Department of Economics

Working Paper Series

Estimation of Country-Pair Data Models Controlling for Clustered Errors: with International Trade Applications

A. Colin Cameron
University of California, Davis

Natalia Golotvina
University of California, Davis

January 31, 2005

Paper # 06-13

We consider cross-section regression models for country-pair data, such as gravity models for trade volume between countries or models of exchange rate volatility, allowing for the presence of country-specific errors. This induces clustered errors in a nonstandard setting. OLS standard errors that ignore this clustering are greatly underestimated. Under the assumption of random country-specific effects we provide analytical results that permit more efficient GLS estimation even in settings where the number of unique country-pairs is very large. We include applications to international data on real exchange rates and on bilateral trade that provided the motivation for this paper. The results are more generally applicable to regression with paired data.

UCDAVIS

Estimation of Country-Pair Data Models Controlling for Clustered Errors: with International Trade Applications

A.Colin Cameron[*] and Natalia Golotvina

University of California - Davis

January 31, 2005

Abstract

We consider cross-section regression models for country-pair data, such as gravity models for trade volume between countries or models of exchange rate volatility, allowing for the presence of country-specific errors. This induces clustered errors in a nonstandard setting. OLS standard errors that ignore this clustering are greatly underestimated. Under the assumption of random country-specific effects we provide analytical results that permit more efficient GLS estimation even in settings where the number of unique country-pairs is very large. We include applications to international data on real exchange rates and on bilateral trade that provided the motivation for this paper. The results are more generally applicable to regression with paired data.

Keywords: clustered errors; random effects; country-pair data; international trade data; exchange rate data.

JEL Classification: C29, F14, F31

---

[*]Address for correspondence: Colin Cameron, Department of Economics, University of California - Davis. email: accameron@ucdavis.edu

1. Introduction

The need to control for intracluster correlation of errors in linear regression in regression errors is well known, with leading references including Kloek (1981) and Moulton (1986, 1990). Examples include clustering on region, such as state, on household and, in the case of panel data, on the individual unit of observation. For the OLS estimator, estimated standard errors computed without regard to clustering can be greatly understated. And more efficient estimators than OLS are possible.

This paper presents valid statistical inference for paired data such as that used in gravity models of international trade. A common example has dependent variable y_{ij} equal to the volume of trade between countries i and j and regressors that include the distance between countries and a measure of the relative sizes of the two countries.

Early applications of gravity models ignored the clustering in errors that arises with such paired data. Matyas (1997, 1998) pointed out the need to incorporate country-specific error components that may be either fixed or random. Cheng and Wall (2003) provide a recent summary.

Models with country-specific fixed effects can be simply estimated by OLS regression in a model that additionally includes country-specific dummies. The number of such dummies is typically small enough that direct OLS estimation is feasible. The usual OLS standard errors can be used, provided these dummies provide an adequate control for clustering.

Statistical inference for random effects models, which unlike fixed effect models permit estimation of coefficients of country-specific regressors, is more challenging. Rose and Engel (2002) provide an example. They estimate a gravity model by OLS and report standard errors that control for heteroskedasticity in the error but not error correlation. They note that these standard errors are likely to be downward biased, but do not report cluster-robust standard errors as the simple formulae for the commonest forms of clustering provided by standard statistical packages are not applicable.

General results for paired data that take into account the specific structure of the error variance-covariance matrix in the presence of random effects are given in section 2. We present both a formula for cluster corrected standard errors of the OLS estimator and a method to implement feasible GLS. In section three we generate data with random components and confirm the results of section 2. The proposed estimation techniques are applied to Rose and Engel (2002) data

on real exchange rate volatility and on trade volume in section four. Section five concludes.

2. Regression with Paired Data

We specialize to dependent variable that is symmetric with $y_{ij} = y_{ji}$, where i and j denote countries and we refer to the pairs as country-pairs, for concreteness, though the results are more generally applicable. Symmetry arises if the dependent variable is the level of bilateral trade or a price index differential. Results can be adapted to the nonsymmetric case $(y_{ij} \neq y_{ji})$ that is appropriate if export data, for example, are being modelled; see subsection 2.6. And we assume that it is not meaningful to model y_{ii}.

The random effects model being studied is a two-way random components model with an unbalanced data set with $n(n-1)/2$ unique country pairs. The two random components are assumed to be drawn from the same distribution, see (2.1) below. We present formulae for the true standard errors of the OLS estimator and additionally show how to estimate by feasible GLS without having to numerically invert the large-dimension error variance matrix.

2.1. The Model

For the i, j^{th} country-pair the regression model is

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \alpha_i + \alpha_j + \varepsilon_{ij}, \quad j = i+1, ..., n, \ i = 1, ..., n-1, \tag{2.1}$$

where α_i is a country-specific error component and ε_{ij} is an idiosyncratic error component. For models with $y_{ij} \neq y_{ji}$, see subsection 2.6, it may be appropriate to more generally specify the model $y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \alpha_i + \delta_j + \varepsilon_{ij}$, with $\delta_j \neq \alpha_j$. But for $y_{ij} = y_{ji}$, the case here, it is natural to impose the constraint $\delta_j = \alpha_j$.

Stacking over j for country i yields

$$\begin{bmatrix} y_{i,i+1} \\ \vdots \\ y_{i,n} \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_{i,i+1}\boldsymbol{\beta} \\ \vdots \\ \mathbf{x}'_{i,n}\boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \alpha_i \\ \vdots \\ \alpha_i \end{bmatrix} + \begin{bmatrix} \alpha_{i+1} \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \varepsilon_{i,i+1} \\ \vdots \\ \varepsilon_{i,n} \end{bmatrix}$$

or

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{P}_i\boldsymbol{\alpha} + \mathbf{M}_i\boldsymbol{\alpha} + \boldsymbol{\varepsilon}_i, \tag{2.2}$$

3

where $\mathbf{y}_i$ and ε_i are $(n-i) \times 1$ vectors, $\mathbf{X}_i$ is an $(n-i) \times k$ matrix, $\boldsymbol{\beta}$ is a $k \times 1$ parameter vector and $\boldsymbol{\alpha}$ is an $n \times 1$ parameter vector. The matrices $\mathbf{P}_i$ and $\mathbf{M}_i$ are $(n-i) \times n$ matrices of zeroes and ones with

$$
\begin{aligned}
\mathbf{P}_i &= \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{O}_{(n-i)\times(i-1)} & \mathbf{E}_{(n-i)\times 1} & \mathbf{O}_{(n-i)\times(n-i)} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{0}'_{n-i}\mathbf{0}_{i-1} & \mathbf{e}_{n-i} & \mathbf{0}_{n-i}\mathbf{0}'_{n-i} \end{bmatrix}
\end{aligned}
\tag{2.3}
$$

and

$$
\begin{aligned}
\mathbf{M}_i &= \begin{bmatrix} 0 & \cdots & & \cdots & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & \cdots & & \cdots & 0 \\ 1 & 0 & & \cdots & 0 \\ 0 & 1 & & \cdots & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{O}_{(n-i)\times i} & \mathbf{I}_{(n-i)\times(n-i)} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{0}_{n-i}\mathbf{0}'_i & \mathbf{I}_{n-i} \end{bmatrix},
\end{aligned}
\tag{2.4}
$$

where $\mathbf{0}_k$ is a $k \times 1$ vector of zeroes, $\mathbf{e}_k$ is a $k \times 1$ vector of ones, $\mathbf{O}_{k\times l} = \mathbf{0}_k\mathbf{0}'_l$ is a $k \times l$ matrix of zeroes, $\mathbf{E}_{k\times l} = \mathbf{e}_k\mathbf{e}'_l$ is a $k \times l$ matrix of ones and $\mathbf{O}_{k\times 0} = \emptyset$ is defined to be a null observation.

Stacking over all countries

$$
\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{n-1} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{n-1} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_{n-1} \end{bmatrix} \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_{n-1} \end{bmatrix} \boldsymbol{\alpha} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{n-1} \end{bmatrix},
$$

or

$$
\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\alpha} + \mathbf{M}\boldsymbol{\alpha} + \varepsilon \\
&= \mathbf{X}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\alpha} + \varepsilon,
\end{aligned}
\tag{2.5}
$$

where $\mathbf{y}$ and ε are $T \times 1$ vectors, $\mathbf{X}$ is a $T \times k$ matrix, $\mathbf{P}$ and $\mathbf{M}$ and $\mathbf{L} = \mathbf{P} + \mathbf{M}$ are $T \times n$ matrices and

$$
T = n(n-1)/2,
\tag{2.6}
$$

is the number of unique country pairs.

We make the standard random effects error components assumptions that

$$\varepsilon_{ij} \sim \text{iid}[0, \sigma_\varepsilon^2] \tag{2.7}$$
$$\alpha_i \sim \text{iid}[0, \sigma_\alpha^2].$$

Note that then

$$\text{Cov}[v_{ij}, v_{kl}] = \begin{cases} 2\sigma_\alpha^2 + \sigma_\varepsilon^2 & i = k, j = l \\ 0 & i \neq k \neq j \neq l, \\ \sigma_\alpha^2 & \text{otherwise.} \end{cases}$$

The model (2.5) is then

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \tag{2.8}$$

where the error term

$$\mathbf{v} = \mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where $\mathbf{L} = (\mathbf{P} + \mathbf{M})$. Given assumptions (2.7),

$$
\begin{aligned}
\boldsymbol{\Omega} &\equiv \text{E}[\mathbf{v}\mathbf{v}'] \\
&= \text{E}[(\mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\varepsilon})(\mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\varepsilon})'] \\
&= \mathbf{L}\text{E}[\boldsymbol{\alpha}\boldsymbol{\alpha}']\mathbf{L}' + \text{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] \\
&= \mathbf{L}\sigma_\alpha^2 \mathbf{I}_N \mathbf{L}' + \sigma_\varepsilon^2 \mathbf{I}_T \\
&= \sigma_\varepsilon^2 [\mathbf{I}_T + (\sigma_\varepsilon^2/\sigma_\alpha^2)\mathbf{L}\mathbf{L}'].
\end{aligned}
$$

Hence the error variance matrix

$$\boldsymbol{\Omega} = \sigma_\varepsilon^2(\mathbf{I}_T + c\mathbf{L}\mathbf{L}'), \tag{2.9}$$

where

$$c = \sigma_\alpha^2/\sigma_\varepsilon^2 \tag{2.10}$$

is the ratio of country-specific error variance to idiosyncratic error variance.

2.2. OLS Estimation

The OLS estimator is

$$\widehat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

The correct variance matrix of the OLS estimator uses the sandwich form with $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. Given $\mathbf{\Omega} = \sigma_\varepsilon^2(\mathbf{I}_T + c\mathbf{L}\mathbf{L}')$ this can be re-expressed as

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}[1 + c\mathbf{X}'\mathbf{L}\mathbf{L}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]. \tag{2.11}$$

If the clustering is erroneously ignored then we use, $\sigma_v^2(\mathbf{X}'\mathbf{X})^{-1}$, or equivalently

$$\mathrm{V}_{\mathrm{wrong}}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}[1 + 2c], \tag{2.12}$$

using $\sigma_v^2 = (\sigma_\varepsilon^2 + 2\sigma_\alpha^2) = \sigma_\varepsilon^2(1 + 2c)$ where c is defined in (2.10).

2.2.1. OLS in Intercept-only Models

Exact results are possible in the case that the only regressor is an intercept. Then $\mathbf{X} = \mathbf{e}_T$, $\mathbf{X}'\mathbf{X} = T$, $\mathbf{X}'\mathbf{L} = \mathbf{e}_T'\mathbf{L} = (n-1)\mathbf{e}_n$, see appendix 6.1, and hence $\mathbf{X}'\mathbf{L}\mathbf{L}'\mathbf{X} = (n-1)^2\mathbf{e}_n'\mathbf{e}_n = 2(n-1)T$ as $T = n(n-1)/2$. Then

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] = \sigma_\varepsilon^2 T^{-1}[1 + 2c(n-1)] \tag{2.13}$$
$$\mathrm{V}_{\mathrm{wrong}}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] = \sigma_\varepsilon^2 T^{-1}[1 + 2c],$$

where $T = n(n-1)/2$. Table 2.1 presents these two variances for different values of T and n.

The variance of the OLS estimator increases as the variance of the country-specific effect increases, for fixed σ_ε^2 and n, and is minimized when $\sigma_\alpha^2 = 0$. This efficiency loss due to the country-specific effect can be quite great. For example, when $n = 50$ and $\sigma_\alpha^2 = 0.1 \times \sigma_\varepsilon^2$ (so $c = 0.1$) the variance of the OLS is eleven times larger than if $c = 0.0$, increasing from 0.0008 to 0.0088.

The variance of the OLS estimator decreases as the number of countries increases, for fixed σ_ε^2 and c, as expected. The interesting question is whether it does so at rate T or at rate n. For $c = 0$ we obtain the usual result that it decreases at rate $T = n(n-1)/2$, since then $\mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] = \sigma_\varepsilon^2 T^{-1}$. For large c the variance decreases much more slowly, at the rate of the number of countries (here $50/10 = 5$), rather than the number of country-pairs ($1225/45 = 27.222$). Even for moderate values of c, such as $c = 0.1$, the rate of decrease is much closer to n than to T. For $c \neq 0$ and large n, $\mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] \simeq 4c\sigma_\varepsilon^2 n^{-1}$. The gains to adding more countries to a data set may be much smaller than thought a priori.

Table 2.1: OLS: Efficiency gain and standard error bias as c and n vary

$c=\sigma_\alpha^2/\sigma_\varepsilon^2$	n Countries	T=n(n-1) Country-pairs	Correct variance	Variance gain	Incorrect variance	Variance ratio τ
0	10	45	0.0222		0.0222	1.000
0	50	1225	0.0008	27.222	0.0008	1.000
0.1	10	45	0.0622		0.0267	2.333
0.1	50	1225	0.0088	7.058	0.0010	9.000
1	10	45	0.4222		0.0667	6.333
1	50	1225	0.0808	5.224	0.0024	33.000
10	10	45	4.0222		0.4667	2.936
10	50	1225	0.8008	5.023	0.0171	46.7143
∞	10	45	–		–	9.000
∞	50	1225	–	5.000	–	49.000

Note: Analytical results for intercept-only random effects model with $\sigma_\varepsilon^2=1$. Variance gain is the gain in OLS variance in going from n=10 to n=50 countries (note that 1225/45=27.222). Variance ratio is the true OLS variance divided by the incorrect OLS variance.

The multiplicative bias of the variance of the OLS estimator is

$$
\begin{aligned}
\tau &= V[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}]/V_{\mathrm{wrong}}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] \\
&= [1 + 2c(n-1)]/[1+2c] \\
&= 1 + \frac{(n-2)}{1+(1/2c)}.
\end{aligned}
$$

The bias is increasing in c and n. For large c, $\tau \simeq (n-1)$, and even for moderate c the bias is very large. For example, if $\sigma_\alpha^2 = 0.1 \times \sigma_\varepsilon^2$ (so $c = 0.1$) then $\tau = 1 + (n-2)/6 = 9$ when $n = 50$.

Now consider the traditional grouped data model with

$$
y_{jg} = \mathbf{x}'_{jg}\boldsymbol{\beta}_g + \alpha_g + \varepsilon_{jg}, \ j = 1,...,n_g, \ g = 1,...,G. \tag{2.14}
$$

Kloek (1981) and Moulton (1990) and others show that for regression on an intercept with balanced clusters of size $M = n_g = T/G$, $V[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] = \sigma^2 T^{-1}[1+\rho(M-1)]$ and $V_{\mathrm{wrong}}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] = \sigma^2 T^{-1}$ where $\sigma^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2$ and $\rho = \sigma_\alpha^2/\sigma^2$. Defining $c = \sigma_\alpha^2/\sigma_\varepsilon^2$, equivalently

$$
\begin{aligned}
V[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] &= \sigma_\varepsilon^2 T^{-1}[1+cM]. \\
V_{\mathrm{wrong}}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] &= \sigma_\varepsilon^2 T^{-1}[1+c].
\end{aligned}
$$

The result (2.13) is qualitatively similar, except that c is replaced by $2c$ since the random effect α_c appears twice in (2.1) and M is replaced by $(n-1)$ rather than the average group size $n/2$.

2.2.2. OLS in Regression Models

For the grouped data model (2.14), Scott and Holt (1982) show that if a single regressor is included then the multiplicative bias of the usual OLS standard error for the slope coefficient is $1 + \widehat{\rho}_x \rho (M-1)$, where $\widehat{\rho}_x$ can be viewed as an estimate of the intraclass correlation of the x_{jg}. The bias of $\mathrm{V}_{\mathrm{wrong}}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}]$ is therefore greater for regressors that do not vary within group, such as the intercept, than for regressors that do vary within group. If $\widehat{\rho}_x = 0$ then there is no bias.

A qualitatively similar result is expected here, since from the appendix

$$\mathbf{X}'\mathbf{L} = (n-1) \left[\begin{array}{ccc} \bar{\mathbf{x}}_1 & \cdots & \bar{\mathbf{x}}_n \end{array} \right],$$

where $\bar{\mathbf{x}}_i = n^{-1} \sum_{j=1}^{n} \mathbf{x}_{ij}$, so $\mathbf{X}'\mathbf{L}\mathbf{L}'\mathbf{X} = (n-1)^2 \sum_{i=1}^{n} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'$, while $\mathbf{X}'\mathbf{X} = \sum_{i=1}^{n-1} \mathbf{X}_i'\mathbf{X}_i = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbf{x}_{ij} \mathbf{x}_{ij}'$, so the inflation factor $\mathbf{X}'\mathbf{L}\mathbf{L}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ in (2.11) is largest when $\bar{\mathbf{x}}_i = \mathbf{x}_{ij}$ for all j.

2.3. GLS Estimation

The GLS estimator is

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}.$$

GLS estimation requires inversion of the $T \times T$ matrix $\boldsymbol{\Omega}$ which is potentially large. For example, with $n = 50$ countries $\boldsymbol{\Omega}$ is a 1225×1225 matrix. Matyas (1997) proposed GLS estimation for country-pair data but did not provide an expression for $\boldsymbol{\Omega}^{-1}$.

In general $(\mathbf{I}_T + \mathbf{A}\mathbf{A}')^{-1} = \mathbf{I}_T - \mathbf{A}(\mathbf{I}_n + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, see for example a more general result in Magnus and Neudecker (1988, p.25). Letting $\mathbf{A} = \sqrt{c}\mathbf{L}$ it follows that for $\boldsymbol{\Omega} = \sigma_\varepsilon^2(\mathbf{I}_T + c\mathbf{L}\mathbf{L}')$,

$$\boldsymbol{\Omega}^{-1} = \sigma_\varepsilon^{-2}[\mathbf{I}_T - c\mathbf{L}(\mathbf{I}_n + c\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'] , \qquad (2.15)$$

which entails inversion of a much smaller $n \times n$ matrix. The GLS estimator is therefore easily computed and has variance matrix

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}] = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}. \qquad (2.16)$$

For matrix $\mathbf{L}$ defined in subsection 2.1 even the matrix inversion within (2.15) can be avoided, as then from appendix 6.2

$$\mathbf{\Omega}^{-1} = \sigma_\varepsilon^{-2} \left[\mathbf{I}_T - \frac{c}{1+c(n-2)} \left(\mathbf{LL}' - \frac{4c}{1+2c(n-1)} \mathbf{e}_T \mathbf{e}_T' \right) \right]. \qquad (2.17)$$

For the intercept-only model $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \bar{y}$, see appendix 6.2. Otherwise the GLS estimator will usually be more efficient than OLS.

2.4. Feasible GLS Estimation

Feasible GLS estimation requires consistent estimates of σ_ε^2 and σ_α^2 (and hence $c = \sigma_\alpha^2 / \sigma_\varepsilon^2$). As usual there are many possible ways to obtain consistent estimates of the variance components. We use the following method.

First, we transform to a model that eliminates the country-specific effects and has error with variance σ_ε^2. Define the matrix $\mathbf{F}$ such that $\mathbf{F}'\mathbf{L} = \mathbf{O}_{T \times T}$ and $\mathbf{F}'\mathbf{F} = \mathbf{I}_T$. Pre-multiplying the model (2.5) by $\mathbf{F}'$ yields

$$\begin{aligned} \mathbf{F}'\mathbf{y} &= \mathbf{F}'\mathbf{X}\boldsymbol{\beta} + \mathbf{F}'\mathbf{L}\boldsymbol{\alpha} + \mathbf{F}'\boldsymbol{\varepsilon} \qquad (2.18) \\ &= \mathbf{F}'\mathbf{X}\boldsymbol{\beta} + \mathbf{F}'\boldsymbol{\varepsilon}, \end{aligned}$$

where $\mathrm{V}[\mathbf{F}'\boldsymbol{\varepsilon}] = \sigma_\varepsilon^2 \mathbf{I}$, since $\mathrm{E}[\mathbf{F}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{F}] = \mathbf{F}'\mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']\mathbf{F} = \mathbf{F}'\sigma_\varepsilon^2 \mathbf{I}\mathbf{F} = \sigma_\varepsilon^2 \mathbf{I}$ since $\mathbf{F}'\mathbf{F} = \mathbf{I}$. [Note that the transformation will eliminate the intercept term since $\mathbf{e}_T = \mathbf{L}\mathbf{e}_n/2$, see appendix 6.1, so $\mathbf{F}'\mathbf{e}_T = \mathbf{F}'\mathbf{L}\mathbf{e}_n/2 = \mathbf{O}$ as $\mathbf{F}'\mathbf{L} = \mathbf{O}$]. Since the transformed errors have variance σ_ε^2, the average of the squared residuals from OLS estimation of the transformed equation (2.18) provides a consistent estimate of σ_ε^2, so

$$\widehat{\sigma}_\varepsilon^2 = \frac{(\mathbf{F}'\mathbf{y} - \mathbf{F}'\mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{F}'\mathbf{y} - \mathbf{F}'\mathbf{X}\widehat{\boldsymbol{\beta}})}{\mathrm{rows}[F] - k}, \qquad (2.19)$$

where k is the number of regressors.

Second, note that for the OLS residual from the model (2.5) $\widehat{\mathbf{v}} = \mathbf{Q}\mathbf{v} = \mathbf{Q}(\mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\varepsilon})$ where $\mathbf{Q} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, the inner product has expected value

$$\begin{aligned} \mathrm{E}[\widehat{\mathbf{v}}'\widehat{\mathbf{v}}] &= \mathrm{E}[\mathrm{trace}[\widehat{\mathbf{v}}'\widehat{\mathbf{v}}]] \\ &= \mathrm{E}[\mathrm{trace}[\{\mathbf{Q}(\mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\varepsilon})\}'\{\mathbf{Q}(\mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\varepsilon})\}]] \\ &= \mathrm{E}[\mathrm{trace}[\{\mathbf{Q}(\mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\varepsilon})\}'\{\mathbf{Q}(\mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\varepsilon})\}]] \\ &= \mathrm{E}[\mathrm{trace}[\boldsymbol{\alpha}'\mathbf{L}\mathbf{Q}\mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}'\mathbf{Q}\mathbf{L}\boldsymbol{\alpha} + \boldsymbol{\alpha}'\mathbf{L}'\mathbf{Q}\boldsymbol{\varepsilon} + \mathbf{Q}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')]] \\ &= \mathrm{trace}[\mathbf{L}\mathbf{Q}'\mathbf{Q}\mathbf{L}]\mathrm{E}[\boldsymbol{\alpha}\boldsymbol{\alpha}'] + \mathrm{trace}\mathbf{Q}\mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] \\ &= \mathrm{trace}[\mathbf{L}\mathbf{Q}'\mathbf{Q}\mathbf{L}]\sigma_\alpha^2 + (T - k)\sigma_\varepsilon^2, \end{aligned}$$

9

using $\mathbf{Q}'\mathbf{Q} = \mathbf{Q}$ and independence of $\boldsymbol{\alpha}$ and $\boldsymbol{\varepsilon}$. This suggests estimate

$$\widehat{\sigma}_\alpha^2 = \frac{\widehat{\mathbf{v}}'\widehat{\mathbf{v}} - (T-k)\widehat{\sigma}_\varepsilon^2}{\mathrm{trace}[\mathbf{L}\mathbf{Q}'\mathbf{Q}\mathbf{L}]} \qquad (2.20)$$

The main challenge in implementation is computation of the matrix $\mathbf{F}$, the orthonormal basis for the null space of the matrix $\mathbf{L}$. In Matlab there is a direct command to do this. In Stata there is no such command. One can form the matrix $\mathbf{F}$ from the eigenvectors corresponding to zero eigenvalues of the matrix $\mathbf{L}$. However, this mechanical approach runs into size problems when n and hence T is very large. Matlab and Stata code used in this paper will be made available.

2.5. Hausman Tests for Fixed Effects

The results so far assume that the country-specific effect α_i is purely random, see (2.7). An alternative assumption is the fixed effects assumption that α_i is correlated with some of the regressors. Then consistent estimation of $\boldsymbol{\beta}$ requires controlling for α_i. In the original model (2.1) the effects $\alpha_1, ..., \alpha_n$ are viewed as parameters to estimate along with $\boldsymbol{\beta}$ and the only error is ε_{it}.

The fixed effects model is easily estimated by introducing n dummy variables $d1, ..., dn$, where $dk_{ij} = 1$ if $k = i$ or if $k = j$ and $dk_{ij} = 0$ otherwise, and by estimating by OLS the country dummy variables model

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \sum\nolimits_{k=1}^{n} \alpha_k dk_{ij} + \varepsilon_{ij}. \qquad (2.21)$$

In most applications n is sufficiently small that this is feasible without the need to instead use a differencing transformation to eliminate the α parameters. Inclusion of the dummies can control for any clustering and one can use the usual OLS standard errors provided the error ε_{ij} is iid.

A limitation of the fixed effects model is that it is no longer possible to identify the coefficients of regressors that for given country i are invariant across country j. For example, if $\mathbf{x}'_{ij}\boldsymbol{\beta} = \mathbf{z}'_{ij}\boldsymbol{\beta}_\mathbf{z} + \mathbf{w}'_i\boldsymbol{\beta}_\mathbf{w} + \mathbf{s}'_j\boldsymbol{\beta}_\mathbf{s}$ then $\boldsymbol{\beta}_\mathbf{w}$ and $\boldsymbol{\beta}_\mathbf{s}$ are not identified. They are estimable in the random effects model, but if the fixed effects assumption is appropriate then the random effects estimator (OLS or GLS) is inconsistent. This is the classical situation for a Hausman test. For the model $y_{ij} = \mathbf{z}'_{ij}\boldsymbol{\beta}_\mathbf{z} + \alpha_i + \alpha_j + \varepsilon_{ij}$, where $\mathbf{z}_{ij}$ are regressors that vary over both i and j, let $\widehat{\boldsymbol{\beta}}_\mathbf{z}$ denote the random effects GLS estimator and $\widetilde{\boldsymbol{\beta}}_\mathbf{z}$ denote the fixed effects estimator. Then

$$\mathbf{H} = (\widetilde{\boldsymbol{\beta}}_\mathbf{z} - \widehat{\boldsymbol{\beta}}_\mathbf{z})' \left[\widehat{\mathrm{V}}[\widetilde{\boldsymbol{\beta}}_\mathbf{z}] - \widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_\mathbf{z}] \right]^{-1} (\widetilde{\boldsymbol{\beta}}_\mathbf{z} - \widehat{\boldsymbol{\beta}}_\mathbf{z})$$

10

is asymptotically chi-squared distributed under the null hypothesis of random effects.

The validity of the test requires that $\widehat{\boldsymbol{\beta}}_{\mathbf{z}}$ is fully efficient under the null hypothesis, since only then does $\mathrm{Cov}[\widetilde{\boldsymbol{\beta}}_{\mathbf{z}}, \widehat{\boldsymbol{\beta}}_{\mathbf{z}}] = \mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathbf{z}}]$ so that $\mathrm{V}[\widetilde{\boldsymbol{\beta}}_{\mathbf{z}} - \widehat{\boldsymbol{\beta}}_{\mathbf{z}}] = \mathrm{V}[\widetilde{\boldsymbol{\beta}}_{\mathbf{z}}] - \mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathbf{z}}]$. This requires $\widehat{\boldsymbol{\beta}}_{\mathbf{z}}$ to be the GLS estimator. Instead many applications erroneously let $\widehat{\boldsymbol{\beta}}_{\mathbf{z}}$ be the OLS estimator, and further compound the error by using a downward-biased estimate of the variance of the OLS estimator.

The GLS estimator given in subsection 2.3 enables correct implementation of the Hausman test. This could potentially change the common empirical finding that the Hausman test rejects the random effects model.

2.6. Extensions

The preceding analysis assumes that data are available for all country pairs (aside from y_{ii}). Data on some pairs may be missing. For example, some countries may not trade with each other. The preceding results still apply with the one change that the corresponding rows of the the matrix $\mathbf{L} = (\mathbf{P} + \mathbf{M})$ need to be dropped. This matrix is then $T \times n$ where now $T < n(n-1)/2$. With this change to $\mathbf{L}$ the analytical expression for $\boldsymbol{\Omega}^{-1}$ is again that given in (2.15), though the specialization (2.17) no longer holds.

The previous results focus on the symmetric case. In the asymmetric case $y_{ij} \neq y_{ji}$ and a natural model is

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \alpha_i + \delta_j + v_{ij}, \ \ j = 1, ..., (j-1), (j+1), ..., n, \ \ i = 1, ..., n.$$

Now there are $n(n-1)$ unique observations and we no longer impose the constraint that $\alpha_i \neq \delta_j$. The preceding algebra can be adapted to this case. Alternatively note that this is a classic two-way error components model, though is unbalanced due to the absence of data for (i, i). The general results of Wansbeek and Kapteyn (1989), summarized in Baltagi (2002, p.170) can be applied. These results may simplify given the particular structure here.

For panel country-pair data y_{ij} becomes y_{ijt}. The panel case is actually simpler than the cross-section case, if one more generally includes individual country-pair effects (not identified with cross-section data alone) rather than individual country effects. Then

$$y_{ijt} = \mathbf{x}'_{ijt}\boldsymbol{\beta} + \gamma_{ij} + v_{ijt},$$

Letting p denote a unique country pair (i, j), this three-index model for y_{ijt} collapses to a two-index model $y_{pt} = \mathbf{x}'_{pt}\boldsymbol{\beta} + \gamma_p + v_{pt}$. Usual panel methods can then

11

Table 3.1: Simulation: Random Effects Model with Intercept and Single Regressor.

	True	OLS	GLS	FGLS	FE
				Estimator	
Intercept	2.0000	2.0081	2.0080	2.0080	2.0313
Correct se		(0.2834)	(0.2842)	(0.2834)	(0.2872)
Simulation se		<0.2825>	<0.2825>	<0.2825>	<2.0039>
White se		[0.0488]			
Default se		{0.0488}			
Slope	2.0000	2.0023	2.0012	2.0012	2.0012
Correct se		(0.0520)	(0.0346)	(0.0346)	(0.0346)
Simulation se		<0.0520>	<0.0354>	<0.0354>	<0.0354>
White se		[0.0581]			
Default se		{0.0582}			
σ_ε^2	1.0000	0.9997			
σ_α^2	1.0000	1.0047			
c	1.0000	1.0050			

Note: Data generated from random effects model with intercept and single regressor. There are 1000 replications. Standard errors for the slope coefficients that correct for clustering (see text) are given in parentheses; the simulation standard error is the standard deviation of the 1000 parameter estimates; White standard errors that correct for heteroskedasticity but not clustering are given in square brackets; and usual default estimates that assume iid errors are given in braces.

be applied for either fixed or random effects models; see Selenga and Shin (2004).

3. Simulation

In the following simulation we estimate a random effects model using the estimation strategy described above.

The data are generated for model (2.1) with

$$y_{ij} = 2 + 2x_{ij} + \alpha_i + \alpha_j + \varepsilon_{ij},$$

where $\alpha_i \sim \mathcal{N}[0,1]$ and $\varepsilon_{ij} \sim \mathcal{N}[0,1]$ so $c = 1.0$, and $x_{ij} \sim \mathcal{N}[0, \sigma_x^2]$ where σ_x^2 is chosen so that $R^2 \simeq 0.50$ from OLS regression. The simulations use sample size $n = 50$ and are performed 1000 times. The results are presented in Table 3.1.

We first consider bias in the OLS standard errors. For the intercept the correct OLS standard error estimate based on (2.11) is 0.2834, much larger than the incorrect estimate of 0.0488 based on (2.12) that ignores clustering, and the multiplicative bias $\sqrt{\tau} = 0.2834/.0488 = 5.81$ which essentially equals the theoretical value of $\sqrt{\tau} = \sqrt{33} = 5.74$ given in Table 2.1 for the intercept-only model. For the slope coefficient the corresponding standard errors are 0.0520 and 0.0582 and $\sqrt{\tau} = 0.89$. There is little bias in the standard error for the slope coefficient here, a result expected from the discussion in subsubsection 2.2.2 since the regressors here are generated independently. In the application below, however, regressors are correlated within country and there can be considerable bias in OLS standard errors of slope coefficients. White's heteroskedastic standard error estimates are similar to those based on (2.12) since they too do not correct for clustering and the data were generated using homoskedastic errors.

Now consider the efficiency gains to GLS and FGLS. There is no real efficiency gain in estimation of the intercept. But there is considerable efficiency gain for the slope coefficient, with standard error falling from 0.0520 for OLS to 0.0346 for FGLS. The FGLS and GLS estimates are almost identical, since the variance components are quite precisely estimated with $\hat{c} = 1.005$ very close to the dgp value of 1.0 used in GLS. For completeness we also present fixed effects estimators. These essentially coincide with the random effects estimator.

As a consistency check we note that the simulation standard errors, the standard deviation of the 1000 estimated coefficients, coincide with the average of "correct" standard errors computed using (2.11) for OLS and (2.16) for FGLS and GLS.

Similar simulations were performed for other values of c and n. For $n = 50$ the multiplicative error in the standard error for the OLS estimated intercept rises from 3.00 for $c = 0.1$ to 6.95 for $c = 10$, similar to the square root of the values in the last column of Table 2.1, while there is again no bias in slope coefficient standard errors since here the generated regressors are independent within countries. FGLS provides negligible efficiency gain for the intercept, while for the slope coefficient with $n = 50$ there is no real gain when $c = 0.1$ but a considerable gain when $c = 10$ (standard error falls from 0.0485 for OLS to 0.0131 for FGLS).

4. Applications

We re-estimate one of the models used by Rose and Engel (2002). The dependent variable y_{ij} is a measure of real exchange volatility; see Rose and Engel (2002, p.1080) for its definition and for definitions of the regressor variables. There are 98 countries and data are available for 3262 unique country pairs, less than theoretical maximum of $98 \times 97/2 = 4753$. The missing data for some country pairs are handled by the modification given in subsection 2.6.

The first column of Table 4.1 reproduces the first column of Table 5 of Rose and Engel (2002). The correct OLS standard errors use (2.11) based on the random effects model (2.1). The variance components estimates are $s_\varepsilon^2 = 0.0004$ and $s_\alpha^2 = 0.0030$, so $\widehat{c} = 8.3448$ which is quite high and indicates a large clustering effect. The default standard errors are greatly downward biased, with the correct standard errors using (2.11) for the slope coefficients being 1.8 (currency union) to 5.6 times larger (Real GDP per Capita) than those calculated using (2.12). White standard errors that correct for heteroskedasticity only are also considerably downwards-biased. Rose and Engel (2002) used White standard errors, with the caveat in their footnote 19 that "These standard errors may be biased downward because of cross-sectional dependence that is not explicitly modeled here (the British-French residual is likely to be highly related to the British-German residual). Thus we urge the reader not to take our standard errors too literally."

The second column of Table 4.1 presents GLS estimates. For the first four regressors, which vary over country-pair, there are great gains to FGLS estimation with a five-fold to ten-fold reduction in the standard errors. For the remaining two regressors there is little efficiency gain to GLS. Note that $\ln(GDP_i \times GDP_j) = \ln GDP_i + \ln GDP_j$, so these two regressors are essentially subscripted by i alone (or by j alone), with no i, j interaction. Essentially they are varying across one dimension rather than both dimensions.

The fifth column of Table 4.1 presents fixed effect model estimates. The coefficients of regressors that vary over one dimension but not both dimensions are then not identified, so Real GDP per Capita and Real GDP are dropped. The resulting fixed effects estimates, obtained by OLS regression on the other four regressors and country dummy variables, are very close to the FGLS estimates of the random effects model. This is a consequence of c being so high in this example, analogous to the panel data result that the random effects estimator goes to the fixed estimator as the variance of the individual specific error α_i becomes large relative to the variance of the idiosyncratic error ε_{it}.

14

Table 4.1: Application: Real Exchange Rate Volatility, Rose and Engel (2002, Table 5).

	Rose and Engel Model		Reduced Model		
	OLS	FGLS	OLS	FGLS	FE
Intercept	0.2793	0.2989*	0.0920	0.0732*	0.0317*
Correct se	(0.1652)	(0.1441)	(0.0493)	(0.0122)	(0.0080)
White se	[0.0345]		[0.0170]		
Default se	{0.0332}		{0.0153}		
Currency Union	-0.0532	-0.0633*	-0.0300	-0.0630*	-0.0645*
Correct se	(0.0394)	(0.0058)	(0.0402)	(0.0058)	(0.0057)
White se	[0.0045]		[0.0061]		
Default se	{0.0332}		{0.0224}		
(Log) Distance	-0.0022	0.0063*	-0.0017	0.0063*	0.0065*
Correct se	(0.0057)	(0.0005)	(0.0058)	(0.0005)	(0.0005)
White se	[0.0020]		[0.0021]		
Default se	{0.0018}		{0.0018}		
Nominal Exchange Rate Volatility	0.3041*	0.1844*	0.3153*	0.1853*	0.1754*
Correct se	(0.0260)	(0.0064)	(0.0258)	(0.0064)	(0.0066)
White se	[0.0127]		[0.0125]		
Default se	{0.0048}		{0.0048}		
(Log Product) Real GDP per Capita	-0.0126*	-0.0184*			
Correct se	(0.0062)	(0.0058)			
White se	[0.0009]				
Default se	{0.0011}				
(Log Product) Real GDP	0.0007	0.0016			
Correct se	(0.0037)	(0.0034)			
White se	[0.0062]				
Default se	{0.0007}				
s_ε^2	0.0004		0.0004		
s_α^2	0.0030		0.0031		
$\widehat{c}$	8.3448		8.5801		

Note: Cross-section data for 3262 country-pairs formed from 98 countries. Table 5 of Rose and Engel multiplies all coefficients by 100. Standard errors for the slope coefficients that correct for clustering (see text) are given in parentheses; White standard errors that correct for heteroskedasticity but not clustering are given in square brackets; and usual default estimates that assume iid errors are given in braces. An asterisk denotes a coefficient with absolute t-value in excess of 2 using the correct standard error.

As an example of a Hausman test of fixed versus random effects, consider the coefficient of currency union. The FGLS and FE coefficients are very close, with difference $\widetilde{\beta}_{FE} - \widehat{\beta}_{RE} = -0.0645 + 0.0630 = -0.0015$ that is small relative to the estimated coefficients and their standard errors (0.0057 and 0.0058), suggesting that a random effects model is adequate. Actual implementation of a Hausman test is problematic, however, as the estimated variance of $\widetilde{\beta}_{FE} - \widehat{\beta}_{RE} = 0.0057^2 - 0.0058^2 < 0$. Such negative differences can arise for the Hausman test, see for example Ruud (1984, p.214). There is a greater difference between the OLS and FE estimates (the difference is 0.0113), but as noted in subsection 2.6 the standard implementation of the Hausman test is no longer valid since the OLS estimator is inefficient.

We conducted a similar analysis of the related model of Tenreyro and Barro (2003) for comovement of prices. Again there was considerable clustering of the errors, with $\widehat{c} = 8.7$, leading to qualitatively similar conclusions of large underestimation of OLS standard errors and large efficiency gains for GLS. Feasible GLS estimates were quite close to fixed effects estimates, which Tenreyro and Barro (2003) also calculate.

As yet another application, similar analysis was applied to dependent variable the natural logarithm of bilateral trade (in US\$), replicating column 1 of Table 3 of Rose and Engel (2002). Then there are 127 countries and data available for 4618 country pairs. The results are given in Table 4.2. A big difference is that for this dependent variable and model $\widehat{c} = 0.2665$, so clustering is much less important. Nonetheless there is still significant bias in standard errors with the correct OLS standard errors being 1.5 to 4 times the default standard errors. White standard errors are within 20 percent of the default OLS standard errors so are also considerably downwards biased. The greatest difference is for the last two regressors which vary over one dimension not both dimensions, as discussed earlier. The efficiency gains of FGLS are not as great as for the real exchange volatility model, due to the lower value of c, but even here the standard errors for the first two regressors, which vary over dimensions, are roughly halved. To enable comparison with a fixed effects model, the regressors real GDP per Capita and Real GDP again need to be dropped. Then $\widehat{c} = 1.9814$ is higher, resulting in increased bias in the default OLS standard errors, greater gains to feasible GLS estimation, and closeness of fixed and random effects estimates of slope coefficients.

16

Table 4.2: Application: Gravity model for bilateral trade, Rose and Engel (2002, Table 3).

	Rose and Engel Model		Reduced Model		
	OLS	FGLS	OLS	FGLS	FE
Intercept	-32.8421*	-32.0411*	18.4809*	20.9810*	15.0397*
Correct se	(1.9174)	(1.6155)	(0.6458)	(0.4467)	(1.8063)
White se	[0.6823]		[0.5012]		
Default se	{0.6004}		{0.5102}		
Currency Union	1.8588*	1.3464*	-3.1279*	1.1416*	1.2737*
Correct se	(0.5359)	(0.3251)	(1.2416)	(0.3365)	(0.3380)
White se	[0.4553]		[0.5514]		
Default se	{0.3666}		{0.6518}		
(Log) Distance	-1.3667	-1.5858*	-0.1539	-1.5907*	-1.5961*
Correct se	(0.0755)	(0.0319)	(0.1955)	(0.0322)	(0.0321)
White se	[0.0348]		[0.0604]		
Default se	{0.0345}		{0.0614}		
(Log Product) Real GDP per Capita	0.7656*	0.7649*			
Correct se	(0.0741)	(0.0678)			
White se	[0.0204]				
Default se	{0.0198}				
(Log Product) Real GDP	0.8670*	0.8455*			
Correct se	(0.0414)	(0.0367)			
White se	[0.0128]				
Default se	{0.0116}				
s_ε^2	2.0295		2.0291		
s_α^2	0.5409		4.0204		
$\widehat{c}$	0.2665		1.9814		

Note: Cross-section data for 4618 country-pairs formed from 127 countries. Table 3 of Rose and Engel multiplies all coefficients by 100. Standard errors for the slope coefficients that correct for clustering (see text) are given in parentheses; White standard errors that correct for heteroskedasticity but not clustering are given in square brackets; and usual default estimates that assume iid errors are given in braces. An asterisk denotes a coefficient with absolute t-value in excess of 2 using the correct standard error.

5. Conclusions

This paper presents methods to obtain correct standard errors for OLS regression using paired data such as country-pair data, and to implement more efficient FGLS

estimation if a random effects model is estimated. Simulation and application demonstrate considerable bias in estimated standard errors that fail to control for the clustering inherent with such data, and the potential for considerable efficiency gain by feasible GLS estimation.

6. Appendix

6.1. Appendix: Analytical Results for OLS

We begin with a summary of results for the $T \times n$ matrix $\mathbf{L} = \mathbf{M} + \mathbf{P}$ defined in section 2.1.

Proposition: For $\mathbf{e}_T$ a $T = n(n-1)/2$ column vector of ones and $\mathbf{e}_n$ a n column vector of ones we have:

1. $\mathbf{e}_T'\mathbf{L}=(n-1)\mathbf{e}_n'$

2. $\mathbf{L}\mathbf{e}_n=2\mathbf{e}_T$

3. $\mathbf{L}'\mathbf{L} =(n-2)\mathbf{I}_n + \mathbf{e}_n\mathbf{e}_n'$

4. $\mathbf{X}'\mathbf{L} =(n-1)\begin{bmatrix} \bar{\mathbf{x}}_1 & \cdots & \bar{\mathbf{x}}_n \end{bmatrix}$
   where $\bar{\mathbf{x}}_j = [(\mathbf{x}_{j,j+1} + \cdots + \mathbf{x}_{j,n}) + (\mathbf{x}_{1,j} + \cdots + \mathbf{x}_{j-1,j})]/(n-1)$

Proof: For property 1,

$$
\begin{aligned}
\mathbf{e}_T'\mathbf{P} &= \begin{bmatrix} \mathbf{e}_{n-1}' & \mathbf{e}_{n-2}' & \cdots & \mathbf{e}_1' \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_{n-1} \end{bmatrix} \\
&= \sum_{i=1}^{n-1} \mathbf{e}_{n-i}'\mathbf{P}_i \\
&= \sum_{i=1}^{n-1} \mathbf{e}_{n-i}' \begin{bmatrix} \mathbf{0}_{n-i}\mathbf{0}_{i-1}' & \mathbf{e}_{n-i} & \mathbf{0}_{n-i}\mathbf{0}_{n-i}' \end{bmatrix} \\
&= \sum_{i=1}^{n-1} \begin{bmatrix} \mathbf{0}_{i-1}' & (n-i) & \mathbf{0}_{n-i}' \end{bmatrix} \\
&= \begin{bmatrix} (n-1) & (n-2) & \cdots & 1 & 0 \end{bmatrix}
\end{aligned}
$$

18

and

$$
\begin{aligned}
\mathbf{e}_T'\mathbf{M} &= \sum_{i=1}^{n-1} \mathbf{e}_{n-i}'\mathbf{M}_i \\
&= \sum_{i=1}^{n-1} \mathbf{e}_{n-i}' \\
&= \sum_{i=1}^{n-1} \begin{bmatrix} \mathbf{0}_i' & \mathbf{e}_{n-i}' \end{bmatrix} \begin{bmatrix} \mathbf{0}_{n-i}\mathbf{0}_i' & \mathbf{I}_{n-i} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 1 & \cdots & & \cdots & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 1 & \cdots & 1 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & \cdots & \cdots & & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 1 & \cdots & (n-2) & (n-1) \end{bmatrix}.
\end{aligned}
$$

Property 1 follows using

$$
\mathbf{e}_T'\mathbf{L} = \mathbf{e}_T'(\mathbf{P} + \mathbf{M}) = \mathbf{e}_T'\mathbf{P} + \mathbf{e}_T'\mathbf{M}.
$$

To obtain property 2 start with $\mathbf{P}_i\mathbf{e}_n$ which is a $(n-i) \times n$ matrix times $n \times 1$ vector yields $(n-i)$ vector

$$
\mathbf{P}_i\mathbf{e}_n = \begin{bmatrix} \mathbf{0}_{n-i}\mathbf{0}_{i-1}' & \mathbf{e}_{n-i} & \mathbf{0}_{n-i}\mathbf{0}_{n-i}' \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{e}_n,
$$

so

$$
\mathbf{P}\mathbf{e}_n = \begin{bmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_{n-1} \end{bmatrix} \mathbf{e}_n = \begin{bmatrix} \mathbf{P}_1\mathbf{e}_n \\ \vdots \\ \mathbf{P}_{n-1}\mathbf{e}_n \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{n-1} \\ \vdots \\ \mathbf{e}_1 \end{bmatrix} = \mathbf{e}_T.
$$

Similarly $\mathbf{M}\mathbf{e}_n = \mathbf{e}_T$ using

$$
\mathbf{M}_i\mathbf{e}_n = \begin{bmatrix} \mathbf{0}_{n-i}\mathbf{0}_i' & \mathbf{I}_{n-i} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{e}_n.
$$

Property 2 follows using $\mathbf{L}\mathbf{e}_n = \mathbf{P}\mathbf{e}_n + \mathbf{M}\mathbf{e}_n$.

For property 3 start with

$$
\mathbf{P}_i'\mathbf{P}_i = \begin{bmatrix} \mathbf{0}_{i-1}\mathbf{0}_{n-i}' \\ \mathbf{e}_{n-i}' \\ \mathbf{0}_{n-i}\mathbf{0}_{n-i}' \end{bmatrix} \begin{bmatrix} \mathbf{0}_{n-i}\mathbf{0}_{i-1}' & \mathbf{e}_{n-i} & \mathbf{0}_{n-i}\mathbf{0}_{n-i}' \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (n-i) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},
$$

19

so
$$\mathbf{P}'\mathbf{P} = \sum_{i=1}^{n-1} \mathbf{P}'_i\mathbf{P}_i = \begin{bmatrix} (n-1) & 0 & \cdots & \cdots & 0 \\ 0 & (n-2) & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 & \vdots \\ \vdots & \vdots & 0 & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

Similarly
$$\mathbf{M}'_i\mathbf{M}_i = \begin{bmatrix} \mathbf{0}_{i-1}\mathbf{0}'_{n-i} \\ \mathbf{I}_{n-i} \end{bmatrix} \begin{bmatrix} \mathbf{0}_{n-i}\mathbf{0}'_{i-1} & \mathbf{I}_{n-i} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-i} \end{bmatrix},$$

so
$$\mathbf{M}'\mathbf{M} = \sum_{i=1}^{n-1} \mathbf{M}'_i\mathbf{M}_i = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 & \vdots \\ \vdots & \vdots & 0 & (n-2) & 0 \\ 0 & \cdots & \cdots & \cdots & (n-1) \end{bmatrix}.$$

Also
$$\mathbf{P}'_i\mathbf{M}_i = \begin{bmatrix} \mathbf{0}_{i-1}\mathbf{0}'_{n-i} \\ \mathbf{e}'_{n-i} \\ \mathbf{0}_{n-i}\mathbf{0}'_{n-i} \end{bmatrix} \begin{bmatrix} \mathbf{0}_{n-i}\mathbf{0}'_{i-1} & \mathbf{I}_{n-i} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{e}'_{n-i} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

so $\mathbf{P}'_i\mathbf{M}_i$ has i^{th} row with last $(n-i)$ entries a one and all others are zero. It follows that
$$\mathbf{P}'\mathbf{M} = \sum_{i=1}^{n-1} \mathbf{P}'_i\mathbf{M}_i = \begin{bmatrix} 0 & 0 & 1 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 1 & \vdots \\ \vdots & \vdots & 0 & 0 & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

Hence $\mathbf{P}'\mathbf{P} + \mathbf{M}'\mathbf{M} = (n-1)\mathbf{I}_n$ and $(\mathbf{P}'\mathbf{M} + \mathbf{M}'\mathbf{P}) = \mathbf{e}_n\mathbf{e}'_n - \mathbf{I}_n$, so $\mathbf{L}'\mathbf{L} = (n-2)\mathbf{I}_n + \mathbf{e}_n\mathbf{e}'_n$.

6.2. Appendix: Analytical Results for GLS

To obtain the expression for $\mathbf{\Omega}^{-1}$ given in (2.17), begin with (2.15) and substitute $\mathbf{L}'\mathbf{L} = (n-2)\mathbf{I}_n + \mathbf{e}_n\mathbf{e}_n'$ from the proposition, so

$$
\begin{aligned}
\mathbf{\Omega}^{-1} &= \sigma_\varepsilon^{-2}[\mathbf{I}_T - c\mathbf{L}[\mathbf{I}_n + c(n-2)\mathbf{I}_n + c\mathbf{e}_n\mathbf{e}_n']^{-1}\mathbf{L}'] \\
&= \sigma_\varepsilon^{-2}[\mathbf{I}_T - c\mathbf{L}[(1 + c(n-2))\mathbf{I}_n + c\mathbf{e}_n\mathbf{e}_n']^{-1}\mathbf{L}'] \\
&= \sigma_\varepsilon^{-2}[\mathbf{I}_T - \mathbf{L}[a\mathbf{I}_n + \mathbf{e}_n\mathbf{e}_n']^{-1}\mathbf{L}'] \text{ where } a = (1 + c(n-2))/c \\
&= \sigma_\varepsilon^{-2}\left[\mathbf{I}_T - \mathbf{L}\left[a^{-1}\left(\mathbf{I}_n - \frac{1}{a+n}\mathbf{e}_n\mathbf{e}_n'\right)\mathbf{L}'\right]\right] \\
&= \sigma_\varepsilon^{-2}\left[\mathbf{I}_T - a^{-1}\left(\mathbf{L}\mathbf{L}' - \frac{1}{a+n}\mathbf{L}\mathbf{e}_n\mathbf{e}_n'\mathbf{L}'\right)\right] \\
&= \sigma_\varepsilon^{-2}\left[\mathbf{I}_T - a^{-1}\left(\mathbf{L}\mathbf{L}' - \frac{1}{a+n}2\mathbf{e}_T 2\mathbf{e}_T'\right)\right] \\
&= \sigma_\varepsilon^{-2}\left[\mathbf{I}_T - \frac{c}{1 + c(n-2)}\left(\mathbf{L}\mathbf{L}' - \frac{4c}{1 + 2c(n-1)}\mathbf{e}_T\mathbf{e}_T'\right)\right],
\end{aligned}
$$

which is (2.17). The fourth equality uses

$$
[a\mathbf{I}_n + \mathbf{e}_n\mathbf{e}_n']^{-1} = a^{-1}\left[\mathbf{I}_n - \frac{1}{a+n}\mathbf{e}_n\mathbf{e}_n'\right],
$$

shown below, the sixth equality uses $\mathbf{L}\mathbf{e}_n = 2\mathbf{e}_T$, see the proposition, and the final equality uses

$$
\begin{aligned}
\frac{4}{a+n} &= \frac{4}{(1 + c(n-2))/c + n} \\
&= \frac{4c}{(1 + c(n-2)) + nc} \\
&= \frac{4c}{(1 + 2c(n-1))}.
\end{aligned}
$$

To verify that $[a\mathbf{I}_n + \mathbf{e}_n\mathbf{e}_n']^{-1} = a^{-1}\left[\mathbf{I}_n - \frac{1}{a+n}\mathbf{e}_n\mathbf{e}_n'\right]$, note that

$$[a\mathbf{I}_n + \mathbf{e}_n\mathbf{e}_n'] \times a^{-1}\left[\mathbf{I}_n - \frac{1}{a+n}\mathbf{e}_n\mathbf{e}_n'\right]$$

$$= a^{-1}\left[a\mathbf{I}_n + \mathbf{e}_n\mathbf{e}_n' - \frac{a}{a+n}\mathbf{e}_n\mathbf{e}_n' - \frac{1}{a+n}\mathbf{e}_n\mathbf{e}_n'\mathbf{e}_n\mathbf{e}_n'\right]$$

$$= a^{-1}\left[a\mathbf{I}_n + \mathbf{e}_n\mathbf{e}_n' - \frac{a}{a+n}\mathbf{e}_n\mathbf{e}_n' - \frac{n}{a+n}\mathbf{e}_n\mathbf{e}_n'\right] \text{ since } \mathbf{e}_n'\mathbf{e}_n = n$$

$$= \mathbf{I}_n.$$

Now consider GLS when $\mathbf{X} = \mathbf{e}_T$, so $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (\mathbf{e}_T'\boldsymbol{\Omega}^{-1}\mathbf{e}_T)^{-1}\mathbf{e}_T'\boldsymbol{\Omega}^{-1}\mathbf{y}$. Then using expression (2.17) for $\boldsymbol{\Omega}^{-1}$,

$$
\begin{aligned}
(\mathbf{e}_T'\boldsymbol{\Omega}^{-1}\mathbf{e}_T')^{-1} &= \sigma_\varepsilon^2\left[T - \frac{c}{1+c(n-2)}\left(2(n-1)T - \frac{4c}{1+2c(n-1)}T^2\right)\right]^{-1} \\
&= \sigma_\varepsilon^2 T^{-1}\left[1 - \frac{c}{1+c(n-2)}\left(2(n-1) - \frac{4c}{1+2c(n-1)}\frac{n(n-1)}{2}\right)\right]^{-1} \\
&= \sigma_\varepsilon^2 T^{-1}\left[1 - \frac{2c(n-1)}{1+c(n-2)}\left(1 - \frac{cn}{1+2c(n-1)}\right)\right]^{-1} \\
&= \sigma_\varepsilon^2 T^{-1}\left[1 - \frac{2c(n-1)}{1+c(n-2)}\left(\frac{1+c(n-2)}{1+2c(n-1)}\right)\right]^{-1} \\
&= \sigma_\varepsilon^2 T^{-1}\left[1 - \frac{2c(n-1)}{1+2c(n-1)}\right]^{-1} \\
&= \sigma_\varepsilon^2 T^{-1}\left[\frac{1}{1+2c(n-1)}\right]^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{e}_T'\boldsymbol{\Omega}^{-1}\mathbf{y} &= \sigma_\varepsilon^{-2}\left[T\bar{y} - \frac{c}{1+c(n-2)}\left(2(n-1)T\bar{y} - \frac{4c}{1+2c(n-1)}T^2\bar{y}\right)\right] \\
&= \sigma_\varepsilon^{-2}\left[T - \frac{c}{1+c(n-2)}\left(2(n-1)T - \frac{4c}{1+2c(n-1)}T^2\right)\right]\bar{y} \\
&= \sigma_\varepsilon^{-2}T\left[\frac{1}{1+2c(n-1)}\right]\bar{y},
\end{aligned}
$$

where the first equality uses

$$\mathbf{L} = \mathbf{e}_T$$
$$\mathbf{X}'\mathbf{y} = \mathbf{e}_T'\mathbf{y} = T\bar{y}$$
$$\mathbf{X}'\mathbf{L}\mathbf{L}'\mathbf{y} = (n-1)\mathbf{e}_n'n \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix} = (n-1)2T\bar{y}$$
$$\mathbf{X}'\mathbf{e}_T\mathbf{e}_T'\mathbf{X} = \mathbf{e}_T'\mathbf{e}_T\mathbf{e}_T'\mathbf{y} = T \times T\bar{y} = T^2\bar{y}.$$

where the third line uses $\mathbf{L}'\mathbf{y}$ given in the proposition. Upon cancellation $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (\mathbf{e}_T'\boldsymbol{\Omega}^{-1}\mathbf{e}_T)^{-1}\mathbf{e}_T'\boldsymbol{\Omega}^{-1}\mathbf{y} = \bar{y}$.

7. References

Cheng, I-H. and H.J. Wall (2003), "Controlling for Heterogeneity in Gravity Models of Trade and Integration," Federal Reserve Bank of St. Louis working paper 1999-010D, revised May 2003.

Kloek, T. (1981), "OLS Estimation in a Model where a Microvariable is Explained by Aggregates and Contemporaneous Disturbances are Equicorrelated," *Econometrica*, 49, 205-07.

Magnus, J.R. and H. Neudecker (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Chichester: Wiley.

Matyas, L. (1997), "Proper Econometric Specification of the Gravity Model," *World-Economy*, 20(3), 363-68.

Matyas, L. (1998), "The Gravity Model: Some Econometric Considerations," *World-Economy*, 21(3), 397-401.

Moulton, B.R. (1986), "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 32, 385-97.

Moulton, B.R. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 72, 334-38.

Rose, A. C. and Engel (2002), "Currency Unions and International Integration," *Journal of Money, Credit and Banking*, 34(4), 1067-1089.

Ruud, P. (1984), "Tests of Specification in Econometrics," *Econometric Reviews*, 3(2), 212-242.

Scott, A.J. and D. Holt (1982), "The Effect of Two-Stage Sampling on Ordinary Least Squares Methods," *Journal of the American Statistical Association*, 77, 848-54.

Serlenga, L. and Y. Shin (2004), "Gravity Models of the Intra-EU Trade: Application of the Hausman-Taylor Estimation in Heterogeneous Panels with Common Time-specific Factors," manuscript, University of Edinburgh.

Tenreyro, S. and R. Barro (2003), "Economic Effects of Currency Unions," NBER Working Paper No. 9435, January.