the right hand sides of equations (59) and (60) into equation (58) yields

$$\begin{aligned} & tr\ M_{b,i}' P_{[\Delta_i]} M_{b,i}\Sigma \\ = \quad & N_i\sigma_0^2 + \sum_{j\neq i}^{m}\sigma_j^2 tr\ \Delta_j' P_{[\Delta_i]}\Delta_j + n\sigma_i^2 \\ & -tr\ \left(X'P_{[\Delta_i]}X\right)^{-1} X' \left(\sigma_0^2 P_{[\Delta_i]} + \sum_{j\neq i}^{m}\sigma_j^2 P_{[\Delta_i]}\Delta_j\Delta_j' P_{[\Delta_i]} + \sigma_i^2\Delta_i\Delta_i'\right) X. \end{aligned}$$

Finally, rearranging the right hand side of the foregoing equation, and given equation (57), yields equation (28):

$$\begin{aligned} E\left(q_{b,i}|X\right) &= \left(N_i - k\right)\sigma_0^2 + \left[n - \zeta_i\right]\sigma_i^2 \\ &+ \sum_{j\neq i}^{m}\left[tr\ \Delta_j' P_{[\Delta_i]}\Delta_j - \zeta_{ij}\right]\sigma_j^2, \end{aligned}$$

where

$$\zeta_i \equiv tr\ \left(X'P_{[\Delta_i]}X\right)^{-1} X'\Delta_i\Delta_i' X$$

and

$$\zeta_{ij} \equiv tr\ \left(X'P_{[\Delta_i]}X\right)^{-1} X'P_{[\Delta_i]}\Delta_j\Delta_j' P_{[\Delta_i]}X.$$

D Condition A.3

The following proves that if all group sizes are of the same order of magnitude then $\overline{v}_i^2$ is uniformly bounded.

Proof. If all group sizes are of the same order of magnitude, then there exist constants $0 < \xi < 1 < \Xi < \infty$ and an integer n_0 such that

$$\xi < \frac{g_i(s)}{g_i(t)} < \Xi$$

for all $n > n_0$, $s, t = 1, ..., N_i$. Hence, $\xi < g_i(s)/\bar{g}_i < \Xi$ and $\xi^2 < g_i^2(s)/\bar{g}_i^2 < \Xi^2$ for all $n > n_0$, $s = 1, ..., N_i$. So,

$$\xi^2 < \frac{\frac{1}{N_i}\sum_{s=1}^{N_i} g_i^2(s)}{\bar{g}_i^2} < \Xi^2$$

for all $n > n_0$. ■

The following example proves two claims on $\overline{v}_i^2$. First, $\overline{v}_i^2$ may be uniformly bounded even if group sizes are heterogeneous in the order of magnitude. Second, A.3 does not rule out $\overline{v}_i^2$ being unbounded.

Example 1 *Consider a one-way design with $m = 1$ and group sizes heterogeneous, so that: $1 < g(i) \equiv \gamma < \infty$ for $i = 2, \ldots, N$ and $0 < \xi < N^{-\delta} g(1) < \Xi < \infty$ for all $N > \overline{N}$, with $1/2 \leq \delta < 1$ and $\overline{N}$ a given integer, i.e. the size of group 1 is of order N^{δ}, whereas the sizes of the remaining groups are fixed. Then,*

$$\overline{v}^2 = \frac{N(N-1)\gamma^2 + Ng^2(1)}{(N-1)^2\gamma^2 + 2\gamma(N-1)g(1) + g^2(1)}$$

is of order $N^{2\delta-1}$. Hence, if $1/2 < \delta < 1$, $\overline{v}^2$ increases with N, although at a slower rate. If $\delta = 1/2$, $\overline{v}^2$ is uniformly bounded. In either case A.3 is met if $N \to \infty$.