

Robust Inference for Dyadic Data^{*}

A. Colin Cameron[†] and Douglas L. Miller[‡]

December 31, 2014

Abstract

In this paper we consider inference with paired or dyadic data, such as cross-section and panel data on trade between two countries. Regression models with such data have a complicated pattern of error correlation. For example, errors for US-UK trade may be correlated with those for any other country pair that includes either the US or UK. We consider models with regressors treated as predetermined and stationary. The standard cluster-robust variance estimator or sandwich estimator for one-way clustering is inadequate. The two-way cluster robust estimator is a substantial improvement, but still understates standard errors. Some studies in social network data analysis have addressed this issue. The network in international trade studies is much denser than in typical network studies, so it becomes especially important to control for dyadic error correlation. In applications with the gravity model of trade we find that even after inclusion of country fixed effects, standard errors that properly control for dyadic error correlation can be several times those being reported using current methods.

Keywords: cluster-robust standard errors; two-way clustering; dyadic data; paired data; network data; country-pair data; gravity model; international trade.

JEL Classification: C12, C21, C23.

^{*}Preliminary version presented at the Winter North American Meetings of the Econometric Society, Boston, January 5, 2015.

[†]Dept. of Economics, University of California - Davis. accameron@ucdavis.edu

[‡]Dept. of Economics, University of California - Davis. dlmiller@ucdavis.edu

1. Introduction

A key component of empirical research is conducting accurate statistical inference. One challenge to this is the possibility of errors being correlated across observations. In this paper we present a variance matrix estimator that provides cluster-robust inference for regressions with paired or dyadic data, such as country-pair data analyzed frequently in international trade applications.

Controlling for clustering can be very important, as failure to do so can lead to massively under-estimated standard errors and consequent over-rejection using standard hypothesis tests. Moulton (1986, 1990) and Bertrand, Duflo and Mullainathan (2004) demonstrated that the need to control for one-way clustering arose in a much wider range of settings than had been appreciated by microeconometricians. Most notably, even modest within-group error correlation can greatly inflate default standard errors for a grouped regressor (one observed at a more highly aggregated level than the dependent variable). The common way to control for clustering is to use one-way “cluster-robust” standard errors that generalize those of White (1980) for independent heteroskedastic errors. Key references include Shah et al. (1977) for clustered sampling, White (1984) for a multivariate dependent variable, Liang and Zeger (1986) for estimation in a generalized estimating equations setting, and Arellano (1987) and Hansen (2007) for linear panel models. Wooldridge (2003) and Cameron and Miller (2011, 2015) provide surveys.

Cluster robust inference for the one-way case has been generalized to two-way and multi-way clustering; see Miglioretti and Heagerty (2006), Cameron, Gelbach and Miller (2011) and Thompson (2011). An example is cross-section data of individual wages on two grouped regressors with different types of groupings, such as occupation-level job injury risk and industry-level job injury risk.

In this paper we consider a different departure from one-way clustering, that due to paired or dyadic data, such as that for trade flows between countries. Then model errors are likely to be correlated between country-pair observations that have a country in common. For example, errors for US-UK trade may be correlated with those for any other country pair that includes either the US or UK. Standard approaches, including using country fixed effects or (for panel data) country-pair fixed effects, do not fully account for this error correlation and lead to standard error estimates that can be misleadingly low.

Cameron and Golotvina (2005) developed a modified two-way random effects model for

dyadic data, but this has the limitation of inference being valid only if the underlying errors are indeed i.i.d. Cameron, Gelbach and Miller (2011, section 4.2) propose using two-way cluster-robust standard errors that for country pair (g, h) cluster on country g and on country h . In the current paper we find this provides considerable improvement over current practice, but does not control for all of the potential error correlation.

We instead use methods that have been proposed in the context of inference for social network data. Leading examples are Frank and Snijders (1994) and Snijders and Borgatti (1999) in the non-regression case, and Fafchamps and Gubert (2007) in regression applications. In many network applications an individual may interact directly with relatively few others. In international trade applications, by contrast, a country may trade with all other countries. This means that failure to control for clustering may have a much larger impact on standard error estimates.

Like the one-way cluster-robust method, our methods assume that the number of clusters (here countries) goes to infinity. It is well-known that standard Wald tests based on one-way and two-way cluster-robust standard errors can over-reject when there are few clusters. Similar problems can be expected to exist for dyadic-robust standard errors, and we consider finite-cluster issues in some detail.

The methods are presented in Section 2. In Section 3 we present Monte Carlo experiments. Section 4 presents two international trade applications that demonstrate the importance of controlling for clustering with dyadic data, even when country fixed effects are included in the model. Section 5 concludes.

2. Cluster-Robust Inference

This section emphasizes the OLS estimator, for simplicity. We begin with reviews of one-way clustering and two-way clustering before considering dyadic clustering. The section concludes with extension from OLS to m-estimators, such as probit and logit, and to GMM estimators.

2.1. One-Way Clustering

The linear model with one-way clustering is

$$y_{ig} = \mathbf{x}'_{ig}\boldsymbol{\beta} + u_{ig}, \quad i = 1, \dots, N, \quad (2.1)$$

where i denotes the i^{th} of N individuals in the sample, g denotes the g^{th} of G clusters, $E[u_{ig}|\mathbf{x}_{ig}] = 0$, and error independence across clusters is assumed so that

$$E[u_{ig}u_{jg'}|\mathbf{x}_{ig}, \mathbf{x}_{jg'}] = 0, \text{ unless } g = g'. \quad (2.2)$$

Errors for individuals belonging to the same group may be correlated, with quite general heteroskedasticity and correlation. Grouping observations by cluster, so $\mathbf{y}_g = \mathbf{X}_g\boldsymbol{\beta} + \mathbf{u}_g$, and stacking over clusters yields $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where \mathbf{y} and \mathbf{u} are $N \times 1$ vectors, and \mathbf{X} is an $N \times K$ matrix. The OLS estimator is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \left(\sum_{g=1}^G \mathbf{X}_g'\mathbf{X}_g \right)^{-1} \sum_{g=1}^G \mathbf{X}_g'\mathbf{y}_g, \quad (2.3)$$

where \mathbf{X}_g has dimension $N_g \times K$ and \mathbf{y}_g has dimension $N_g \times 1$, with N_g observations in cluster g .

Under commonly assumed restrictions on moments and heterogeneity of the data, $\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ has a limit normal distribution with variance matrix

$$\left(\lim_{G \rightarrow \infty} \frac{1}{G} \sum_{g=1}^G E[\mathbf{X}_g'\mathbf{X}_g] \right)^{-1} \left(\lim_{G \rightarrow \infty} \frac{1}{G} \sum_{g=1}^G E[\mathbf{X}_g'\mathbf{u}_g\mathbf{u}_g'\mathbf{X}_g] \right) \left(\lim_{G \rightarrow \infty} \frac{1}{G} \sum_{g=1}^G E[\mathbf{X}_g'\mathbf{X}_g] \right)^{-1}. \quad (2.4)$$

The earliest work posited a model for the cluster error variance matrices $\boldsymbol{\Omega}_g = V[\mathbf{u}_g|\mathbf{X}_g] = E[\mathbf{u}_g\mathbf{u}_g'|\mathbf{X}_g]$, in which case $E[\mathbf{X}_g'\mathbf{u}_g\mathbf{u}_g'\mathbf{X}_g] = E[\mathbf{X}_g'\boldsymbol{\Omega}_g\mathbf{X}_g]$ can be estimated given a consistent estimate $\hat{\boldsymbol{\Omega}}_g$ of $\boldsymbol{\Omega}_g$. Given these strong assumptions more efficient feasible GLS estimation is additionally possible.

Current applied studies instead use the cluster-robust variance matrix estimate

$$\hat{V}[\hat{\boldsymbol{\beta}}] = c \times (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{g=1}^G \mathbf{X}_g'\hat{\mathbf{u}}_g\hat{\mathbf{u}}_g'\mathbf{X}_g \right) (\mathbf{X}'\mathbf{X})^{-1}, \quad (2.5)$$

where $\hat{\mathbf{u}}_g = \mathbf{y}_g - \mathbf{X}_g\hat{\boldsymbol{\beta}}$ and $c = G/(G-1)$ or $c = [G/(G-1)] \times [(N-1)/(N-k)]$ is a finite-sample adjustment. This provides a consistent estimate of the variance matrix if $G^{-1} \sum_{g=1}^G \mathbf{X}_g'\hat{\mathbf{u}}_g\hat{\mathbf{u}}_g'\mathbf{X}_g - G^{-1} \sum_{g=1}^G E[\mathbf{X}_g'\mathbf{u}_g\mathbf{u}_g'\mathbf{X}_g] \xrightarrow{p} \mathbf{0}$ as $G \rightarrow \infty$. White (1984, p.134-142) presented formal theorems for a multivariate dependent variable, directly applicable to balanced clusters. Liang and Zeger (1986) proposed this method for estimation in a generalized estimating equations setting, Arellano (1987) proposed this method for the fixed effects estimator in linear panel models, and Rogers (1993) popularized this method in applied econometrics by incorporating it in Stata. Note that (2.5) does not require specification of a model for $\boldsymbol{\Omega}_g$, and thus it permits quite general forms of $\boldsymbol{\Omega}_g$.

If the primary source of clustering is due to group-level common shocks, a useful approximation is that for the j^{th} regressor the default OLS variance estimate based on $s^2 (\mathbf{X}'\mathbf{X})^{-1}$, where s is the estimated standard deviation of the error, should be inflated by $\tau_j \simeq 1 + \rho_{x_j}\rho_u(\bar{N}_g - 1)$, where ρ_{x_j} is a measure of the within cluster correlation of x_j , ρ_u is the within cluster error correlation, and \bar{N}_g is the average cluster size. Moulton (1986, 1990) pointed out that in many settings the adjustment factor τ_j can be large even if ρ_u is small. For example, if $N_G = 81$, $\rho_{x_j} = 1$, and $\rho_u = 0.10$ then default OLS standard errors should be deflated by approximately $\sqrt{9} = 3$. In general it is especially important to control for clustering when both the errors and regressors are correlated within cluster and there are many observations per cluster.

A helpful informal presentation of (2.5) is that

$$\widehat{\mathbf{V}}[\widehat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{X})^{-1}\widehat{\mathbf{B}}(\mathbf{X}'\mathbf{X})^{-1}, \quad (2.6)$$

where the central matrix

$$\begin{aligned} \widehat{\mathbf{B}} &= c \times \sum_{g=1}^G \mathbf{X}'_g \widehat{\mathbf{u}}_g \widehat{\mathbf{u}}'_g \mathbf{X}_g \\ &= c \times \mathbf{X}' \begin{bmatrix} \widehat{\mathbf{u}}_1 \widehat{\mathbf{u}}'_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{u}}_2 \widehat{\mathbf{u}}'_2 & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \widehat{\mathbf{u}}_G \widehat{\mathbf{u}}'_G \end{bmatrix} \mathbf{X} \\ &= c \times \mathbf{X}' (\widehat{\mathbf{u}}\widehat{\mathbf{u}}' * \mathbf{S}^G) \mathbf{X}, \end{aligned} \quad (2.7)$$

where c is the finite-sample adjustment, $*$ denotes element-by-element multiplication and \mathbf{S}^G is an $N \times N$ indicator matrix, or selection matrix, with ij^{th} entry equal to one if the i^{th} and j^{th} observation belong to the same cluster and equal to zero otherwise. An intuitive explanation of the asymptotic theory is that the indicator matrix \mathbf{S}^G must zero out a large amount of $\widehat{\mathbf{u}}\widehat{\mathbf{u}}'$, or, asymptotically equivalently, $\mathbf{u}\mathbf{u}'$. Here there are $N^2 = (\sum_{g=1}^G N_g)^2$ terms in $\widehat{\mathbf{u}}\widehat{\mathbf{u}}'$ and all but $\sum_{g=1}^G N_g^2$ of these are zeroed out. For fixed N_g , $(\sum_{g=1}^G N_g^2/N^2) \rightarrow 0$ as $G \rightarrow \infty$. In particular, for balanced clusters $N_g = N/G$, so $(\sum_{g=1}^G N_g^2)/N^2 = 1/G \rightarrow 0$ as $G \rightarrow \infty$.

An equivalent expression to (2.7) is

$$\widehat{\mathbf{B}} = c \times \sum_{i=1}^N \sum_{j=1}^N \mathbf{1}[g = g'] \times \widehat{u}_{ig} \widehat{u}_{jg'} \mathbf{x}_{ig} \mathbf{x}'_{jg'}, \quad (2.8)$$

where the indicator $\mathbf{1}[A]$ equals 1 if event A occurs and equals 0 otherwise.

2.2. Two-Way Clustering

Now consider situations where each observation may belong to more than one “dimension” of groups. For instance, if there are two dimensions of grouping, each individual i will belong to a group $g \in \{1, 2, \dots, G\}$, as well as to a group $h \in \{1, 2, \dots, H\}$. The regression model is now

$$y_{igh} = \mathbf{x}'_{igh}\boldsymbol{\beta} + u_{igh}, \quad i = 1, \dots, N, \quad (2.9)$$

where we assume that for $i \neq j$

$$E[u_{igh}u_{jg'h'} | \mathbf{x}_{igh}, \mathbf{x}_{jg'h'}] = 0, \text{ unless } g = g' \text{ or } h = h'. \quad (2.10)$$

If errors belong to the same group (along either dimension), they may have an arbitrary correlation.

The intuition for the variance estimator in this case is a simple extension of (2.8) for one-way clustering. We keep those elements of $\widehat{\mathbf{u}}\widehat{\mathbf{u}}'$ where the i^{th} and j^{th} observations share a cluster in any dimension. This yields $\widehat{\mathbf{V}}[\widehat{\boldsymbol{\beta}}]$ in (2.6) where

$$\widehat{\mathbf{B}} = \sum_{i=1}^N \sum_{j=1}^N \mathbf{1}[g = g' \text{ and/or } h = h'] \times \widehat{u}_{igh}\widehat{u}_{jg'h'}\mathbf{x}_{igh}\mathbf{x}'_{jg'h'}. \quad (2.11)$$

Now the indicator for sharing a cluster in any dimension can be calculated as the sum of an indicator for whether $g = g'$ and an indicator for whether $h = h'$, and then subtracting an indicator for whether $g = g'$ and $h = h'$ to avoid double counting. Thus in (2.6)

$$\begin{aligned} \widehat{\mathbf{B}} = & c_1 \times \sum_i \sum_j \mathbf{1}[g = g'] \widehat{u}_{igh}\widehat{u}_{jg'h'}\mathbf{x}_{igh}\mathbf{x}'_{jg'h'} + c_2 \times \sum_i \sum_j \mathbf{1}[h = h'] \widehat{u}_{igh}\widehat{u}_{jg'h'}\mathbf{x}_{igh}\mathbf{x}'_{jg'h'} \\ & - c_3 \times \sum_{i=1}^N \sum_{j=1}^N \mathbf{1}[(g, h) = (g', h')] \widehat{u}_{igh}\widehat{u}_{jg'h'}\mathbf{x}_{igh}\mathbf{x}'_{jg'h'}, \end{aligned} \quad (2.12)$$

where c_1 , c_2 and c_3 are finite-sample adjustments. It follows that the three components can be separately computed by OLS regression of \mathbf{y} on \mathbf{X} with variance matrix estimates based on: (1) clustering on $g \in \{1, 2, \dots, G\}$; (2) clustering on $h \in \{1, 2, \dots, H\}$; and (3) clustering on $(g, h) \in \{(1, 1), \dots, (G, H)\}$. $\widehat{\mathbf{V}}[\widehat{\boldsymbol{\beta}}]$ is the sum of the first and second components, minus the third component.

2.3. Dyadic or Paired Clustering

Now consider cross-section dyadic data, such as trade between countries g and h . We consider specialization to the bidirectional trade volume case. Then countries do not trade with themselves, so $y_{gg} = 0$ and we drop pairs with $g = h$. And for trade volume that is bidirectional, rather than unidirectional data on import volume or on export volume, $y_{gh} = y_{hg}$ so we additionally drop pairs (g, h) for which $g > h$ to avoid duplication. It follows that

$$y_{gh} = \mathbf{x}'_{gh}\boldsymbol{\beta} + u_{gh}, \quad h = g + 1, \dots, G, \quad g = 1, \dots, G - 1, \quad (2.13)$$

where G is the number of countries. If data are available for all pairs then a dataset with G countries has a total of $G(G - 1)/2$ country-pair observations.

Under dyadic correlation the errors u_{gh} and $u_{g'h'}$ are assumed to be correlated if observations have either component of the dyad in common, and are otherwise uncorrelated. Thus we assume that

$$E[u_{gh}u_{g'h'}|\mathbf{x}_{gh}, \mathbf{x}_{g'h'}] = 0, \text{ unless } g = g' \text{ or } h = h' \text{ or } g = h' \text{ or } h = g'. \quad (2.14)$$

To make ideas concrete, consider the case of cross-section data on bilateral trade between four countries, numbered 1, 2, 3, and 4. Then there are six country-pair observations, namely (1, 2), (1, 3), (1, 4), (2, 3), (2, 4) and (3, 4), and there are 36 error correlations.

Clustering on country-pair controls for correlation when $(g, h) = (g', h')$. Then only the diagonal entries in the table below are nonzero; these are denoted cp. Clustering on country-pair coincides in the cross-section case with using heteroskedastic robust standard errors.

$(g, h) \setminus (g', h')$	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
(1,2)	cp	2way	2way	dyad	dyad	0
(1,3)	2way	cp	2way	2way	0	dyad
(1,4)	2way	2way	cp	0	x	x
(2,3)	dyad	2way	0	cp	x	dyad
(2,4)	dyad	0	2way	2way	cp	x
(3,4)	0	dyad	2way	dyad	2way	cp

Two-way cluster-robust standard errors, with clustering on g and on h , additionally control for possible correlation when $g = g'$ and/or $h = h'$. These additional correlations are denoted by 2way in the table.

Dyadic clustering additionally picks up cases where $g = h'$ or $h = g'$. These additional cases are denoted **dyad** in the table.

The table indicates a potential problem when there are relatively few countries - many of the pairs may be correlated. In this example, extreme as G is so small, 30 out of 36 terms in the error variance matrix are nonzero.

Even when there are a relatively large number of countries, a substantial fraction of the error correlations may be nonzero. When data are available for all country pairs there are $[G(G-1)/2]^2$ error correlations and some algebra reveals that given (2.14) there are $[G(G-1)/2] \times (2G-3)$ potential nonzero correlations. It follows that the fraction of the matrix of correlations that are potentially correlated is

$$(2G-3)/[G(G-1)/2] = (4-6/G)/(G-1). \quad (2.15)$$

With $G = 10$, for example, there are 45 country pairs and 38% of the entries in the 45×45 correlation matrix are potentially nonzero. Similar figures for $G = 30$ and $G = 100$ are, respectively, 13% and 4%.

For large G there are $4/(G-1)$ potential nonzero error correlations compared to $2/(G-1)$ with two-way clustering and only $2/[G(G-1)]$ if errors are uncorrelated.

2.3.1. Dyadic-Robust Variance Estimator

The first estimator for clustering extends the two-way method, by adding in correlations when $g = h'$ or $h = g'$. Then $\widehat{V}[\widehat{\beta}]$ is given in (2.6) with

$$\widehat{\mathbf{B}} = c \times \sum_{i=1}^N \sum_{j=1}^N \mathbf{1}[g = g' \text{ or } h = h' \text{ or } g = h' \text{ or } h = g'] \times \widehat{u}_{igh} \widehat{u}_{jg'h'} \mathbf{x}_{igh} \mathbf{x}_{jg'h'}', \quad (2.16)$$

where $c = [(G-1)/(G-2)]/[(N-1)/(N-k)]$ is a finite-cluster adjustment.¹ Unlike the two-way cluster-robust method, the dyadic-robust method does not appear to lend itself to a simple calculation analogous to that described after (2.12).

This method is used by Fafchamps and Gubert (2007), who motivate it as an extension of the method of Conley (1999) for spatial correlation. Fafchamps and Gubert (2007, p.330) state that “Monte Carlo simulations indicate that standard errors corrected for dyadic correlation can be much larger than uncorrected ones. The bias is particularly large when the average degree is high. Correcting standard errors is thus essential when estimating any dyadic regression. In our case, the magnitude of the correction is relatively small because the

¹The adjustment factor c for bidirectional data is explained in Section 3.1. If unidirectional data are instead used then $c = [G/(G-1)]/[(N-1)/(N-k)]$.

average degree is low.” Here degree is the number of links and in their study each individual had relatively few links.

For cross-section dyadic data there is only one observation for each dyad. When there are multiple observations per dyad, as is the case for panel data, formula (2.16) again applies.

2.3.2. Node-Jackknife Variance Estimator

The delete-one-node estimate $\hat{\beta}_{(-g)}$ is obtained by dropping country g and any pair with that country (i.e. for given g all pairs (g, h) and (h, g) for $h = 1, \dots, G$ are dropped).

Then the node-jackknife estimate of the variance matrix of $\hat{\beta}$ is

$$\hat{V}[\hat{\beta}] = \frac{G-2}{2G} \sum_{g=1}^G (\hat{\beta}_{(-g)} - \bar{\hat{\beta}})(\hat{\beta}_{(-g)} - \bar{\hat{\beta}})', \quad (2.17)$$

where $\bar{\hat{\beta}} = \frac{1}{G} \sum_{g=1}^G \hat{\beta}_{(-g)}$. This is proposed in the non-regression case by Frank and Snijders (1994) and Snijders and Borgatti (1999). The multiplier of the sum is a dyadic data variant of the multiplier $\frac{N-1}{N}$ for independent data. (This weight in turn differs from $\frac{1}{N-1}$ because each data set is similar to the other since only one observation is changed).

Frank and Snijders (1994) propose this multiplier under a particular sampling scheme and it is an open question as to how well this variance estimate will work in the setting of this paper.

2.3.3. Dyadic Random Effects Estimator

Cameron and Golotvina (2005) consider the following dyadic variant of a two-way random effects model for dyadic data. For the $(g, h)^{th}$ country-pair the regression model is

$$y_{gh} = \mathbf{x}_{gh}'\beta + \alpha_g + \alpha_h + \varepsilon_{gh}, \quad h = 1, \dots, g-1, \quad g = 1, \dots, G, \quad (2.18)$$

where α_g and α_h are country-specific error components and ε_{gh} is an idiosyncratic error component. Unlike a two-way random effects model, symmetry of α_g and α_h is imposed, so there are just G draws of α .

The variance components are assumed to be i.i.d. with

$$\begin{aligned} \varepsilon_{gh} &\sim \text{iid}[0, \sigma_\varepsilon^2] \\ \alpha_g &\sim \text{iid}[0, \sigma_\alpha^2], \end{aligned} \quad (2.19)$$

Then for error $v_{gh} = \alpha_g + \alpha_h + \varepsilon_{gh}$

$$\text{Cov}[v_{gh}, v_{g'h'}] = \begin{cases} 2\sigma_\alpha^2 + \sigma_\varepsilon^2 & g = h, g' = h' \\ \sigma_\alpha^2 & g = g', h \neq h', g \neq g', h = h' \\ 0 & g \neq g', h \neq h'. \end{cases} \quad (2.20)$$

Cameron and Golotvina (2005) compute standard errors for the OLS estimator assuming the d.g.p. is that given in (2.18)-(2.20). They also present a feasible GLS estimator given this model. And they provide a default OLS variance inflation factor in the case of an intercept-only model.

2.4. Dyadic-robust Inference with Few Countries

The various cluster-robust variance matrix estimates rely on asymptotic theory that assumes that the number of clusters goes to infinity. For one-way clustering, finite-cluster modifications of (2.5) are typically used, since without modification the cluster-robust standard errors are biased downwards. Cameron, Gelbach, and Miller (2008) reviewed various small-sample corrections that have been proposed in the literature, for both standard errors and for inference using resultant Wald statistics. For example, Stata uses the multiplier c defined after (2.5) and uses $t(G - 1)$ critical values rather than standard normal critical values. Improved finite-cluster inference with one-way clustering is an active area of research; Cameron and Miller (2015) provide a summary.

For two-way cluster-robust inference, Cameron, Gelbach and Miller (2011, Table 1) found even larger over-rejection rates than in the one-way case when there are few clusters. Cameron, Gelbach and Miller (2008) and Imbens and Kolesar (2012) have proposed few-cluster methods for the one-way case that do not readily extend to two-way clustering. At a minimum one should use a variance matrix estimate with finite-cluster correction such as that in (2.12) and use the Student's t -distribution with $\min(G, H) - 1$ degrees of freedom.

For dyadic error correlation we expect the finite-cluster problem to be even greater still, as formula (2.12) introduces additional terms in the computation of $\hat{\mathbf{B}}$. Additionally, the simulations and applications in this paper use bidirectional data. Then clusters are unbalanced, and recent work by Carter, Schnepel and Steigerwald (2013) and MacKinnon and Webb (2013) finds that in the one-way case the few cluster problem becomes more pronounced when clusters are unbalanced. Finite-cluster performance is investigated in the Monte Carlos of Section 3.

Another practical matter is that the dyadic robust estimator $\widehat{V}[\widehat{\beta}]$ may not be positive-semidefinite, especially when G is small. A similar problem arises in the two-way cluster-robust case. A positive-semidefinite matrix can be created by employing a technique used in the time series HAC literature, such as in Politis (2011). This uses the eigendecomposition of the estimated variance matrix and converts any negative eigenvalue(s) to zero. Specifically, decompose the variance matrix into the product of its eigenvectors and eigenvalues: $\widehat{V}[\widehat{\beta}] = U\Lambda U'$, with U containing the eigenvectors of \widehat{V} , and $\Lambda = \text{Diag}[\lambda_1, \dots, \lambda_d]$ containing the eigenvalues of \widehat{V} . Then create $\Lambda^+ = \text{Diag}[\lambda_1^+, \dots, \lambda_d^+]$, with $\lambda_j^+ = \max(0, \lambda_j)$, and use $\widehat{V}^+[\widehat{\beta}] = U\Lambda^+U'$ as the variance estimate.

2.5. Dyadic Clustering for m-estimators and GMM Estimators

The preceding analysis considered the OLS estimator. More generally we can consider dyadic clustering for other regression estimators commonly used in econometrics. The results for the dyadic-robust variance estimator are qualitatively the same as for OLS.

We begin with an m-estimator that solves $\sum_{i=1}^N \mathbf{h}_i(\widehat{\theta}) = \mathbf{0}$. Examples include nonlinear least squares estimation, maximum likelihood estimation, and instrumental variables estimation in the just-identified case. For the logit MLE $\mathbf{h}_i(\beta) = (y_i - \Lambda(\mathbf{x}_i'\beta))\mathbf{x}_i$, where $\Lambda(\cdot)$ is the logistic c.d.f.

Under standard assumptions, $\widehat{\theta}$ is asymptotically normal with estimated variance matrix

$$\widehat{V}[\widehat{\theta}] = \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}'^{-1}, \quad (2.21)$$

where $\widehat{\mathbf{A}} = \sum_i \frac{\partial \mathbf{h}_i}{\partial \theta'} \big|_{\widehat{\theta}}$, and $\widehat{\mathbf{B}}$ is an estimate of $V[\sum_i \mathbf{h}_i]$.

For one-way clustering $\widehat{\mathbf{B}} = \sum_{g=1}^G \widehat{\mathbf{h}}_g \widehat{\mathbf{h}}_g'$ where $\widehat{\mathbf{h}}_g = \sum_{i=1}^{N_g} \widehat{\mathbf{h}}_{ig}$. Clustering may or may not lead to parameter inconsistency, depending on whether $E[\mathbf{h}_i(\theta)] = \mathbf{0}$ in the presence of clustering. As an example consider a probit model with one-way clustering. One approach, called a population-averaged approach in the statistics literature, is to assume that $E[y_{ig}|\mathbf{x}_{ig}] = \Phi(\mathbf{x}_{ig}'\beta)$, even in the presence of clustering. An alternative approach is a random effects approach. Let $y_{ig} = 1$ if $y_{ig}^* > 0$ where $y_{ig}^* = \mathbf{x}_{ig}'\beta + \varepsilon_g + \varepsilon_{ig}$, the idiosyncratic error $\varepsilon_{ig} \sim \mathcal{N}[0, 1]$ as usual, and the cluster-specific error $\varepsilon_g \sim \mathcal{N}[0, \sigma_g^2]$. Then it can be shown that $E[y_{ig}|\mathbf{x}_{ig}] = \Phi(\mathbf{x}_{ig}'\beta/\sqrt{1 + \sigma_g^2})$, so that the moment condition is no longer $E[y_{ig}|\mathbf{x}_{ig}] = \Phi(\mathbf{x}_{ig}'\beta)$. When $E[\mathbf{h}_i(\theta)] \neq \mathbf{0}$ the estimated variance matrix is still that in (2.21), but the distribution of the estimator will be instead centered on a pseudo-true value (White,

1982). Note that for the probit model the average partial effect is nonetheless consistently estimated (Wooldridge, 2002, p. 471).

Our concern is with dyadic clustering. The analysis of the preceding section carries through, with $\hat{u}_{gh}\mathbf{x}_{gh}$ in (2.16) replaced by $\hat{\mathbf{h}}_i$. Then $\hat{\boldsymbol{\theta}}$ is asymptotically normal with estimated variance matrix $\hat{V}[\hat{\boldsymbol{\theta}}] = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}'^{-1}$, with $\hat{\mathbf{A}}$ defined after (2.21) and $\hat{\mathbf{B}}$ defined as in (2.16) with $\hat{u}_{igh}\mathbf{x}_{igh}$ replaced by $\hat{\mathbf{h}}_{igh}$.

Finally we consider GMM estimation for over-identified models. A leading example is linear two stage least squares with more instruments than endogenous regressors. Then $\hat{\boldsymbol{\theta}}$ minimizes $Q(\boldsymbol{\theta}) = \left(\sum_{i=1}^N \mathbf{h}_i(\boldsymbol{\theta})\right)' \mathbf{W} \left(\sum_{i=1}^N \mathbf{h}_i(\boldsymbol{\theta})\right)$, where \mathbf{W} is a symmetric positive definite weighting matrix. Under standard regularity conditions $\hat{\boldsymbol{\theta}}$ is asymptotically normal with estimated variance matrix given dyadic clustering

$$\hat{V}[\hat{\boldsymbol{\theta}}] = \left(\hat{\mathbf{A}}'\mathbf{W}\hat{\mathbf{A}}\right)^{-1} \hat{\mathbf{A}}'\mathbf{W}\tilde{\mathbf{B}}\mathbf{W}\hat{\mathbf{A}} \left(\hat{\mathbf{A}}'\mathbf{W}\hat{\mathbf{A}}\right)^{-1}, \quad (2.22)$$

where $\hat{\mathbf{A}} = \sum_i \frac{\partial \mathbf{h}_i}{\partial \boldsymbol{\theta}'} \big|_{\hat{\boldsymbol{\theta}}}$, and $\tilde{\mathbf{B}}$ is an estimate of $V[\sum_i \mathbf{h}_i]$ that can be computed as in (2.16) with $\hat{u}_{igh}\mathbf{x}_{igh}$ replaced by $\hat{\mathbf{h}}_{igh}$.

3. Monte Carlo Exercises

We consider two Monte-Carlo exercises for OLS regression with cross-section dyadic data with a bidirectional relationship and no relationship with oneself. An example is bidirectional trade flow data in a single year, and we use that terminology with a dyad being a country-pair and the two components of the dyad being country 1 and country 2.

3.1. Monte Carlo Setup

In both cases the data generating process is of the form

$$y_{gh} = \beta_1 + \beta_2 x_{gh} + u_{gh}, \quad h = g + 1, \dots, G, \quad g = 1, \dots, G - 1. \quad (3.1)$$

There are $N = G(G - 1)/2$ observations and $k = 2$ regression parameters. In the first Monte Carlo the error u_{gh} is i.i.d. and in the second Monte Carlo there is dyadic correlation due to country-specific random effects.

The regressor x_{gh} is constructed to be similar to a log-distance measure in a gravity model of trade. Specifically, let (z_{1g}, z_{2g}) denote the coordinates of country g , where z_{1g} and z_{2g} are i.i.d. draws from the uniform distribution. Then $x_{gh} = \ln(\sqrt{(z_{1g} - z_{1h})^2 + (z_{2g} - z_{2h})^2})$.

The parameters are estimated by OLS regression of y_{gh} on an intercept and x_{gh} . Standard errors for the OLS coefficients are computed in the following ways

1. IID: OLS default assuming i.i.d. errors
2. HETROB: heteroskedastic-robust (same as PAIRS: one-way cluster-robust clustering on country-pair (g, h))
3. CTRY1: one-way cluster-robust with clustering on country 1 (g)
4. TWOWAY: two-way cluster-robust clustering on countries 1 and 2 (g, h)
5. DYADS: dyadic-robust
6. NJACK: leave-one-node-out jackknife.

The first three methods use Stata command regress to compute standard errors. So methods 1-2 use the usual finite sample adjustment factor of $N/(N - k)$.

For this d.g.p. with bidirectional data there are only $G - 1$ country 1 clusters as there are no (G, h) pairs (and similarly there are only $G - 1$ country 2 clusters as there are no $(g, 1)$ pairs). As a result, the finite sample adjustments given after (2.5) and after (2.11) need to be modified. Define $G^* = G - 1$. Then method 3 uses finite-sample adjustment factor $c = G^*/(G^* - 1) \times [N/(N - k)]$. Method 4 inflates the three components in (2.11) by, respectively, $c_1 = G^*/(G^* - 1) \times [N/(N - k)]$, $c_2 = c_1$, and $c_3 = N/(N - k)$. Method 5 again inflates by $G^*/(G^* - 1) \times [N/(N - k)]$. Method 6 is computed as in (2.17).

Additionally a two-sided five percent significance test for β_2 is performed. The critical values used are from the $t(N - k)$ distribution for methods 1-2 and from the $t(G^* - 1)$ distribution for methods 3-6, with $G^* = G - 1$ for the d.g.p.'s used here.

We report results for $G = 100, 30$, and 10 . This corresponds to sample sizes of, respectively, 4950, 435 and 45 dyads.

There were 4,000 simulations, so the 95% simulation interval for a test with true size 0.05 is (0.043, 0.057).

3.2. Independent and Identically Distributed Errors

Here in model (3.1) the error u_{gh} is i.i.d. $\mathcal{N}[0, 1]$, and $\beta_1 = \beta_2 = 0$. In this case as $G \rightarrow \infty$ all six standard error estimates should be correct and tests of $\beta_2 = 0$ at 5% should have actual rejection rates of 5%. Results are reported in Table 1.

For $G = 100$, the various standard error estimates are all close to the standard deviation of $\hat{\beta}_2$ across simulations of 0.0226. The cost of using more robust methods, unnecessary for this d.g.p., is increased variability in the standard error estimate. For example the standard deviations of standard errors computed assuming one-way clustering and dyadic clustering equal, respectively, 0.0020 and 0.0033, compared to 0.0005 for the default. At the same time even 0.0033 is small relative to the true standard deviation of $\hat{\beta}_2$ of 0.0226.

Turning to test size, with $G = 100$, the dyadic-robust standard errors do lead to mild over-rejection with a rejection rate of 0.064. The other methods all lead to rejection rates within the 95% simulation interval of (0.043, 0.057).

For $G = 30$, the various robust standard errors all under-estimate the standard deviation of $\hat{\beta}_2$ across simulations of 0.0773. The under-estimation increases the more “robust” the method used. In particular, the dyadic-robust standard error has a standard deviation of 0.0697, a substantial under-estimation of 10%.

This under-estimation of the standard error leads to test over-rejection. For $G = 30$ the dyadic-robust method has rejection rate of 0.117, while the two-way robust method, for example, has rejection rate of 0.078.

For $G = 10$, the results are qualitatively similar to those with $G = 30$, though with greater under-estimation of standard errors and greater test size distortion. This poor performance is not surprising. There are only 45 observations. The dyadic-robust method then permits 38% of the terms in the 45×45 error covariance matrix to be nonzero, and even the one-way cluster robust method allows 14% of the terms to be nonzero. The dyadic-robust method in some cases leads to a non-positive definite estimated variance matrix of $\hat{\beta}$, necessitating the eigendecomposition adjustment presented in section 2.4, and in several other cases the estimated standard error of $\hat{\beta}_2$ was essentially equal to zero.

Finally, note that the node jackknife works very well for $G = 10$, $G = 30$ and $G = 100$, with average standard error very close to the simulation standard deviation of $\hat{\beta}_2$ and test size very close to 0.05.

In summary, the dyadic-robust variance matrix estimate works well when $G = 100$, but works poorly when $G = 30$ or $G = 10$.

3.3. Spatially Correlated Errors

Now let the error in model (3.1) be generated by a spatially correlated process, with

$$u_{gh} = \alpha_g + \alpha_h + 0.25 \times \varepsilon_{gh}$$

where α_g is i.i.d. uniform and ε_{gh} is i.i.d. $\mathcal{N}[0, 1]$. The coefficients $\beta_1 = 8$ and $\beta_2 = -1$. For this dyadic correlated error d.g.p., introduced in section 2.3.3, the error correlation for dyads that have a country in common is 0.447, since then (2.20) yields $\text{Cor}^2[u_{gh}, u_{g'h'}] = (1/12)/[(2/12) + 0.25] = 0.2$.

In this case as $G \rightarrow \infty$ the dyadic-robust standard error estimates should be correct and the test of $\beta_2 = -1$ at 5% should have actual rejection rates of 5%. The remaining methods, with the exception of the node jackknife, are expected to under-estimate standard errors and lead to tests that over-reject. Results are reported in Table 2.

For $G = 100$, the dyadic-robust standard errors are closest on average to the standard deviation of $\hat{\beta}_2$ across simulations of 0.0248, and the rejection rate of 0.057 is close to 0.05. By comparison the other methods greatly under-estimate the standard errors and lead to large test over-rejection. As expected, the performance of the other methods improves the greater the “robustness” of the variance estimation method, with dyadic best, followed in order by two-way cluster, one-way cluster and non-clustered methods. Furthermore the differences across the methods are substantial.

For $G = 30$ there is a similar ordering of the performance of the methods. Now, however, even the dyadic-robust method suffers from considerable under-estimation of the standard deviation of $\hat{\beta}_2$ (0.0458 compared to 0.0525) and test rejection rate of 0.095.

For $G = 10$ all methods under-estimate standard errors and lead to test over-rejection. Surprisingly the various cluster-robust methods perform even worse than methods assuming i.i.d. errors or independent heteroskedastic errors. Clearly $G = 10$ is too low for the asymptotic theory to kick in.

Finally, note that the node jackknife works very well for $G = 10$, and reasonably well for $G = 30$. For $G = 100$ it is still a substantial improvement on using one-way cluster-robust standard errors, but does not perform as well as either two-way or dyad-robust methods.

In summary, the dyadic-robust variance matrix estimate works well when $G = 100$, works better than the other methods when $G = 30$, but works poorly when $G = 10$.

4. Empirical Examples

We consider two examples - one cross-section and one panel. Both examples use data obtained from Andrew [Rose's](#) website. We commend Andrew Rose for making the data and Stata programs for his many papers readily available to other researchers. It should be clear

that the methods used in his papers are the standard methods used in the international trade literature at the time the papers were published. For example, in the panel paper we study, Rose (2004, p.98) states that “To make my argument as persuasive as possible I use widely accepted techniques, a conventional empirical methodology, and standard data sets.”

4.1. Cross-section Example

Table 3A and 3B

We consider a cross-section application with dependent variable the natural logarithm of bilateral trade (in US\$), replicating column 1 of Table 3 of Rose and Engel (2002). Then there are $G = 127$ countries and data are available for $N = 4618$ of the potential $8,001$ country pairs (equals $127 \times 126/2$).

The model estimated by OLS is

$$y_{gh} = \mathbf{x}'_{gh}\boldsymbol{\beta} + u_{gh}, \quad (4.1)$$

with standard errors computed using the six methods detailed in Section 3.1, except now the one-way clustered standard errors are computed in two different ways, by clustering on country 1 (as before) and by clustering on country 2.

The results are given in Table 3A. The heteroskedastic-robust standard errors equal those reported in Rose and Engel (2002). In this cross-section data example they are equivalent to those obtained by clustering on country pairs. Allowing for increasing amounts of clustering leads to progressively larger standard errors. For example, for the log distance regressor the heteroskedastic-robust standard error is 0.0349, the one-way cluster-robust standard error is 0.0646 or 0.0912, depending on whether one clusters on country 1 or country 2, the two-way cluster-robust standard error is 0.1062 and the dyad-robust standard error is 0.1215. Similar increases occur for the log product regressors. In particular, the dyad-robust standard errors are roughly 3.5 times the heteroskedastic robust standard errors. The standard error magnification is smaller for the standard error of the currency union regressor, the key regressor in the study of Rose and Engel (2002). Note that this binary regressor is non-zero for only 24 of the 4,618 observations.²

Some of the correlation in errors may be absorbed by the inclusion of country fixed effects.

²We use exactly the same data as used by Rose and Engel (2002). In fact for the OLS regression reported here, and in Table 3 of Rose and Engel (2002), three country pairs appear twice in the dataset, and all these duplicates were in a currency union. If these duplicates are dropped then the currency union coefficient falls to 0.9769, still very large, and remains statistically significant at the 5% level.

We estimate the model

$$y_{gh} = \alpha_g + \delta_h + \mathbf{x}_{gh}'\boldsymbol{\beta} + u_{gh}, \quad (4.2)$$

where α_g is a fixed effect for the first country in the pair and δ_h is a fixed effect for the second country in the pair.

There are several ways to implement this fixed effects regression. These lead to the same coefficient estimates but different standard error estimates due to different finite sample degrees of freedom adjustments. The fastest method is to Frisch-Waugh out the fixed effects and perform OLS regression on the residuals. Thus OLS regress y_{gh} on a full set of country 1 and country 2 dummy variables (in Stata `reg y i.cty1 i.cty2`) and save the residuals as r_y_{gh} . Similarly OLS regress each component of \mathbf{x}_{gh} on a full set of country 1 and country 2 dummy variables, leading to residual vector $\mathbf{r_x}_{gh}$. Then OLS regress r_y_{gh} on $\mathbf{r_x}_{gh}$ and compute the various standard errors as before. Let $k_{\mathbf{x}}$ denote the number of regression parameters other than the fixed effects (here $k_{\mathbf{x}} = 2$). Then finite-sample adjustment factors that include $N - k$ will use $N - k_{\mathbf{x}}$, whereas direct OLS estimation of (4.2) with the $2(G^* - 1)$ country dummies uses $N - (k_{\mathbf{x}} + 2(G - 1))$. The latter method will lead to a larger adjustment factor and hence larger standard errors.

Table 3B presents results when country 1 fixed effects and country 2 fixed effects are included. The only identified coefficients are those for currency union and log distance, since the coefficients of regressors that are invariant within country 1 or invariant within country 2 are not identified. There is actually an efficiency improvement, due in part to the root mean square error (RMSE) falling from 1.75 to 1.35. But the dyadic-robust standard errors remain much larger than the heteroskedastic-robust standard errors, being 2.84 times larger for log distance and 1.55 times larger for currency union.

Clearly the inclusion of country-specific effects does not account for all the error correlation. It is still necessary to control for dyadic error correlation. The two-way cluster-robust and node-jackknife methods go a long way towards doing so. But the dyad-robust standard errors are still larger – for regressors other than currency union roughly 10-15% larger than two-way cluster-robust and 30% larger than node jackknife.

Table 3C presents feasible GLS estimates of a dyadic random effects variant of the model (4.2), under the strong assumption that $\delta_g = \alpha_g$ and that α_g and u_{gh} are i.i.d. errors. This is the dyadic random effects error correlation model of Section 2.3.3. The FGLS coefficients for currency union and log distance lie between the OLS and fixed effects coefficients, while the other two regressors have coefficients similar to those obtained without either fixed or

random country effects. There appears to be an efficiency gain in using the random effects estimator, but this may be illusory as the standard errors given in the last column of Table 3C are only valid if α_g and u_{gh} are i.i.d. Indeed, for the coefficients of currency union and log distance, these standard errors are close to the i.i.d. standard errors for the FE estimator, rather than to the larger dyadic-robust standard errors for the FE estimator.

4.2. Panel Example

We apply various standard error estimators to data from Rose (2004), replicating column 1 of his Table 1. This study uses annual data on real bilateral trade flows between countries g and h . There are 187 countries and 52 years (1948-1999) of data. In principal there are as many as 819,156 observations (equals $52 \times 178 \times 177/2$). The data analyzed here have 234,597 observations.

The model is

$$y_{ght} = \mathbf{x}'_{ght}\boldsymbol{\beta} + u_{ght}, \quad (4.3)$$

where y_{ght} is the natural logarithm of real bilateral trade trade between countries g and h in time period t . Some regressors vary over time while others do not. The regressors include a full set of time dummies.

Table 4A presents results for the model when no fixed effects are included. The coefficients and cluster-robust standard errors, obtained by one-way clustering on country pair, equal those given in Column 1 of Table 1 of Rose (2004). These country-pair cluster-robust standard errors are 3-3.5 times the heteroskedastic-robust standard errors. While they allow for error correlation over time, they do not pick up all potential error correlations as, for example, it is assumed that the error for a (1,2) pair is independent of the error for a (1,3) pair. Allowing for more general clustering, either one-way on country 1 or country 2, or two-way on country 1 and 2 leads to much larger standard errors. The preferred dyadic-robust standard errors are roughly 10%-25% larger than the node-jackknife standard errors, 10%-25%, and 2-4 times larger than cluster-robust standard errors that cluster on country pair.

Clearly it is important to control properly for the error correlation. Rose (2004) focused on the first two regressors, respectively, a binary indicator for whether both g and h are GATT/WTO members and a binary indicator for whether either g and h are GATT/WTO members. When dyadic-robust standard errors are used the key conclusion of Rose (2004),

that these two regressors are statistically insignificant remains unchanged, though 95% confidence intervals for these two slope parameters will be much broader.

Some of the dyadic correlation in errors may be absorbed by the inclusion of country fixed effects. There are various ways that country fixed effects may be included in the panel case. As in the cross-section case one can include country 1 fixed effects and country 2 fixed effects. Here we consider a richer model, with country-pair fixed effects. The model is now

$$y_{ght} = \alpha_{gh} + \mathbf{x}'_{ght}\boldsymbol{\beta} + u_{ght}, \quad (4.4)$$

where α_{gh} are country-pair fixed effects.

In the current example there are potentially as many as $178 \times 177/2 = 15,753$ country-pair fixed effects. Now we Frisch-Waugh out both the country-pair fixed effects and the time dummies, and perform OLS regression on the residuals. Thus perform fixed effects estimation of y_{ght} on the time dummy variables (in Stata `xtreg y d*, fe (i.ctypair)` where `d*` denotes the time dummies). Similarly perform fixed effects estimation of each component of \mathbf{x}_{ght} (other than the time dummies) on the time dummy variables, leading to residual vector $\mathbf{r}\mathbf{x}_{ght}^*$. Then OLS regress $r\text{-}y_{gh}$ on $\mathbf{r}\mathbf{x}_{ght}^*$. As in the cross-section case this will lead to smaller finite-sample adjustment factor $N - k$ than if (4.4) is directly estimated by OLS.

Table 4B presents results when country-pair fixed effects are included. Only eight of the seventeen regressors have coefficients that are identified, as the remaining regressors are time-invariant (i.e. $x_{ght} = x_{gh}$ for all t). The dyadic-robust standard errors are now roughly 10%-30% larger than the node-jackknife standard errors, 10%-25% larger than two-way cluster-robust standard errors, and 2-3 times larger than cluster-robust standard errors that cluster on country pair.

Even with country-pair fixed effects included, the country-pair cluster-robust standard errors greatly under-estimate the standard error and overstate the estimator precision. When dyadic-robust standard errors are used the key conclusion of Rose (2004), that the first two regressors are statistically insignificant remains unchanged, though 95% confidence intervals for these two slope parameters will be much broader.³

³Our fixed effects estimates differ from those given in column 4 of Table 1 of Rose (2004). The key first two regressors have coefficients of 0.15 and 0.05, compared to our 0.13 and 0.06. Other coefficients differ more substantially and, surprisingly, Rose(2004) reports nonzero coefficients for the regressors that are not identified even though he states that his fixed effects estimates use country-pair fixed effects.

5. Conclusion

Failure to control properly for error correlation in models with country-pair data can lead to greatly under-estimated standard errors and over-stated t-statistics. In the two empirical examples dyadic-robust standard errors were often several times larger than country-pair cluster-robust standard errors, even after inclusion of country fixed effects. More generally such a large difference in reported standard errors may arise with dyadic data when each individual is paired with many other individuals, so that the network is a dense network.

It is well-known that one-way and two-way cluster-robust standard errors lead to standard Wald tests that over-reject when there are few clusters. Similar problems exist for dyadic-robust standard errors when there are few underlying individuals forming the dyads (in our notation when the number of countries G is small even though the number of observed dyads N may be large). Our Monte Carlos suggest that with $G = 100$ there is no problem, but with fewer countries there can be considerable over-rejection.

References

- Arellano, M. (1987), "Computing Robust Standard Errors for Within-Group Estimators," *Oxford Bulletin of Economics and Statistics*, 49, 431-434.
- Bertrand, M., E. Duflo, and S. Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 119, 249-275.
- Cameron, A.C., Gelbach, J.G., and D.L. Miller (2008), "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics* 90, 414-427.
- Cameron, A.C., Gelbach, J.G., and D.L. Miller (2011), "Robust Inference with Multi-way Clustering," *Journal of Business and Economic Statistics*, 29, 238-249.
- Cameron, A.C., and N. Golotvina (2005), "Estimation of Country-Pair Data Models Controlling for Clustered Errors: with International Trade Applications," U.C.-Davis Economics Department Working Paper No. 06-13.
- Cameron, A.C., and D.L. Miller (2011), "Robust Inference with Clustered Data" (with Douglas Miller), A. Ullah and D. E. Giles eds., *Handbook of Empirical Economics and Finance*, 2011, 1-28, CRC Press.
- Cameron, A.C., and D.L. Miller (2015), "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, forthcoming.
- Carter, A.V., K.T. Schnepel, and D.G. Steigerwald (2013), "Asymptotic Behavior of a t Test Robust to Cluster Heterogeneity," University of California - Santa Barbara.
- Conley, T.G. (1999), "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92, 1-45.
- Fafchamps, M., and F. Gubert (2007), "The Formation of Risk Sharing Networks," *Journal of Development Economics*, 83, 326-350.
- Frank, O. and T.A.B. Snijders (1994), "Estimating the Size of Hidden Populations under Snowball Sampling," *Journal of Official Statistics*, 10, 53-67.
- Hansen, C. (2007), "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large," *Journal of Econometrics*, 141, 597-620.

- Imbens, G.W., and M. Kolesar (2012), “Robust Standard Errors in Small Samples: Some Practical Advice,” NBER Working Paper 18478.
- Liang, K.-Y., and S.L. Zeger (1986), “Longitudinal Data Analysis Using Generalized Linear Models,” *Biometrika*, 73, 13-22.
- MacKinnon, J., and M.D. Webb (2013), “Wild Bootstrap Inference for Wildly Different Cluster Sizes,” Queens Economics Department Working Paper No. 1314.
- Miglioretti, D.L., and P.J. Heagerty (2006), “Marginal Modeling of Nonnested Multilevel Data using Standard Software,” *American Journal of Epidemiology*, 165(4), 453-463.
- Moulton, B.R. (1986), “Random Group Effects and the Precision of Regression Estimates,” *Journal of Econometrics*, 32, 385-397.
- 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.
- Politis, D.N. (2011), “Higher-order Accurate, Positive Semi-definite Estimation of Large-sample Covariance and Spectral Density Matrices,” *Econometric Theory*, 27, 704-744.
- Rogers, W.H. (1993), “Regression Standard Errors in Clustered Samples,” *Stata Technical Bulletin*, 13, 19-23.
- Rose, A.K.(2004), “Do We Really Know That the WTO Increases Trade?” *American Economic Review*, 94, 98-114.
- Rose, A.K. and C. Engel (2002), “Currency Unions and International Integration,” *Journal of Money, Credit and Banking*, 34(4), 1067-1089.
- Shah, Bbabubhai V., M. M. Holt and Ralph E. Folsom (1977), “Inference About Regression Models from Sample Survey Data.” *Bulletin of the International Statistical Institute Proceedings of the 41st Session*, 47(3), 43-57.
- Snijders, T.A.B., and S.B. Borgatti (1999), “Non-Parametric Standard Errors and Tests for Network,” *Connections*, 22, 161-70.
- Thompson, S. (2011), “Simple Formulas for Standard Errors that Cluster by Both Firm and Time,” *Journal of Financial Economics*, 99, 1-10.
- White, H. (1980), “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a

Direct Test for Heteroskedasticity,” *Econometrica*, 48, 817-838.

White, H. (1984), *Asymptotic Theory for Econometricians*, San Diego, Academic Press.

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

Wooldridge, J.M. (2003), “Cluster-Sample Methods in Applied Econometrics,” *American Economic Review*, 93, 133-138.

TABLE 1: SIMULATION with IID ERRORSCross-section OLS of y_{gh} on intercept and scalar x_{gh}

Number of "Countries"	100		30		10	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
COEFF	0.0008	0.0226	-0.0016	0.0773	-0.0030	0.2598
SE_IID	0.0227	0.0005	0.0773	0.0049	0.2510	0.0475
SE_HETROB	0.0227	0.0007	0.0768	0.0072	0.2370	0.0628
SE_CTRY1	0.0226	0.0020	0.0755	0.0133	0.2270	0.0847
SE_TWOWAY	0.0225	0.0027	0.0740	0.0179	0.2173	0.1020
SE_DYAD	0.0221	0.0033	0.0697	0.0220	0.1890	0.1106
SE_NJACK	0.0227	0.0017	0.0772	0.0123	0.2596	0.0903
REJ_IID	0.050		0.045		0.049	
REJ_HETROB	0.050		0.050		0.079	
REJ_CTRY1	0.050		0.059		0.087	
REJ_TWOWAY	0.053		0.078		0.142	
REJ_DYAD	0.064		0.117		0.219	
REJ_NJACK	0.048		0.051		0.052	
Sample size	N = 4950		N = 435		N = 45	

4,000 Monte Carlo simulations

COEFF is the fitted slope coefficient (D.g.p. value is 0).

SE_IID is default standard errors assuming i.i.d. errors

SE_HETROB = SE_PAIRS is heteroskedastic robust standard error

SE_CTRY1 is one-way cluster robust standard error with clustering on country 1 (g)

SE_TWOWAY is two-way cluster robust standard error (g and h)

SE_DYAD is dyadic cluster-robust standard error

SE_NJACK is node-jackknife cluster-robust standard error

REJ_ is rejection rate for two-sided test that $\beta = 0$ at 5%Critical values use $t(N-2)$ for IID and ROBUST, and $t(G^*-1) = t(G-2)$ for the remainder.

TABLE 2: SIMULATION with SPATIALLY CORRELATED ERRORS

Cross-section OLS of y_{gh} on intercept and scalar x_{gh}

Number of "Countries"	100		30		10	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
COEFF	-1.0001	0.0248	-0.9991	0.0525	-0.9998	0.1290
SE_IID	0.0108	0.0004	0.0360	0.0031	0.1116	0.0225
SE_HETROB	0.0108	0.0005	0.0359	0.0039	0.1059	0.0283
SE_CTRY1	0.0164	0.0022	0.0400	0.0084	0.1014	0.0373
SE_TWOWAY	0.0205	0.0027	0.0435	0.0111	0.0972	0.0453
SE_DYAD	0.0238	0.0033	0.0458	0.0138	0.0854	0.0483
SE_NJACK	0.0190	0.0023	0.0446	0.0089	0.1219	0.0428
REJ_IID	0.402		0.174		0.092	
REJ_HETROB	0.402		0.172		0.119	
REJ_CTRY1	0.191		0.123		0.116	
REJ_TWOWAY	0.099		0.106		0.161	
REJ_DYAD	0.057		0.095		0.230	
REJ_NJACK	0.128		0.081		0.060	

4,000 Monte Carlo simulations

COEFF is the fitted slope coefficient (D.g.p. value is -1).

SE_IID is default standard errors assuming i.i.d. errors

SE_HETROB = SE_PAIRS is heteroskedastic robust standard error

SE_CTRY1 is one-way cluster robust standard error with clustering on country 1 (g)

SE_TWOWAY is two-way cluster robust standard error (g and h)

SE_DYAD is dyadic cluster-robust standard error

SE_NJACK is node-jackknife cluster-robust standard error

REJ_ is rejection rate for two-sided test that $\beta = -1$ at 5%

Critical values use $t(N-2)$ for IID and ROBUST, and $t(G^*-1) = t(G-2)$ for the remainder.

0.872381

TABLE 3A: COUNTRY-PAIR CROSS-SECTION DATA EXAMPLE: NO FIXED EFFECTS
Comparison of Standard Errors computed in various ways

	COEFF	g and h	IID	HETROB	ST. ERROR				SE RATIOS					
					CTRY1	CTRY2	TWOWAY	DYAD	NJACK	SIG 5%	DYAD /HET	DYAD /CTRY1	DYAD /2WAY	DYAD /JACK
Currency Union	1.8588	Yes	0.3666	0.4556	0.4269	0.7764	0.7119	0.6888	0.8035	Yes	1.51	1.61	0.97	0.86
Log_Distance	-1.3667	Yes	0.0345	0.0349	0.0646	0.0912	0.1062	0.1215	0.092	Yes	3.48	1.88	1.14	1.32
Log_product_real_GDP	0.7656		0.0198	0.0204	0.0427	0.0540	0.0657	0.0758	0.0582	Yes	3.72	1.78	1.15	1.30
Log_product_real_GDP_pc	0.867		0.0116	0.0128	0.0353	0.0274	0.0428	0.0469	0.036	Yes	3.66	1.33	1.10	1.30
Constant	-34.8421		0.6004	0.6827	1.6821	1.2356	1.9713	2.1609	1.6647	Yes	3.17	1.28	1.10	1.30
Country 1 and 2 dummies	No													
Observations	4618													
RMSE	1.75													
R-squared	0.708													

g and h is YES if regressor varies over (g,h) rather than just g or just h

IID is default standard errors assuming i.i.d. errors

HETROB = PAIRS is heteroskedastic robust standard error

CTRY1 is one-way cluster robust standard error with clustering on country 1 (g)

CTRY2 is one-way cluster robust standard error with clustering on country 1 (h)

TWOWAY is two-way cluster robust standard error (g and h)

DYAD is dyadic cluster-robust standard error

NJACK is node-jackknife cluster-robust standard error

SIG 5% is Yes if statistically different from 5% based on dyadic cluster-robust standard errors

Critical values use $t(N-2)$ for IID and ROBUST, and $t(G^*-1) = t(G-2)$ for the remainder.

DYAD/HET is ratio of dyadic-robust and heteroskedastic-robust standard errors

DYAD/CTRY1 is ratio of dyadic-robust and one-way cluster-robust on country 1 standard errors

DYAD/2WAY is ratio of dyadic-robust and two-way cluster-robust robust standard errors

DYAD/JACK is ratio of dyadic-robust and node-jackknife standard errors

TABLE 3B: COUNTRY-PAIR CROSS-SECTION DATA EXAMPLE: COUNTRY 1 AND 2 FIXED EFFECTS
Comparison of Standard Errors computed in various ways

	COEFF	g and h	ST. ERROR							SE RATIOS				
			IID	HETROB	CTRY1	CTRY2	TWOWAY	DYAD	NJACK	SIG 5%	DYAD /HET	DYAD /CTRY1	DYAD /2WAY	DYAD /JACK
Currency Union	0.8952	Yes	0.3377	0.3597	0.4227	0.4825	0.5284	0.559	0.5089	No	1.55	1.32	1.06	1.10
Log_Distance	-1.6939	Yes	0.0386	0.0421	0.0696	0.0903	0.1059	0.1196	0.0910	Yes	2.84	1.72	1.13	1.31
Log_product_real_GDP	--		--	--	--	--	--	--	--	--	--	--	--	--
Log_product_real_GDP_pc	--		--	--	--	--	--	--	--	--	--	--	--	--
Constant	0.0000		0.0198	0.0198	0	0	0.0008	0.0007	0	--	--	--	--	--
Country 1 and 2 dummies	Yes													
Observations	4618													
RMSE	1.35													
R-squared	0.828													

g and h is YES if regressor varies over (g,h) rather than just g or just h

IID is default standard errors assuming i.i.d. errors

HETROB is heteroskedastic robust standard error

CTRY1 is one-way cluster robust standard error with clustering on country 1 (g)

CTRY2 is one-way cluster robust standard error with clustering on country 1 (h)

TWOWAY is two-way cluster robust standard error (g and h)

DYAD is dyadic cluster-robust standard error

NJACK is node-jackknife cluster-robust standard error

SIG 5% is Yes if statistically different from 5% based on dyadic cluster-robust standard errors

Critical values use $t(N-2)$ for IID and ROBUST, and $t(G^*-1) = t(G-2)$ for the remainder.

DYAD/HET is ratio of dyadic-robust and heteroskedastic-robust standard errors

DYAD/CTRY1 is ratio of dyadic-robust and one-way cluster-robust on country 1 standard errors

DYAD/2WAY is ratio of dyadic-robust and two-way cluster-robust robust standard errors

DYAD/JACK is ratio of dyadic-robust and node-jackknife standard errors

TABLE 3C: COUNTRY-PAIR CROSS_SECTION DATA EXAMPLE: RANDOM EFFECTS
Comparison of Standard Errors computed in various ways

	NO COUNTRY EFFECTS				FIXED EFFECTS			RANDOM EFFECTS		
	COEFF		STANDARD ERROR		COEFF	ST. ERROR		COEFF	ST. ERROR	
	g and h		HETROB	DYAD	COEFF	IID	DYAD	COEFF	DEFAULT	
Currency Union	Yes	1.8588	0.4556	0.6888	0.8952	0.3377	0.559	1.3463	0.3251	
Log_Distance	Yes	-1.3667	0.0349	0.1215	-1.6939	0.0386	0.1196	-1.5858	0.0319	
Log_product_real_GDP		0.7656	0.0204	0.0758	--	--	--	0.7649	0.0678	
Log_product_real_GDP_pc		0.867	0.0128	0.0469	--	--	--	0.8455	0.0367	
Constant		-34.8421	0.6827	2.1609	0.0000	0.0198	0.0007	-32.0411	1.6155	
Country 1 and 2 fixed effects	No				Yes			No		
Country 1 and 2 dyadic random effects	No				No			Yes		
Observations		4618			4618			4618		
RMSE		1.7526								
R-squared		0.708			0.828					

g and h is YES if regressor varies over (g,h) rather than just g or just h

HETROB = PAIRS is heteroskedastic robust standard error

IID is default standard errors assuming i.i.d. errors

DYAD is dyadic cluster-robust standard error

DEFAULT is FGLS standard error assuming the country random effects and idiosyncratic error are all i.i.d.

TABLE 4A: COUNTRY-PAIR PANEL DATA EXAMPLE WITH NO COUNTRY-PAIR FIXED EFFECTS
Comparison of Standard Errors computed in various ways

	COEFF	TIME	IID	HETROB	PAIRS	CTRY1	CTRY2	TWOWAY	DYAD	SE RATIOS			
										DYAD /HET	DYAD /PAIRS	DYAD /2WAY	DYAD /JACK
Both_in_GATTorWTO	-0.0423	Yes	0.0159	0.0182	0.0528	0.1222	0.1066	0.1533	0.1864	10.24	3.53	1.22	1.23
One_in_GATTorWTO	-0.0583	Yes	0.0154	0.0178	0.0486	0.0952	0.0709	0.1083	0.1256	7.04	2.58	1.16	1.20
GSP	0.8585	Yes	0.0111	0.0094	0.0318	0.0978	0.0741	0.1185	0.1308	13.97	4.11	1.10	1.31
Log_Distance	-1.1190	Yes	0.0061	0.0061	0.0221	0.0510	0.0495	0.0676	0.0781	12.90	3.54	1.16	1.22
Log_product_real_GDP	0.9159	Yes	0.0026	0.0027	0.0096	0.0236	0.0278	0.0352	0.0435	16.40	4.53	1.24	1.22
Log_product_real_GDP_pc	0.3214	Yes	0.0039	0.0040	0.0141	0.0327	0.0377	0.0479	0.0520	13.07	3.70	1.09	1.26
Regional_FTA	1.1988	Yes	0.0360	0.0291	0.1062	0.2217	0.1855	0.2688	0.3327	11.42	3.13	1.24	1.26
Currency_Union	1.1181	Yes	0.0374	0.0346	0.1217	0.1652	0.1757	0.2082	0.2353	6.80	1.93	1.13	1.13
Common_language	0.3125	Yes	0.0110	0.0111	0.0403	0.0808	0.0655	0.0959	0.1109	10.02	2.75	1.16	1.24
Land_border	0.5257		0.0266	0.0259	0.1107	0.1478	0.1500	0.1791	0.2071	8.00	1.87	1.16	1.18
Number_landlocked	-0.2706		0.0093	0.0096	0.0312	0.0692	0.0789	0.1002	0.1096	11.43	3.51	1.09	1.26
Number_islands	0.0419		0.0095	0.0092	0.0361	0.1047	0.0830	0.1287	0.1538	16.63	4.26	1.20	1.18
Log_product_land_area	-0.0967		0.0020	0.0021	0.0080	0.0251	0.0254	0.0348	0.0433	21.08	5.41	1.24	1.20
Common_colonizer	0.5846		0.0162	0.0194	0.0672	0.1261	0.1083	0.1521	0.1777	9.17	2.65	1.17	1.22
Currently_colonized	1.0751	Yes	0.1067	0.0704	0.2347	0.4618	0.2560	0.4729	0.4802	6.83	2.05	1.02	0.55
Ever colony	1.1638		0.0312	0.0230	0.1172	0.1927	0.1149	0.1914	0.2086	9.08	1.78	1.09	1.05
Common_country	-0.0163		0.2623	0.2049	1.0811	0.8846	1.0774	0.8800	0.8596	4.20	0.80	0.98	0.47
Constant	-24.9600		0.2870	0.2225	0.4071	0.9811	0.9756	1.3224	1.5526	6.98	3.81	1.17	1.25
Year dummies	Yes												
Country-pair dummies	No												
Observations	234597												
R-squared	0.648												

TIME is YES if regressor varies over time as well as over country pair.

IID is default standard errors assuming i.i.d. errors

HETROB is heteroskedastic robust standard error

PAIRS is one-way cluster robust standard error with clustering on country-pair (g,h)

CTRY1 is one-way cluster robust standard error with clustering on country 1 (g)

CTRY2 is one-way cluster robust standard error with clustering on country 1 (h)

TWOWAY is two-way cluster robust standard error (g and h)

DYAD is dyadic cluster-robust standard error

NJACK is node-jackknife cluster-robust standard error

SIG 5% is Yes if statistically different from 5% based on dyadic cluster-robust standard errors
Critical values use $t(N-2)$ for IID and ROBUST, and $t(G^*-1) = t(G-2)$ for the remainder.

DYAD/HET is ratio of dyadic-robust and heteroskedastic-robust standard errors

DYAD/PAIRS is ratio of dyadic-robust and one-way cluster-pairs robust standard errors

DYAD/2WAY is ratio of dyadic-robust and two-way cluster-robust standard errors

DYAD/JACK is ratio of dyadic-robust and node-jackknife standard errors

TABLE 4B: COUNTRY-PAIR PANEL DATA EXAMPLE WITH COUNTRY-PAIR FIXED EFFECTS
Comparison of Standard Errors computed in various ways

	COEFF	TIME	IID	HETROB	PAIRS	CTRY1	CTRY2	TWOWAY	DYAD	NJACK	SIG 5%	SE RATIOS			
												DYAD /HET	DYAD /PAIRS	DYAD /2WAY	DYAD /JACK
Both_in_GATTorWTO	0.1271	Yes	0.0185	0.0190	0.0419	0.0883	0.0773	0.1096	0.1325	0.1011	No	6.96	3.16	1.21	1.31
One_in_GATTorWTO	0.0600	Yes	0.0158	0.0169	0.0372	0.0754	0.0509	0.0830	0.0949	0.0742	No	5.62	2.55	1.14	1.28
GSP	0.1754	Yes	0.0124	0.0105	0.0277	0.0697	0.0510	0.0818	0.0828	0.0631	Yes	7.89	2.99	1.01	1.31
Log_Distance	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Log_product_real_GDP	0.4425	Yes	0.0177	0.0174	0.0465	0.1046	0.0985	0.1360	0.1513	0.1150	Yes	8.69	3.25	1.11	1.32
Log_product_real_GDP_pc	0.2368	Yes	0.0166	0.0165	0.0449	0.0968	0.0948	0.1279	0.1415	0.1079	No	8.57	3.15	1.11	1.31
Regional_FTA	0.7639	Yes	0.0384	0.0254	0.0732	0.1286	0.1177	0.1582	0.1717	0.1379	Yes	6.76	2.34	1.09	1.24
Currency_Union	0.6314	Yes	0.0482	0.0457	0.1141	0.1292	0.1117	0.1270	0.1605	0.1500	Yes	3.51	1.41	1.26	1.07
Common_language	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Land_border	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Number_landlocked	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Number_islands	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Log_product_land_area	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Common_colonizer	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Currently_colonized	0.2957	Yes	0.0847	0.0439	0.1531	0.0730	0.1558	0.0784	0.0973	0.1381	Yes	2.22	0.64	1.24	0.70
Ever_colony	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Common_country	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Constant	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	--	--	--	--	--
Year dummies	Yes														
Country-pair dummies	Yes														
Observations	234597														
R-squared	0.853														

TIME is YES if regressor varies over time as well as over country pair.

IID is default standard errors assuming i.i.d. errors

HETROB is heteroskedastic robust standard error

PAIRS is one-way cluster robust standard error with clustering on country-pair (g,h)

CTRY1 is one-way cluster robust standard error with clustering on country 1 (g)

CTRY2 is one-way cluster robust standard error with clustering on country 1 (h)

TWOWAY is two-way cluster robust standard error (g and h)

DYAD is dyadic cluster-robust standard error

NJACK is node-jackknife cluster-robust standard error

SIG 5% is Yes if statistically different from 5% based on dyadic cluster-robust standard errors

Critical values use $t(N-2)$ for IID and ROBUST, and $t(G^*-1) = t(G-2)$ for the remainder.

DYAD/HET is ratio of dyadic-robust and heteroskedastic-robust standard errors

DYAD/PAIRS is ratio of dyadic-robust and one-way cluster-pairs robust standard errors

DYAD/JACK is ratio of dyadic-robust and node-jackknife standard errors

DYAD/2WAY is ratio of dyadic-robust and two-way cluster-robust standard errors

DYAD/JACK is ratio of dyadic-robust and node-jackknife standard errors