Robust Inference for Dyadic Data with Application to the Gravity Model of Trade*

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

June 2023

Abstract

In this paper we consider inference with paired or dyadic data, such as cross-section and panel data on trade between pairs of countries. Regression models with such data can 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. The standard cluster-robust variance estimator or sandwich estimator based on one-way clustering on dyads is inadequate. The two-way cluster-robust estimator with clustering on each pair in the dyad is a substantial improvement, but still understates standard errors. Instead one should use dyadic-robust standard errors. Qualitatively similar issues arise in social network data analysis, but the consequences are especially severe in international trade studies since trade networks are typically very dense. Applications with the gravity model of trade rarely use dyadic-robust standard errors, though panel data applications do include rich sets of fixed effects that could potentially control for dyadic correlation. Using several leading applications we find that even when country-pair and country-time fixed effects are included the failure to additionally use dyadic-robust standard errors can lead to reported standard errors that are several times too small.

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

JEL Classification: C12, C21, C23.

---

*A preliminary version titled "Robust Inference for Dyadic Data" see Cameron (2014), was 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

[‡]School of Public Policy and Dept. of Economics, Cornell University. dlm386@cornell.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 consider extension of cluster-robust methods for one-way and two-way clustering to regressions with paired or dyadic data, such as country-pair data analyzed frequently in international trade applications.

In the simplest case of one-way clustering, observations are grouped so that model errors are independent across groups but are potentially correlated within groups. In that case Moulton (1986, 1990) and Bertrand, Duflo and Mullainathan (2004) demonstrated that the need to control for such 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 standard way to control for one-way clustering is to use "cluster-robust" standard errors that generalize those of White (1980) for independent heteroskedastic errors. Key references include Shah et al. (1977) for clustered sampling, Liang and Zeger (1986) for estimation in a generalized estimating equations setting, and Arellano (1987) and Hansen (2007) for linear panel models.

Cluster-robust inference for the one-way case has been generalized to two-way and multi-way clustering in independent papers by Miglioretti and Heagerty (2006), Cameron, Gelbach and Miller (2011) and Thompson (2011). Davezies, D'Haultfoueille and Guyonvarch (2021), MacKinnon, Nielsen and Webb (2021) and Menzel (2021) provide relevant theory. 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. Cameron and Miller (2015) and MacKinnon, Nielsen and Webb (2023) provide surveys.

In this paper we consider a different extension of one-way clustering, that due to paired or dyadic data, such as that for trade flows between countries. Then the observational unit is a country pair (g, h) if data are cross-sectional, or country pair by time period (g, h, t) if panel data are available. Model errors are likely to be correlated between country-pair observations that have a country in common. For example, errors for the US-UK pair may be correlated with those for any other country pair that includes either the US or UK. Such correlation is the consequence of, for example, a random effects model $y_{gh} = \mathbf{x}'_{gh}\boldsymbol{\beta} + \alpha_g + \alpha_h + u_{gh}$ where α_g and α_h are random effects for country g and country h.

An obvious approach is to use two-way clustering where the first grouping is defined by the first country in the pair and the second grouping is defined by the second country in the country pair. This picks up correlation of the US-UK pair with any country pair ordered with US as the first country and/or UK as the second country. But two-way clustering fails to pick up correlation with any country pair ordered with UK as the first country or with US as the second country.

Dyadic-robust inference controls for this additional source of correlation. The dyadic-robust variance estimator was proposed in an applied paper by Fafchamps and Gubert (2007), its consistency was proven in a limited setting by Aronow, Samii and Assenova (2015), and the asymptotic normality of the resultant Wald t-statistic was proven for OLS by Tabord-Meehan (2019) under more general conditions. Davezies, D'Haultfoueille and Guyonvarch (2021), MacKinnon, Nielsen and Webb (2021), Menzel (2021), Graham (2020a) and Graham (2020b) provide theory.[1]

The diffusion of cluster-robust inference theory to practice has been slow. Almost all economics studies that should control for one-way clustering now do so. Many practitioners with two-way clustering still erroneously one-way cluster on the pair (such as employer-employee) leading to considerably under-estimated standard errors. For dyadic data exceptionally few papers currently control for dyadic correlation. Carlson, Incerti and Aronow (2021) reanalyze 22 papers published in *International Organization*. Most studies used heteroskedastic-robust or some sort of cluster-robust standard errors, but none of these papers used dyadic-robust standard errors. Dyadic-robust standard errors for key regressors were on average twice as large.

In this paper we consider application of dyadic-robust inference to estimation of the gravity model of international trade. Following the seminal paper by Baier and Bergstrand (2007) it is common to include a rich set of fixed effects, to enable causal interpretation of key regression coefficients. In particular, for panel data it is standard to include country-pair fixed effects and time dummies. Additionally time dummies may be interacted with the first country in the pair and with the second country in the pair.

Subsequent inference implicitly presumes that inclusion of a rich set of fixed effects is sufficient to control for any correlation of errors across country pairs, so that valid inference could be based on one-way clustering on country pairs. We find in empirical applications

[1] Leung (2023) instead considers the case where the network can be partitioned into clusters, in which case one-way cluster-robust methods can be used.

that this is not the case – one-way clustering on country pairs leads to underestimation of standard errors and Wald t tests that over-reject. This is consistent with experience with standard panel data analysis where the inclusion of individual-specific fixed effects does not eliminate within-individual error correlation so that it is still necessary to obtain standard errors that cluster on the panel unit. For short panel data Arellano (1987) explicitly covered the case of fixed effects, and MacKinnon et al. (2023, p.276) present a standard one-way factor model as an example where cluster-robust standard errors need to be used.

Furthermore, in international trade applications the appropriate dyadic-robust standard errors and test p-values can be much larger than those obtained by one-way clustering on country pairs. This is a consequence of the dyads forming a dense network, since on average a given country trades with many other countries. For example, with bidirectional trade data on a cross-section of 100 countries that all trade with each other there are $100 \times 99/2 = 4950$ unique country-pair observations, but the information content can be much less than that of 4950 independent observations.

Cluster-robust inference does not require modelling the within-cluster correlation, but does rely on the assumption that the number of clusters, rather than the number of observations, 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. We find that a similar problem arises for dyadic-robust standard errors when the dyads are based on few underlying units (countries in the gravity model application) where "few" can be quite large.

The methods are presented in Section 2. In Section 3 we present Monte Carlo experiments. Section 4 presents several international trade applications that demonstrate the importance of controlling for clustering with dyadic data, even with panel data and a rich set of fixed effects. Section 5 concludes.

2. Robust Inference

This section reviews one-way and two-way cluster robust methods for OLS before moving to the dyadic case.

2.1. Cluster-robust standard errors and inference

Consider the linear regression model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i, \quad i = 1, ..., N,$$

where i denotes the i^{th} of N individuals in the sample, $\mathbf{x}_i$ is $K \times 1$ and $\mathrm{E}[u_i|\mathbf{x}_i] = 0$.

The OLS estimator is

$$\widehat{\boldsymbol{\beta}} = \left(\sum\nolimits_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \sum\nolimits_{i=1}^{N} \mathbf{x}_i y_i = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y}.$$

The OLS estimator has variance conditional on $\mathbf{X}$

$$\begin{aligned} \mathrm{V}(\widehat{\boldsymbol{\beta}}) &= \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathrm{Var}\left[\sum\nolimits_{i=1}^{N} \mathbf{x}_i u_i\right] \left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1} \sum\nolimits_{i=1}^{N} \sum\nolimits_{j=1}^{N} \mathrm{E}[\mathbf{x}_i\mathbf{x}_j' u_i u_j] \left(\mathbf{X}'\mathbf{X}\right)^{-1}, \end{aligned}$$

using $\mathrm{E}[u_i|\mathbf{x}_i] = 0$.

Various cluster-robust estimates of the variance take the form

$$\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}) = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(c_N \sum\nolimits_{i=1}^{N} \sum\nolimits_{j=1}^{N} \mathbf{1}_{ij}\mathbf{x}_i\mathbf{x}_j'\widehat{u}_i\widehat{u}_j\right) \left(\mathbf{X}'\mathbf{X}\right)^{-1}, \tag{2.1}$$

where $\mathbf{1}_{ij}$ is an indicator function that selects only terms with $\mathrm{E}[\mathbf{x}_i\mathbf{x}_j' u_i u_j] \neq 0$, and $c_N \rightarrow 1$ is a degrees-of-freedom correction. The indicator function $\mathbf{1}_{ij}$ varies with the type of clustering.

For one-way clustering the errors are potentially correlated within clusters $g = 1, ..., G$ but are assumed to be independent across clusters. Then in (2.1) we set $\mathbf{1}_{ij} = \mathbf{1}[(i, j)$ both in cluster $g]$, and most often $c_N = G/(G-1)$ or $c_N = [G/(G-1)] \times [(N-1)/(N-K)]$. This estimator proposed by Liang and Zeger (1986) and Arellano (1987) is a natural generalization of White (1980) who considered models with independent heteroskedastic standard errors; then in (2.1) $\mathbf{1}_{ij} = \mathbf{1}[i = j]$ and $c_N = N/(N-K)$.

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). For two-way clustering with errors correlated within clusters $g = 1, ..., G$ and within clusters $h = 1, ..., H$, we let $\mathbf{1}_{ij} = \mathbf{1}[(i, j)$ both in cluster g and/or cluster $h]$ in (2.1). Then the variance can be computed as

$$\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}) = \widehat{\mathrm{V}}_G + \widehat{\mathrm{V}}_H - \widehat{\mathrm{V}}_{G\cap H} \tag{2.2}$$

where $\widehat{\mathrm{V}}_G$, $\widehat{\mathrm{V}}_H$ and $\widehat{\mathrm{V}}_{G\cap H}$ are one-way cluster-robust variance estimates with clustering on, respectively, G, H and $G \cap H$.

For test of $H_0 : \beta_k = \beta_{k0}$ the Wald t-statistic is

$$W_{k0} = \frac{\widehat{\beta}_k - \beta_{k0}}{se(\widehat{\beta}_k)},$$

where $se(\widehat{\beta}_k)$ is the square root of the k^{th} diagonal entry in $\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}})$ defined in (2.1). Under appropriate assumptions

$$W_{k0} \xrightarrow{d} N(0,1) \text{ under } H_0 : \beta_k = \beta_{k0}.$$

The asymptotic theory for the Wald statistic is complicated as rates of convergence of $\mathrm{Var}\left[\sum_{i=1}^{N} \mathbf{x}_i u_i\right]$ vary with the nature of the within-cluster correlation. If within-cluster correlation declines with some measure of distance between observations, then the rate of convergence is in the sample size N. More often in microeconometrics applications there is no such decline in the within-cluster correlation; a leading example is a random effects model that implies equicorrelated errors within cluster. Then the rate of convergence is in the number of clusters. With few clusters there can be considerable downwards bias in standard error estimates and hypothesis test over-rejection, even if coefficient estimates are reasonably precise due to many observations per cluster.

2.2. Dyadic model and correlation

A dyad based on G units such as G countries is a pair indexed by (g, h), where g and h may each take integer values between 1 and G. In practice not all possible dyads may arise. We define $\mathcal{D}$ to be the set of dyads with M elements and $(g, h) \in \mathcal{D}$.

Given a single observation per dyad we consider estimation of $\boldsymbol{\beta}$ in the linear model

$$y_{gh} = \mathbf{x}'_{gh}\boldsymbol{\beta} + u_{gh}. \tag{2.3}$$

For notational simplicity this section presents results for this cross-section case. Results generalize immediately to multiple observations per dyad with arbitrary correlation allowed across errors for observations in the same dyad. This covers panel data in which case

$$y_{ght} = \mathbf{x}'_{ght}\boldsymbol{\beta} + u_{ght}.$$

For international trade data on G countries, y_{gh} is a measure of trade between country g and country h and the dyad is the country pair (g, h). If the trade measure is bidirectional, such as imports plus exports, then there are at most $G(G-1)/2$ unique dyads, since $y_{gh} = y_{hg}$ and countries do not trade with themselves. If the trade measure is unidirectional, such as imports only (or exports only), then there are at most $G(G-1)$ unique dyads.

We assume that the only potential error correlation is that between dyads (g, h) with at least one of g and h in common. Thus we assume that

$$\mathrm{E}[u_{gh}u_{g'h'}|\mathbf{x}_{gh}, \mathbf{x}_{g'h'}] = 0, \text{ unless } g = g' \text{ or } h = h' \text{ or } g = h' \text{ or } h = g'. \quad (2.4)$$

For example, the (US,UK) error is assumed to be independent of errors for any country pair not involving the US or UK, but may be correlated with the error for any country pair that involves either the US or the UK. The random effects model $y_{gh} = \mathbf{x}'_{gh}\boldsymbol{\beta} + \alpha_g + \alpha_h + u_{gh}$ where α_g and α_h are independent random effects has the property (2.4).

To make ideas concrete, consider the case of cross-section data on bidirectional trade between four countries, numbered $1, 2, 3$, and 4. Then there are six unique country-pair observations, namely $(1, 2)$, $(1, 3)$, $(1, 4)$, $(2, 3)$, $(2, 4)$ and $(3, 4)$, and there are $6^2 = 36$ error correlations. Table 2.1 illustrates the potential error correlations.

Table 2.1: Dyadic correlation with four countries and bidirectional trade.

(g,h)\(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	2way	2way
(2,3)	**dyad**	2way	0	cp	2way	**dyad**
(2,4)	**dyad**	0	2way	2way	cp	2way
(3,4)	0	**dyad**	2way	**dyad**	2way	cp

Note: One-way clustering on country-pair picks up terms labelled cp; two-way cluster robust additionally picks up terms labelled 2way; dyad-robust additionally picks up terms labelled dyad.

Clustering on country-pair controls for correlation when $(g, h) = (g', h')$. Then only the diagonal entries in Table 2.1 are nonzero; these are denoted cp. Clustering on country-pair coincides in the cross-section case with using heteroskedastic-robust standard errors.

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.

2.3. Default standard errors under dyadic correlation

Under one-way clustering with all regressors perfectly correlated within cluster, balanced clusters and $Cor(u_i, u_j) = \rho$ for i and j in the same cluster, the default OLS variance matrix $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$ should be inflated by $1 + \rho(N_G - 1)$ where N_G is the number of observations per cluster.

For dyadic data we obtain a similar result using the results of Cameron and Golotvina (2005) who considered OLS and feasible GLS estimation for the following dyadic variant of a two-way random effects model for unidirectional dyadic data. For the $(g, h)^{th}$ country-pair the regression model is

$$y_{gh} = \mathbf{x}_{gh}'\boldsymbol{\beta} + \alpha_g + \alpha_h + \varepsilon_{gh}, \quad h = 1, ..., g-1, \quad g = 1, ..., G, \tag{2.5}$$

where ε_{gh} is an idiosyncratic error and α_g and α_h are country-specific error components with 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 $\varepsilon_{gh} \sim (0, \sigma_\varepsilon^2)$ and $\alpha_g \sim (0, \sigma_\alpha^2)$. Then the error $v_{gh} = \alpha_g + \alpha_h + \varepsilon_{gh}$ has variance $(2\sigma_\alpha^2 + \sigma_\varepsilon^2)$, covariance σ_α^2 across dyads with $g = g', h \neq h', g \neq g'$ or $h = h'$, and zero covariance otherwise. The implied dyadic error correlation for dyads with a unit in common is then $\rho = \sigma_\alpha^2/(2\sigma_\alpha^2 + \sigma_\varepsilon^2)$.

Cameron and Golotvina (2005) show that for the intercept-only model with $\mathbf{x}_{gh}'\boldsymbol{\beta} = \beta$ in (2.5), i.i.d. errors α_g and ε_{gh}, and data available for all possible dyads, the default variance estimate obtained by assuming independence ($\sigma_\alpha^2 = 0$) should be inflated by the multiple $\frac{1+\tau(G-1)}{1+\tau} = 1 + \frac{\tau}{1+\tau}(G-2)$ where $\tau = \frac{2\sigma_\alpha^2}{\sigma_\varepsilon^2} = \frac{2\rho}{1-2\rho}$. For example, if the dyadic correlation is $\rho = 0.1$ then $\tau = 0.25$ and the default variance estimate should be inflated by $\{1+0.8(G-2)\}$. It follows that with $\rho = 0.1$ and $G = 12$ the default standard errors should be inflated by the multiple $\sqrt{9} = 3$.

So even very mild within-dyad error correlation can lead to dyadic-corrected standard errors being much larger than default standard errors that assume independence across dyads.

The intercept-only model is an extreme case as the regressor is a constant that is necessarily common to dyads that share a unit in common. The variance inflation factor will be less for regressors that are not so highly correlated across dyads that share a unit in common.

2.4. Dyadic-robust standard errors and inference

Dyadic-robust inference does not require specification of a model for the dyadic correlation.

In the case of one observation per dyad the dyadic-robust estimate of the variance matrix takes the form

$$\widehat{\mathrm{V}}_{dyad}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1} \left(c_N \sum\nolimits_{g,h} \sum\nolimits_{g',h'} \mathbf{1}_{ghg'h'} \mathbf{x}_{gh} \mathbf{x}'_{g'h'} \widehat{u}_{gh} \widehat{u}_{g'h'} \right) (\mathbf{X}'\mathbf{X})^{-1}, \qquad (2.6)$$

where

$$\mathbf{1}_{ghg'h'} = \mathbf{1}[g = g' \text{ or } h = h' \text{ or } g = h' \text{ or } h = g'], \qquad (2.7)$$

and $c_N \rightarrow 1$ is a finite-sample adjustment.

We use $c_N = [(G-1)/(G-2)] \times [(N-1)/(N-K)]$ where G is the number of countries and N is the number of observations. This is similar to the correction used in the one-way cluster case, except we $(G-1)/(G-2)$ rather than $G/(G-1)$ as countries do not trade with themselves. We additionally propose an adjustment when G is small.

This method was proposed 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 average degree is low." Here degree is the number of links and in their study each individual had relatively few links.

The consistency of $\widehat{\mathrm{V}}_{dyad}(\widehat{\boldsymbol{\beta}})$ was proven in a limited setting by Aronow, Samii and Assenova (2015). The asymptotic normality of the resultant Wald t-statistic was proven by Tabord-Meehan (2019) under quite general conditions that permit a wide range of dyadic correlation and network density; see also Graham (2020). A key result is that in the case of a dense network and dyadic correlation that is not declining in any distance measure the rate of convergence is in G, the number of units (here countries), rather than in the much larger number of dyads.

Computation of one-way and two-way cluster-robust standard errors is straightforward; in the latter case one use three easily computed separate one-way cluster-robust variance estimates. For the dyadic-robust case Aronow et al. (2015) provide a qualitatively similar decomposition though forming the separate clusters is more complicated. For OLS regression

Bisbee (2019) and Carlson (2021a) provide R packages and Balcazar (2020) and Carlson (2021b) provide Stata commands. The dyadic-robust variance estimate, like the two-way cluster robust estimate, is not guaranteed to be positive semi-definite and we propose a correction in such cases.

It is important to note that dyadic-robust inference is different from two-way-robust, as should be clear from Section 2.2. Several studies confuse the two.[2]

2.5. Dyadic-robust Inference with Few Dyadic Units

As one moves progressively through variance estimates that are in turn heteroskedastic robust, one-way cluster robust, two-way cluster robust and dyadic robust, increasing numbers of terms in the double sum $\sum_i \sum_j \mathrm{E}[\mathbf{x}_i \mathbf{x}_j' u_i u_j]$ or $\sum_{g,h} \sum_{g',h'} \mathrm{E}[\mathbf{x}_{gh} \mathbf{x}_{g'h'} u_{gh} u_{g'h'}]$ are estimated rather than set to zero. This adds noise to the variance estimate that disappears asymptotically, but that increasingly leads to poor finite sample performance when there are few cluster units or dyadic units.

Return to Table 2.1 for bidirectional data with only $G = 4$ countries. In this example, extreme as G is so small, 30 out of 36 terms in the dyadic error variance matrix are nonzero, so almost all model errors are potentially correlated with each other. These additional terms likely lead to underestimation of $\sum_{g,h} \sum_{g',h'} \mathrm{E}[\mathbf{x}_{gh} \mathbf{x}_{g'h'} u_{gh} u_{g'h'}]$ since we know that if all 36 terms are nonzero then $\sum_{g,h} \sum_{g',h'} \mathbf{x}_{gh} \mathbf{x}_{g'h'}' \widehat{u}_{gh} \widehat{u}_{g'h'} = 0$ since $\sum_{g,h} \mathbf{x}_{gh} \widehat{u}_{gh} = 0$.

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 (2.4) 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 nonzero is $(2G-3)/[G(G-1)/2] = (4-6/G)/(G-1)$. 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%.

In the limit for large G and either unidirectional or bidirectional models where all countries trade with all countries there are $4/(G-1)$ potential nonzero error correlations compared

[2] Cameron, Gelbach and Miller (2011, page 239) stated that "Fafchamps and Gubert (2006) analyze networks among individuals, where a person-pair is the unit of observation. In this context they describe the two-way robust estimator in the setting of dyadic models." And section 4.2 gave a bilateral trade example. Post-publication we understood the difference between two-way and dyadic correlation, leading to Cameron and Miller (2014).

to $2/(G-1)$ with two-way clustering and only $2/[G(G-1)]$ if errors are uncorrelated.

Simulations reveal that for finite G one should at a minimum use a finite-cluster degrees-of-freedom adjustment to the variance matrix and for Wald tests of a single restriction additionally use p-values and critical values based on a t distribution rather than the the standard normal.

For one-way clustering with G clusters it is standard to use the correction factor $c_N = G/(G-1)$ or $c_N = [G/(G-1)] \times [(N-1)/(N-K)]$ in (2.1), and to use the $t(G-1)$ distribution rather than the standard normal distribution for tests. With few clusters these adjustments still lead to over-rejection and a popular alternative method is the Wild cluster bootstrap which provides an asymptotic refinement to the Wald test; see Cameron, Gelbach and Miller (2008), MacKinnon and Webb (2018), and Djogbenou, MacKinnon, and Nielsen (2019).

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. At a minimum one should use a variance matrix estimate with finite-cluster correction and use the Students t-distribution with $\min(G, H)-1$ degrees of freedom. The analog for the dyadic case is $G-1$. And MacKinnon, Nielsen, and Webb (2021) propose a bootstrap for two-way clustering.

Menzel (2021) proposes a bootstrap with asymptotic refinement for dyadic data.

Like many dyadic-theory papers, Menzel (2021) makes the assumption of exchangeable arrays for model errors. In that case, Marrs, Fosdick and McCormick (2023) show that the error variance-covariance matrix has at most six distinct terms in the dyadic cross-section case and at most twelve terms in the dyadic panel case, and provide R software to obtain dyadic-standard errors, under the restrictive assumption of exchangeable arrays that, for example, rules out heteroskedasticity.

2.6. Degrees-of-freedom adjusted dyadic standard errors

We propose adjusting the dyad-robust standard errors by a multiple that is obtained by a parametric simulation. This multiple varies with each regressor.

Specifically consider estimation of the coefficient of the k^{th} regressor and let $\widehat{v}_k$ denote the dyad-robust estimate of the variance of $\widehat{\beta}_k$. Generate S samples and for each sample obtain parameter estimates $\widehat{\beta}_k^{(s)}$, $k = 1, ..., K$, and corresponding dyad-robust variance estimates $\widehat{v}_k^{(s)}$ of $\widehat{\beta}_k^{(s)}$. A variance estimate inflation factor is obtained as the ratio of the variance of

$\widehat{\beta}_k^{(s)}$ across the S simulations to the average across the S simulations of the (finite-sample biased) dyad-robust variance estimates $\widehat{v}_k^{(s)}$. Then

$$\begin{aligned} \widehat{v}_{k,corrected} &= c_k \times \widehat{v}_k \\ c_k &= \frac{\frac{1}{S-1}\sum_{s=1}^{S}(\widehat{\beta}_k^{(s)} - \overline{\widehat{\beta}}_k)^2}{\frac{1}{S}\sum_{s=1}^{S}\widehat{v}_k^{(s)}}, \end{aligned}$$

where $\overline{\widehat{\beta}}_k = \frac{1}{S}\sum_{s=1}^{S}\widehat{\beta}_k^{(s)}$.

The S generated samples use the original regressor values $\mathbf{x}_{gh}$, while the dependent variable is generated from the model

$$y_{gh}^{(s)} = \mathbf{x}_{gh}'\widehat{\boldsymbol{\beta}} + u_{gh}^{(s)} \tag{2.8}$$

where $\widehat{\boldsymbol{\beta}}$ is the original OLS estimate and $u_{gh}^{(s)}$ are i.i.d. normal with mean 0 and standard deviation equal to the standard error from the original OLS regression. This method provides a correct adjustment factor c_j if model errors are indeed i.i.d.

2.7. Degrees-of-freedom adjusted t tests

We propose computing Wald test p-values and critical values from a t distribution with an adjusted degrees of freedom where the adjustment varies with each regressor.

Again obtain S samples by simulation from (2.8). Since each sample is generated with $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, we perform a two-sided Wald test of the hypothesis that $\beta_k = \widehat{\beta}_k$ using

$$W_k^{(s)} = \frac{\widehat{\beta}_k^{(s)} - \widehat{\beta}_k}{\sqrt{\widehat{v}_k^{(s)}}}.$$

The S values of $W_k^{(s)}$ provide an empirical estimate of the distribution of the Wald test statistic.

We are interested in the tail behavior of the distribution and propose basing inference on the $t(d)$ distribution where d is chosen to matching tail probabilities of the $t(d)$ distribution to tail probabilities of the empirical distribution of $W_k^{(s)}$, $s = 1, ..., S$. Evaluation in Monte Carlos is currently not complete.

2.8. Node-Jackknife Variance Estimator

The delete-one-node estimate $\widehat{\boldsymbol{\beta}}_{(-g)}$ 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 $h=1,...,G$ are dropped).

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

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}]=\frac{G-2}{2G}\sum_{g=1}^{G}(\widehat{\boldsymbol{\beta}}_{(-g)}-\overline{\widehat{\boldsymbol{\beta}}})(\widehat{\boldsymbol{\beta}}_{(-g)}-\overline{\widehat{\boldsymbol{\beta}}})', \tag{2.9}$$

where $\overline{\widehat{\boldsymbol{\beta}}}=\frac{1}{G}\sum_{g=1}^{G}\widehat{\boldsymbol{\beta}}_{(-g)}$. 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 that is not appropriate in this setting. Graham (2020b, section 4.4) provides an adjustment that is asymptotically equivalent to the dyadic-robust variance estimate.

2.9. 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 $\sum_{g,h}\mathbf{m}_{gh}(\widehat{\boldsymbol{\theta}})=\mathbf{0}$. Examples include nonlinear least squares estimation, maximum likelihood estimation, and instrumental variables estimation in the just-identified case. For the logit MLE $\mathbf{m}_{gh}(\boldsymbol{\beta})=(y_{gh}-\Lambda(\mathbf{x}_{gh}'\boldsymbol{\beta}))\mathbf{x}_{gh}$, where $\Lambda(\cdot)$ is the logistic c.d.f.

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

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\theta}}]=\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}'^{-1}, \tag{2.10}$$

where $\widehat{\mathbf{A}}=\sum_{gh}\left.\frac{\partial\mathbf{m}_{gh}}{\partial\boldsymbol{\theta}'}\right|_{\widehat{\boldsymbol{\theta}}}$, and $\widehat{\mathbf{B}}$ is an estimate of $\mathrm{V}[\sum_{gh}\mathbf{m}_{gh}]$. A robust estimator is $\widehat{\mathbf{B}}=c_N\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}_{ij}\mathbf{x}_i\mathbf{x}_j\widehat{u}_i\widehat{u}_j$ where $\mathbf{1}_{ij}$ is an indicator function that selects only terms with $\mathrm{E}[\mathbf{x}_i\mathbf{x}_ju_iu_j]\neq 0$, and $c_N\rightarrow 1$ is a degrees-of-freedom correction.

For dyadic clustering the analysis of the preceding sections carries through. Then an m-estimator solves $\sum_{i=1}^{N}\mathbf{m}_{gh}(\widehat{\boldsymbol{\theta}})=\mathbf{0}$ and we use (2.6)-(2.7) with $\mathbf{X}'\mathbf{X}$ replaced by $\sum_{g,h}\left.\frac{\partial\mathbf{m}_{gh}}{\partial\boldsymbol{\theta}'}\right|_{\widehat{\boldsymbol{\theta}}}$ and $\mathbf{x}_{gh}\mathbf{x}_{g'h'}\widehat{u}_{gh}\widehat{u}_{g'h'}$ (2.6) replaced by $\widehat{\mathbf{m}}_{gh}\widehat{\mathbf{m}}_{g'h'}$.

For GMM estimation for over-identified models, such as over-identified linear two stage least squares. $\widehat{\boldsymbol{\theta}}$ minimizes $Q(\boldsymbol{\theta}) = \left(\sum_{g,h} \mathbf{m}_{gh}(\boldsymbol{\theta})\right)' \mathbf{W} \left(\sum_{g,h} \mathbf{m}_{gh}(\boldsymbol{\theta})\right)$, where $\mathbf{W}$ is a symmetric positive definite weighting matrix. Under standard regularity conditions $\widehat{\boldsymbol{\theta}}$ is asymptotically normal with estimated variance matrix given dyadic clustering

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\theta}}] = \left(\widehat{\mathbf{A}}'\mathbf{W}\widehat{\mathbf{A}}\right)^{-1} \widehat{\mathbf{A}}'\mathbf{W}\widetilde{\mathbf{B}}\mathbf{W}\widehat{\mathbf{A}} \left(\widehat{\mathbf{A}}'\mathbf{W}\widehat{\mathbf{A}}\right)^{-1}, \tag{2.11}$$

where $\widehat{\mathbf{A}} = \sum_i \frac{\partial \mathbf{h}_i}{\partial \boldsymbol{\theta}'}\Big|_{\widehat{\boldsymbol{\theta}}}$, and $\widetilde{\mathbf{B}}$ is an estimate of $\mathrm{V}[\sum_i \mathbf{h}_i]$ that can be computed as in (2.6) with $\widehat{u}_{igh}\mathbf{x}_{igh}$ replaced by $\widehat{\mathbf{h}}_{igh}$.

The leading nonlinear model example for international trade applications, due to Santos Silva and Tenreyro (2006), fits a gravity model with dependent variable in levels (rather than logs) using an exponential mean model with multiplicative fixed effects. Estimation uses the Poisson quasi-MLE. Graham (2020) provides a dyadic empirical application.

2.10. Many Fixed Effects

There are several ways to implement fixed effects regression. These can lead to the same coefficient estimates but different standard error estimates due to different finite sample degrees of freedom adjustments. The Frisch-Waugh method partials out the fixed effects and performs OLS regression on the residuals.

In the case of country 1 fixed effects and country 2 fixed effects, 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 $\mathbf{r}\_\mathbf{x}_{gh}$. Then OLS regress r_y_{gh} on $\mathbf{r}\_\mathbf{x}_{gh}$ 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 (2.3) 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.

In the example of Rose (2004) studied below there are potentially as many as $178 \times 177/2 = 15,753$ country-pair fixed effects as well as time fixed effects. Then 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_y_{gh} on $\mathbf{r}\_\mathbf{x}^*_{ght}$. As in the preceding paragraph this will lead to smaller finite-sample adjustment factor $N-k$ than if (2.3) is directly estimated by OLS.

An alternative method when there are many fixed effects is to use the reghdfe Stata command due to Correia (2022). We use this when simulations are computationally intensive.

Inclusion of fixed effects can be more problematic in nonlinear models, due to the incidental parameters problem if a fixed effect estimate is based on few observations. For an exponential mean model with multiplicative fixed effects Jochmans (2017) proposes an alternative differencing estimator that does not have this problem, though his inference method ignores correlations in scores across observations (Jochmans (2017, footnote 7).

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.

The dyadic-robust standard errors are asymptotically valid for both unidirectional and bidirectional data. For bidirectional data, the dataset is much more unbalanced by country. For example, in Table 1 there are three observations for country 1, two observations for country 2 and 3 observations for country 3. Recent research for one-way clustering finds that the "few clusters" problem of asymptotic theory not fully kicking in is greater when cluster sizes are unequal; see, for example, Carter, Schnepel and Steigerwald (2017) and MacKinnon and Webb (2017). So we expect worse finite sample performance in simulations with bidirectional trade flows (the case here) than with unidirectional trade flows.

3.1. Monte Carlo Setup

In both cases the data generating process is of the form

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

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 from Table 2.1 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 (2.7) 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 (2.2) 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 (2.9).

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 $G \to \infty$ 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 A.

For $G = 100$, the various standard error estimates are all close to the standard deviation of $\widehat{\beta}_2$ 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 $\widehat{\beta}_2$ 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 $\widehat{\beta}_2$ 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 $\widehat{\boldsymbol{\beta}}$, necessitating the eigendecomposition adjustment presented in section 2.4, and in several other cases the estimated standard error of $\widehat{\beta}_2$ 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 $\widehat{\beta}_2$ 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. Random Effects Model Errors

Now let the error in model (3.1) be generated by a random effects model, with

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

where α_g and α_g are 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 $\text{Cor}^2[u_{gh}, u_{g'h'}] = \sigma^2_\varepsilon/(2\sigma^2_\alpha + \sigma^2_\varepsilon) = (1/12)/[(2/12) + 0.25] = 0.2$.

In this case as $G \to \infty$ 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 B.

For $G = 100$, the dyadic-robust standard errors are closest on average to the standard deviation of $\widehat{\beta}_2$ 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 $\widehat{\beta}_2$ (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 one cross-section data and four panel data examples. We commend the papers' authors for making their data and programs available to other researchers. It should be clear that the methods used in these 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." All papers predate Cameron and Miller (2014) and Aronow et al. (2015) that drew attention to the need for dyadic-robust inference.

The papers have data on a large number of countries (respectively, 127, 187, 96, 122 and 190 countries). The first two studies use bidirectional trade flows while the last three use unidirectional trade flows. The "few-cluster" problem is unlikely to arise.

The gravity model of trade with panel data is

$$\ln y_{ght} = \mathbf{x}'_{ght}\boldsymbol{\beta} + \text{fixed effects} + u_{ght} \tag{4.1}$$

where y_{ght} is the value of trade between countries g and h at time t. The regressors include log of distance between countries g and h (hence the term "gravity"), output measures such as log GDP in country g and in country h, and indicator variables for shared border or common language. Many studies focus on the effect of shared membership in organizations fostering free-trade, notably WTO (World Trade Organization) and GATT (General Agreement on Tariff and Trade).

The fixed effects can be some combination of none, time, country 1, country 2, country pairs, country1 by time, country 2 by time. Depending on the fixed effects included as regressors, control variables such as log distance, log GDP and shared border may no longer be identified.

4.1. Cross-section Example - RE: Rose and Engel (2002)

Rose and Engel (2002) estimate a gravity model using bidirectional trade for 127 countries in a single year. Data are available for 4,618 of the potential 8,001 country pairs (equals $127 \times 126/2$). The key regressor is an indicator variable for joint membership in a currency union, a binary regressor is non-zero for only 24 of the 4,618 observations. Inference is based on heteroskedastic-robust standard errors, which are identical to those from one-way cluster on country pair since in this cross-section case there is only one observation per country pair.

Table 1 re-estimates RE Table 3 column 1 which includes no fixed effects. Dyad-robust standard errors are 1.53 times heteroskedastic-robust standard errors for currency union, and more than three times heteroskedastic-robust standard errors for the other three regressors.[3]

Table 2 re-estimates RE Table 3 column 1 with country 1 fixed effects and country 2 fixed effects added as regressors. (With a single cross-section there is no role for time fixed effects. Country-pair fixed effects cannot be included as there are as many of these as observations.) Then currency union has coefficient that is halved and smaller dyad-robust standard error that is 1.31 times the heteroskedastic-robust standard error. The log product regressors are not identified as $\ln(x_g \times x_h) = \ln x_g + \ln x_h$ and country g and country h fixed effects are included.

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 method goes a long way towards doing so, but the dyad-robust standard errors are still on average about 10-15% larger than two-way cluster-robust.

4.2. Panel Example 1 - R: Rose (2004)

Rose (2004) estimates a gravity model using bidirectional trade flows for 187 countries over 52 years (1948-1999). Data are available for 234,597 of the potential 819,156 observations (equals $52 \times 178 \times 177/2$). The key regressors are indicator variables for one or both countries in GATT/WTO. Inference is based on standard errors that one-way cluster on country-pair.

Table 3 re-estimates R Table 1 column 1 which includes time fixed effects. Dyad-robust standard errors are 3.10 times one-way cluster on country pairs standard errors for one country in GATT/WTO, and 2.86 times for two countries in GATT/WTO.

The GATT/WTO indicator variables have a surprising negative sign. Rose (2004) found that when country-pair fixed effects are added the coefficients have the expected positive sign (R Table 1 column 4).

Table 4 re-estimates R Table 4, with country-pair fixed effects rather than country fixed effects.[4] Dyad-robust standard errors are 3.14 times one-way cluster on country

[3] We 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.

[4] Our 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 coeffcients differ

pairs standard errors for one country in GATT/WTO, and 2.57 times for two countries in GATT/WTO.

Again 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 method goes a long way towards doing so, but the dyad-robust standard errors are still on average about 20% larger than two-way cluster-robust.

4.3. Panel Example 2 - BB: Baier and Bergstrand (2007)

Baier and Bergstrand (2007) propose fixed effects estimation with panel data to control for the endogeneity of free trade agreements. The data are unidirectional trade for 96 countries over 9 years (1960, 1965, 1970, ..., 2000). The key regressor is NOFTA, a binary variable equal to one if there is no free-trade agreement between countries g and h. Baier and Bergstrand used default (i.i.d.) standard errors. The key result of the paper is that inclusion of rich sets of fixed effects leads to a three-fold to five-fold increase in the coefficient of NOFTA.

Table 5 re-estimates BB Table 4 column 4 which includes country-pair (dyad) fixed effects and time fixed effects, leading to dropping time-invariant binary regressors for common language, common land border and distance.

Tables 6 and 7 re-estimate BB Table 5 columns 1 and 5 which include dyad fixed effects and two sets of country-time fixed effects, leading to additionally dropping log GDP for the two countries.

4.4. Panel Example 3 - DT: Dutt and Traca (2010)

Dutt and Traca (2010) use unidirectional trade for 122 countries over 13 years (1989 to 2001) or over 25 years (1980 to 2004) for manufacturing industry. Standard errors are one-way cluster on country 1 (footnote 18 of the article notes that this is better than one-way clustering son country pairs). Interest lies in the coefficients of the first three regressors.

We consider models with country-pair (dyad) fixed effects and time fixed effects included as regressors. Table 8 re-estimates DT Table 2 Panel A Column 4 using sectoral value added data for sector value added and sector consumption. Table 9 re-estimates DT Table 2 Panel B Column 4 using GDP of the exporter and importer countries.

---

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.

4.5. Panel Example 4 - DMV: Dutt, Mihov and Van Zandt (2013)

Dutt, Mihov and Van Zandt (2013) use unidirectional trade for 190 countries over 19 years (1988 to 2006). Standard errors are one-way cluster on country-pair (dyad). DMV decompose value of exports (y_{ght}) into number of products (N_{ght}) times the average value of exports per product ($\bar{y}_{ght}$). Then $\ln y_{ght} = \ln N_{ght} + \ln \bar{y}_{ght}$. Gravity model regressions at the extensive major use $\ln N_{ght}$ as the dependent variable and those at the intensive major use $\ln \bar{y}_{ght}$ as the dependent variable. The authors use robust standard errors that one-way cluster on country pair (dyad).

The first two models we consider have country-year fixed effects. Table 10 re-estimates DMV Table 1 Column 1 for the extensive margin. Table 11 re-estimates DMV Table 1 Column 2 for the intensive margin.

The remaining two models we consider have country-pair (dyad) fixed effects in addition to country-year fixed effects. Table 12 re-estimates DMV Table 1 Column 3 for the extensive margin. Table 13 re-estimates DMV Table 1 Column 4 for the intensive margin. For these two models we obtain different, though qualitatively similar, estimates from those given by DMV (we are currently checking for error at our end).

4.6. Summary of Empirical Examples

Figures 5.1 and 5.2 provide graphical summaries that present standard error ratios for all 93 regressor coefficients estimated across 13 models in five papers.

From Figure 5.1, dyadic-robust standard errors are on average 5.70 times i.i.d. (default) standard errors, 4.01 times heteroskedastic-robust standard errors and 2.56 times one-way cluster on country-pair (dyad) standard errors.

So while one-way clustering on country-pair is an improvement on using default or heteroskedastic-robust standard errors, the standard errors are still inadequate. The average ratio of one-way cluster standard errors to dyad-robust standard errors by study is 2.75 for Rose and Engel, 2.86 for Rose, 2.43 for Baier and Bergstrand, 3.10 for Dutt and Traca, and 1.89 for Dutt, Mihov and Van Zandt.

From Figure 5.2, dyadic-robust standard errors are on average 1.35 times one-way cluster on country 1, 1.50 times one-way cluster on country 2, 1.05 times two-way cluster on country 1 and country 2, and 1.10 times "corrected" dyadic standard errors.

So one-way clustering on country 1 (or on country 2) is a great improvement on one-way cluster on country pair. Better still is two-way cluster robust, though these are still less than the preferred dyad-robust. The average ratio of two-way cluster (on country 1 and on country 2) standard errors to dyad-robust standard errors by study is 1.09 for Rose and Engel, 1.14 for Rose, 1.13 for Baier and Bergstrand, 0.98 for Dutt and Traca, and 0.98 for Mihov and Van Zandt.

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 rich sets of fixed effects such as country-pair and country-time 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.

Baier, S.L. and J.H. Bergstrand (2007), "Do Free Trade Agreements Actually Increase Members' International Trade?," *Journal of International Economics*, 71, 72-95.

Balcazar, F. (2020), "dyadclust - Linear Regression using Dyadic Clustering," https://github.com/cfbalcazar/dyadclust

Bertrand, M., E. Duflo, and S. Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 119, 249-275.

Bisbee, J. (2019), "Dyad robust standard error estimator," https://rdrr.io/github/jbisbee1/dyadRobust/

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," 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*, 50(2), 317-373.

Carlson (2021a), "Stata-dcr," https://github.com/jscarlson/stata-dcr.

Carlson (2021b), "Dyadic Clustering–Robust (DCR) Variance Estimation per Aronow et al. (2015)," https://github.com/jscarlson/dcr.

Carlson, J., Incerti, T., and P.M. Aronow (2021), "Dyadic Clustering in International Relations," arXiv preprint arXiv:2109.03774.

Carter, A.V., K.T. Schnepel, and D.G. Steigerwald (2017), Asymptotic Behavior of a t Test

Robust to Cluster Heterogeneity," *Review of Economics and Statistics,* 99(4), 698–709.

Conley, T.G. (1999), "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92, 1-45.

Correia, S. (2022), "REGHDFE: Linear Models with Many Fixed Effects," http://scorreia.com/software/reg

Djogbenou, A., J.G. MacKinnon, and M.Ø. Nielsen (2019), "Asymptotic theory and wild bootstrap inference with clustered errors," *Journal of Econometrics*, 212, 393-412.

Dutt, P., I. Mihov and T. Van Zandt (2013), "*The effect of WTO on the extensive and the intensive margins of trade*," Journal of International Economics, 91, 204-219.

Dutt, P. and D. Traca, (2010), "Corruption and Bilateral Trade Flows: Extortion or Evasion," *Review of Economics and Statistics*, 92(4), 843-860.

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.

Graham, B. (2020a),"Dyadic Regression", in *The Econometric Analysis of Network Data*, B. Graham & A. de Paula (Eds.), 23–40, Amsterdam: Academic Press.

Graham, B. (2020b), "Network Data", in *Handbook of Econometrics 7A*, S. Durlauf, L. Hansen, J. Heckman and R. Matzkin, (Eds.), 111 - 218, Amsterdam: North-Holland.

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 (2016), "Robust Standard Errors in Small Samples: Some Practical Advice," *Review of Economics and Statistics*, 98, 701-712.

Jochmans (2017), "Two-way Models for Gravity," *Review of Economics and Statistics*, 99(3), 478–485.

Liang, K.-Y., and S.L. Zeger (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13-22.

Leung, M.P. (2023), "Network Cluster-robust Inference", *Econometrica*, 91(2), 641-667.

MacKinnon, J.G., M.Ø. Nielsen, and M.D. Webb (2021), "Wild Bootstrap and asymptotic inference with multiway clustering," *Journal of Business and Economic Statistics*, 39, 505-519.

MacKinnon, J.G., Nielsen, M.Ø. and M.D. Webb (2023), "Cluster-robust inference: A guide to empirical practice", *Journal of Econometrics*, 232(2), 272-299.

MacKinnon, J., and M.D. Webb (2017), "Wild Bootstrap Inference for Wildly Different Cluster Sizes." *Journal of Applied Econometrics*, 32, 233–254.

Mars, F.W., B.K. Fosdick and T.H. McCormick (2023), "Regression of exchangeable relational arrays," *Biometrika*, 110(1), 265-272.

Menzel, K. (2021), "Bootstrap with cluster-dependence in two or more dimensions," *Econometrica*, 89, 2143–2188.

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.

Santos Silva, J.M.C. and S. Tenreyro (2006), "The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641-658.

Shah, B.V., M.M. Holt and R.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.

Vertical lines separate the studies:

(1) Rose & Engel; (2) Rose; (3) Bergstrand & Baier; (3) Dutt & Traca; (5) DMV

Lower horizontal line is 1.0

Upper horizontal line is average of ratio of dyadic-robust to alternative standard errors

Dyadic standard errors are on average 5.70 times i.i.d. (default) standard errors

Dyadic standard errors are on average 4.01 times heteroskedastic-robust standard errors

Dyadic standard errors are on average 2.56 times one-way cluster on dyad-pair.

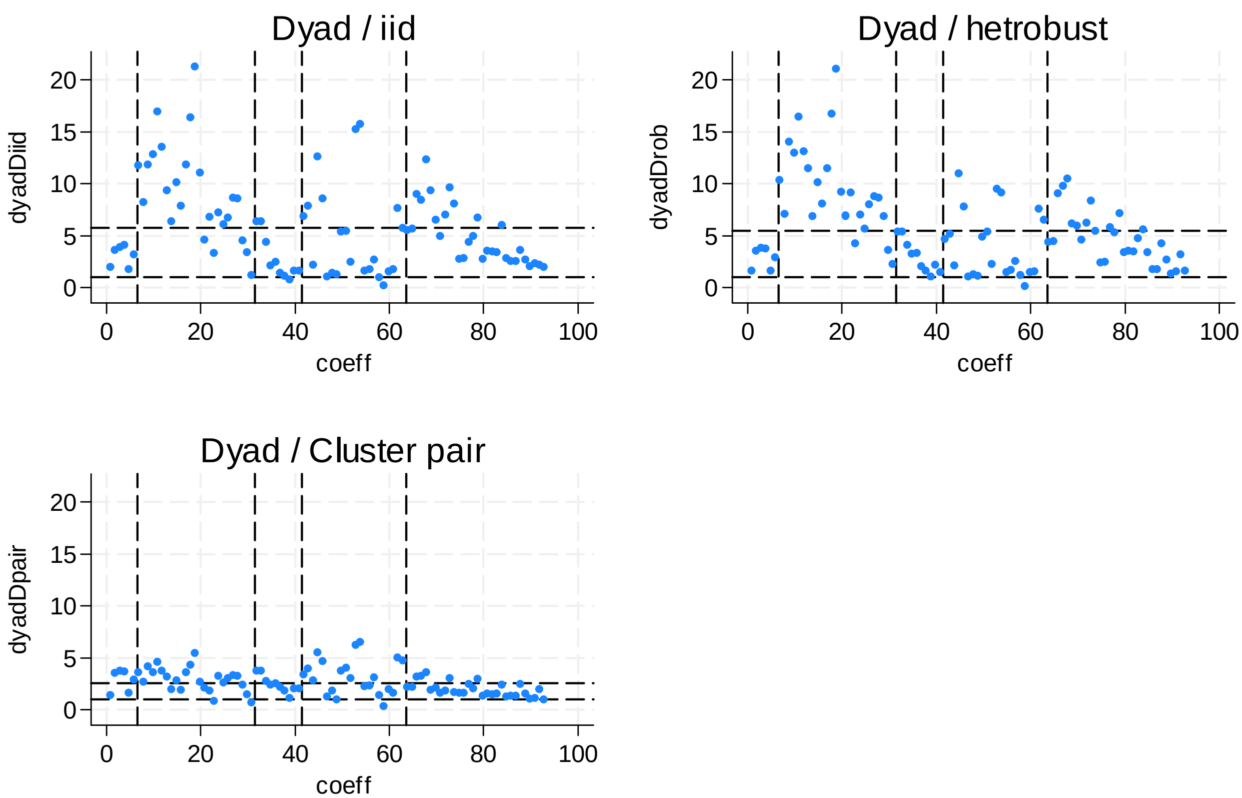

Figure 5.1: Ratio of dyadic-robust to alternative standard errors

Vertical lines separate the studies:

(1) Rose & Engel; (2) Rose; (3) Bergstrand & Baier; (3) Dutt & Traca; (5) DMV

Lower horizontal line is 1.0

Upper horizontal line is average of ratio of dyadic-robust to alternative standard errors

Dyadic standard errors are on average 1.35 times one-way cluster on country 1

Dyadic standard errors are on average 1.50 times one-way cluster on country 2

Dyadic standard errors are on average 1.05 times two-way cluster on country 1 and country 2

Dyadic standard errors are on average 1.10 times "corrected" dyadic standard errors

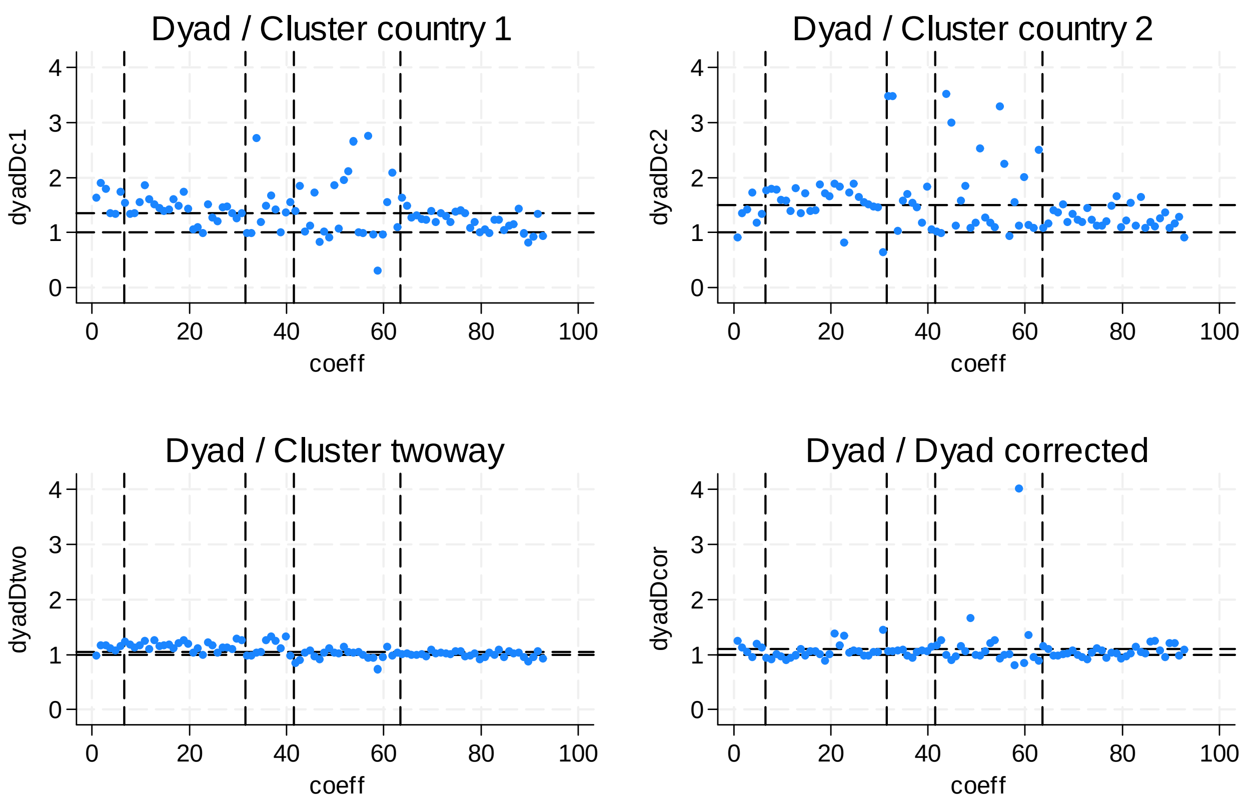

Figure 5.2: Ratio of dyadic-robust to alternative standard errors

TABLE A: SIMULATION from IID DATA (1,000 repetitions)

OLS of y_ij on intercept and scalar x_ij

Number of "Countries"	**10**		**30**		**100**	
	Mean	Std. Dev.	Mean	Std. Dev	Mean	Std. Dev
COEFF	-0.0038	0.1514	-0.0026	0.0473	-0.0005	0.0136
SE_IID	0.1527	0.0230	0.0480	0.0023	0.0142	0.0002
SE_HETROB	0.1476	0.0304	0.0480	0.0023	0.0142	0.0003
SE_CTRY1	0.1408	0.0462	0.0470	0.0077	0.0142	0.0012
SE_TWOWAY	0.1362	0.0576	0.0463	0.0110	0.0141	0.0017
SE_DYAD	0.1167	0.0551	0.0439	0.0117	0.0139	0.0018
SE_NJACK	0.1547	0.0474	0.0478	0.0072	0.0142	0.0011
REJ_IID	0.128		0.040		0.043	
REJ_HETROB	0.081		0.037		0.046	
REJ_CTRY1	0.300		0.014		0.046	
REJ_TWOWAY	0.747		0.061		0.056	
REJ_DYAD	0.976		0.058		0.048	
REJ_NJACK	0.204		0.034		0.043	

SE_IID is default standard errors assuming i.i.d. errors
SE_HETROB is heteroskedastic robust standard error
SE_CTRY1 is one-way cluster robust standard error with clustering on country 1 (i)
SE_TWOWAY is two-way cluster robust standard error
SE_DYADS is dyadic cluster-robust standard error
SE_NJACK is node-jackknife cluster-robust standard error
REJ_ is rejection rate for two-sided test that b = simulation average 0 at 5% using |t| > 1.96

TABLE B: SIMULATION from random effects error (1,000 repetitions)

OLS of y_ij on intercept and scalar x_ij

Number of "Countries"	**10**		**30**		**100**	
	Mean	Std. Dev.	Mean	Std. Dev	Mean	Std. Dev
					-	
COEFF	-47.015	7.6787	-45.6364	3.0343	0.00681	1.2582
SE_IID	2.5421	0.3645	0.8857	0.0618	0.2680	0.0083
SE_HETROB	4.9281	1.5620	2.2839	0.5873	0.7482	0.0939
SE_CTRY1	4.9927	1.6871	2.4637	0.6209	0.9567	0.1201
SE_TWOWAY	5.0093	1.9178	2.6214	0.7237	1.1240	0.1547
SE_DYAD	4.4136	1.8046	2.5997	0.7671	1.1885	0.1713
SE_NJACK	6.5367	3.0992	2.6936	0.7349	1.0534	0.1314
REJ_IID	0.581		0.574		0.663	
REJ_HETROB	0.362		0.162		0.202	
REJ_CTRY1	0.383		0.079		0.137	
REJ_TWOWAY	0.424		0.059		0.079	
REJ_DYAD	0.516		0.054		0.066	
REJ_NJACK	0.369		0.061		0.083	

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

SE_HETROB is heteroskedastic robust standard error

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

SE_TWOWAY is two-way cluster robust standard error

SE_DYADS is dyadic cluster-robust standard error

SE_NJACK is node-jackknife cluster-robust standard error

REJ_ is rejection rate for two-sided test that b = simulation average 0 at 5% using |t| > 1.96