



ARTICLE

Dyadic Clustering in International Relations

Jacob Carlson¹, Trevor Incerti²  and P. M. Aronow³ 

¹Department of Economics, Harvard University, Cambridge, MA, USA; ²Department of Political Science, University of Amsterdam, Amsterdam, Netherlands; ³Departments of Political Science, Statistics & Data Science, Biostatistics, and Economics, Yale University, New Haven, CT, USA.

Corresponding author: P. M. Aronow; Email: p.aronow@yale.edu

(Received 12 April 2022; revised 10 April 2023; accepted 18 April 2023; published online 3 October 2023)

Abstract

Quantitative empirical inquiry in international relations often relies on dyadic data. Standard analytic techniques do not account for the fact that dyads are not generally independent of one another. That is, when dyads share a constituent member (e.g., a common country), they may be statistically dependent, or “clustered.” Recent work has developed dyadic clustering robust standard errors (DCRSEs) that account for this dependence. Using these DCRSEs, we reanalyzed all empirical articles published in *International Organization* between January 2014 and January 2020 that feature dyadic data. We find that published standard errors for key explanatory variables are, on average, approximately half as large as DCRSEs, suggesting that dyadic clustering is leading researchers to severely underestimate uncertainty. However, most (67% of) statistically significant findings remain statistically significant when using DCRSEs. We conclude that accounting for dyadic clustering is both important and feasible, and offer software in R and Stata to facilitate use of DCRSEs in future research.

Keywords: dyadic data; robust standard errors; meta-science

Edited by: Prof. Xun Pang

1. Introduction

Analysis of dyadic data—data for which each observation represents a pair of units, for example, countries—is common in quantitative empirical research in international relations. Seminal theories such as the democratic trade hypothesis (Bliss and Russett 1998; Dixon and Moon 1993; Green, Kim, and Yoon 2001; Mansfield, Milner, and Rosendorff 2000), democratic peace theory (Dafoe 2011; Gartzke 2007; Imai and Lo 2021; Oneal and Russett 2001), liberal peace theory (Oneal and Russett 2001), and democratic alliance formation (Gibler and Wolford 2006; Simon and Gartzke 1996) claim empirical support (or lack thereof) from the analysis of dyadic data.

Dyadic data have a unique dependency structure—one where repeated observations of dyads are likely correlated with one another (as in panel datasets) *and* dyads that share a common member are likely correlated with one another. In particular, because multiple dyads can share members, model errors can be correlated across dyads. Only accounting for the correlations between repeated observations of dyads (e.g., by using dyad clustered standard errors or fixed effects) and ignoring correlations across dyads assumes dyad-level events are independent. This assumption contradicts substantive knowledge of the dependencies between many types of dyads common to the social sciences, such as dyads of countries. As Erikson, Pinto, and Rader (2017) note, “when a nation undergoes a pro-democratic revolution or, alternatively, when democratic leaders are deposed in a coup, the change ripples through all the nation’s many dyads.”

The idea that dyadic data exhibit a unique clustering structure that needs to be addressed methodologically in empirical work is not novel to political scientists. Random effects models

have been proposed for dyads (Cameron and Golotvina 2005), Erikson *et al.* (2017) proposed a permutation testing framework that accounts for dyadic structure, and fully parametric analyses have accounted for dyadic structure and network structure (Hays, Kachi, and Franzese Jr. 2010). Previous research has therefore determined that failing to properly account for dyadic clustering may result in underestimation of the size of standard errors and confidence intervals (Aronow, Samii, and Assenova 2015; Cameron and Miller 2014; Erikson *et al.* 2017). However, these methodological insights have not yielded a corresponding change in the way in which applied scholars conduct their research.

Recent work has developed standard error estimators that account for dyadic clustering. Building from Fafchamps and Gubert (2007), Cameron and Miller (2014), Aronow *et al.* (2015), and Tabord-Meehan (2019) have developed and studied dyadic cluster-robust standard errors (DCRSEs). Using these DCRSEs, we reanalyzed all articles published in *International Organization* over the course of just over 6 years (January 2014 to January 2020) that feature dyadic data, none of which originally implemented DCRSEs in their primary dyadic analyses. We find that DCRSEs are on average approximately twice as large as published standard errors, but that most findings remain statistically significant. While the literature therefore dramatically understates uncertainty, the estimated coefficients are usually large enough to remain statistically significant at conventional levels. To facilitate accounting for dyadic clustering in future research, we also offer software in both R and Stata that perform calculation of DCRSEs.

Our primary contributions are therefore: (1) to empirically assess the degree to which uncertainty has been underestimated in previous research due to the presence of dyadic clustering and (2) to increase the accessibility of potential solutions to dyadic clustering for applied scholars.

Note, however, that DCRSEs are not a panacea—the underlying theory and data generating process of the empirical setting should be taken into account prior to choosing an estimation strategy. When there is dependence between non-incident dyads (i.e., dyads that do not share a common member), DCRSEs will still underestimate uncertainty, just as only clustering on repeated dyads underestimates uncertainty when there is dependence across incident dyads. DCRSEs should therefore not be considered a replacement for approaches to clustering that (appropriately) account for greater amounts of dependence in the data. The reanalysis we present therefore offers a lower bound on the severity of the consequences of inadequate clustering practices in previous research, or, in other words, reveals the extent to which dyadic clustering *alone* is jeopardizing the reliability of research.

2. Why Is Dyadic Clustering a Problem?

Dyadic data contain a dependency structure whereby repeated observations of dyads are allowed to be correlated with one another, and, importantly, dyads that share any common member are *also* allowed to be correlated with one another.

It may be helpful to illustrate the substantive assumptions implicit in assuming that dyad-level events are independent with common examples from IR theory. As Cranmer and Desmarais (2016) and Maoz *et al.* (2019) note, assuming dyadic independence in WWII-era conflict implies that the conflict between Germany and Poland is unrelated to the conflict between Germany and Great Britain. This assumption is not realistic, as we know that Great Britain used the German invasion of Poland to justify its declaration of war on Germany. Similarly, Neumayer and Plümper (2010) and Poast (2016) note that bilateral trade or investment treaties are influenced by the other treaties each member may already be a part of. Assuming independence would, for example, imply that a bilateral trade deal between the US and the UK is unrelated to post-Brexit UK–EU trade negotiations. We provide an illustration using bilateral trade flows in the following section. Note that we set aside separate issues relating to analysis of time-series cross-sectional data (Beck and Katz 1995; Blackwell and Glynn 2018).

Table 1. Assumed dependence by clustering type.

Dyad	US–UK 1	US–UK 2	US–FR 1	UK–ES 1
US–UK 1	Dependent	Dependent	Independent	Independent
US–UK 2	Dependent	Dependent	Independent	Independent
US–FR 1	Independent	Independent	Dependent	Independent
UK–ES 1	Independent	Independent	Independent	Dependent
(a) Clustering by repeated dyad only				
Dyad	US–UK 1	US–UK 2	US–FR 1	UK–ES 1
US–UK 1	Dependent	Dependent	Dependent	Dependent
US–UK 2	Dependent	Dependent	Dependent	Dependent
US–FR 1	Dependent	Dependent	Dependent	Independent
UK–ES 1	Dependent	Dependent	Independent	Dependent
(b) Dyadic clustering				

2.1. Toy Example: Bilateral Trade

Consider an example in which we observe trade volume for a set of country-country-year dyads. For U.S.–U.K. trade volume, any observations that include either the US or the UK may be correlated with observations of any other dyad that also includes either the US or the UK, respectively. Table 1 illustrates the difference in assumptions about the dependencies between countries under traditional clustering by repeated dyad only, and with full dyadic clustering. Table 1 highlights that under clustering by repeated dyad only, all country groups that do not share *both* members are assumed to be independent. By contrast, under dyadic clustering, only country groups that share no members are assumed to be independent.

This clustering structure affects statistical inference. To illustrate this, suppose we were interested in characterizing the variance of the average level of commerce between the US and the UK (Y_{US-UK}) and between the US and France (Y_{US-FR}). We can compute the sample mean of their outcomes: $\hat{\mu} = \frac{1}{2}(Y_{US-UK} + Y_{US-FR})$. If we were to assume that dyads were statistically independent of one another, the variances are simply additive (i.e., there is no covariance term) and we could compute the variance of $\hat{\mu}$ as

$$\begin{aligned}
 V[\hat{\mu}]_{naive} &= V\left[\frac{1}{2}(Y_{US-UK} + Y_{US-FR})\right] \\
 &= \frac{1}{4}(V[Y_{US-UK}] + V[Y_{US-FR}]).
 \end{aligned}$$

However, with dyadic data, this may not be reasonable, especially when country-level factors are likely to impact dyadic outcomes. To see this, suppose that the true data-generating process is additive among countries, so that

$$\begin{aligned}
 Y_{US,UK} &= aX_{US} + bX_{UK} + U_{US,UK}, \\
 Y_{US,FR} &= cX_{US} + dX_{FR} + U_{US,FR},
 \end{aligned}$$

where the X and the U are independent, and all X are pairwise-independent. In this instance, the true variance of $\hat{\mu}$ is

$$\begin{aligned}
 V[\hat{\mu}] &= V\left[\frac{1}{2}(Y_{US-UK} + Y_{US-FR})\right] \\
 &= \frac{1}{4}(V[Y_{US-UK}] + V[Y_{US-FR}]) + \frac{1}{2}\text{Cov}[Y_{US-UK}, Y_{US-FR}] \\
 &= V_{naive} + \frac{1}{2}\text{Cov}[aX_{US} + bX_{UK} + U_{US,UK}, cX_{US} + dX_{FR} + U_{US,FR}] \\
 &= V_{naive} + \frac{ac}{2}V[X_{US}].
 \end{aligned}$$

When a and c are of the same sign—that is, the dyads share a positive correlation—the naive characterization of the variance understates the true sampling variability. Our setting did not involve any “network effects” between countries: the problem emerges solely from the fact that a single country (here, the US) is, mechanically, present in more than one dyad in the data.

3. Standard Error Estimation with Dyadic Data

Our approach to standard error estimation largely follows the logic of Cameron and Miller (2014), in which errors are likely correlated between dyad observations that have a country in common. To ease exposition, we consider the linear model of Y_{ijt} on regressors x_{ijt} ,

$$Y_{ijt} = \beta x_{ijt} + u_{ijt},$$

where Y_{ijt} is the level of commerce between countries i and j in time period or observation t and β is the slope that we obtain from fitting this model to the entire population. Under the exogeneity condition $E[u_{ijt} | x_{ijt}] = 0$ and the usual regularity conditions, the parameters of this model can be estimated using ordinary least squares.

The question is then how to estimate uncertainty in this model. Generically, the variance of an estimated parameter from a linear model can be represented in a symmetric form that resembles a sandwich, with two identical “bread” matrices and a “meat” matrix multiplied together in the order of bread, meat, and bread again (Aronow and Miller 2019; Davidson and MacKinnon 2004; Greene 2002):

$$V_{sandwich} = (X^T X)^{-1} X^T \Omega X (X^T X)^{-1},$$

where X denotes the matrix of regressors, and Ω is the variance–covariance matrix of model errors, such that $\Omega_{ijt, i'j't'} = E[u_{ijt} u_{i'j't'}]$. Robust sandwich estimators are formed by assuming that some elements of Ω are equal to zero, and then substituting residuals for errors, and means for expectations. Thus, the empirical variance–covariance matrix of model residuals, $\hat{\Omega}$, can be plugged into the above expression of variance to arrive at a variance estimator of the following form:¹

$$\hat{V} = (X^T X)^{-1} X^T \hat{\Omega} X (X^T X)^{-1}.$$

So long as there is not “too much” clustering (for a precise statement, see Aronow, Crawford, and Zubizarreta 2018), \hat{V} will be a consistent estimator of the variance–covariance matrix of the sampling distribution of $\hat{\beta}$. The question then becomes, what restrictions on Ω are suitable for the problem at hand? The simplest case is to assume that there is no dependence across observations whatsoever in the data, which is the assumption for non-clustered RSEs: if $i \neq i', j \neq j',$ or $t \neq t',$ then $E[u_{ijt} u_{i'j't'}] = 0,$ or that the errors are uncorrelated. With six observations, Table 2 demonstrates the variance–covariance structure assumed by the naive approach:

In practice, the naive approach is widely recognized as inappropriate in the context of international relations. It is expected that, for example, changes in bilateral trade relations will have impacts beyond the immediate dyad and across time periods.

¹Both the Eicker–Huber–White heteroskedasticity-consistent variance estimator and the Liang and Zeger cluster robust variance estimator are sandwich-like, “plug-in” variance estimators of this form, each making differing assumptions about Ω .

Table 2. Naive approach (no clustering).

$ijt / i'j't'$	111	112	211	212	221	222
111	$E[u_{111}^2]$	0	0	0	0	0
112	0	$E[u_{112}^2]$	0	0	0	0
211	0	0	$E[u_{211}^2]$	0	0	0
212	0	0	0	$E[u_{212}^2]$	0	0
221	0	0	0	0	$E[u_{221}^2]$	0
222	0	0	0	0	0	$E[u_{222}^2]$

Table 3. Clustering by repeated dyad.

$ijt / i'j't'$	111	112	211	212	221	222
111	$E[u_{111}^2]$	$E[u_{111}u_{112}]$	0	0	0	0
112	$E[u_{112}u_{111}]$	$E[u_{112}^2]$	0	0	0	0
211	0	0	$E[u_{211}^2]$	$E[u_{211}u_{212}]$	0	0
212	0	0	$E[u_{212}u_{211}]$	$E[u_{212}^2]$	0	0
221	0	0	0	0	$E[u_{221}^2]$	$E[u_{221}u_{222}]$
222	0	0	0	0	$E[u_{222}u_{221}]$	$E[u_{222}^2]$

Table 4. Dyadic clustering.

$ijt / i'j't'$	111	112	211	212	221	222
111	$E[u_{111}^2]$	$E[u_{111}u_{112}]$	$E[u_{111}u_{211}]$	$E[u_{111}u_{212}]$	0	0
112	$E[u_{112}u_{111}]$	$E[u_{112}^2]$	$E[u_{112}u_{211}]$	$E[u_{112}u_{212}]$	0	0
211	$E[u_{211}u_{111}]$	$E[u_{211}u_{112}]$	$E[u_{211}^2]$	$E[u_{211}u_{212}]$	$E[u_{211}u_{221}]$	$E[u_{211}u_{222}]$
212	$E[u_{212}u_{111}]$	$E[u_{212}u_{112}]$	$E[u_{212}u_{211}]$	$E[u_{212}^2]$	$E[u_{212}u_{221}]$	$E[u_{212}u_{222}]$
221	0	0	$E[u_{221}u_{211}]$	$E[u_{221}u_{212}]$	$E[u_{221}^2]$	$E[u_{221}u_{222}]$
222	0	0	$E[u_{222}u_{211}]$	$E[u_{222}u_{212}]$	$E[u_{222}u_{221}]$	$E[u_{222}^2]$

The most common approach is clustering by dyad, where it is assumed that errors are correlated when $i = i'$ and $j = j'$, regardless of time period. With the same six observations, section 3 demonstrates the additional clustering permitted.

In Table 3, we can see that while this approach does account for within-dyad correlations across time, all observations in the matrix of model errors that do not share both members of the dyad are still assumed to be uncorrelated (i.e., $E[u_{ijt}u_{i'j't'}] = 0$).

By contrast, our approach only assumes independence when country pairs share no members. With the same six observations, we can see the clustering permitted by the dyadic clustering approach.

In the matrix shown in Table 4—full dyadic clustering—we can now see that the only observations assumed to be independent are those for which the dyad does not share any members with other observations.

More specifically, our approach to DCRSEs follows Aronow *et al.* (2015), which, in practice, allows for the DCR variance estimator to be decomposed entirely using robust variance estimators that are

readily computed using popular statistical software packages.² The DCR variance estimator in this decomposed form is

$$\hat{V}_r = \sum_{i=1}^N \hat{V}_{c,i} - \hat{V}_D - (N-2) \hat{V}_0,$$

where \hat{V}_r is the estimated DCR variance–covariance matrix for longitudinal data; $\hat{V}_{c,i}$ is the estimated dyad-member- i -specific clustered variance–covariance matrix; \hat{V}_D is the estimated repeated-dyad clustered variance–covariance matrix; and \hat{V}_0 is the estimated heteroskedasticity-consistent variance–covariance matrix. Taking the square root of the diagonal of the DCR variance–covariance matrix yields DCRSEs for all model parameters.³

Limit theorems for dyadic data (Tabord-Meehan 2019) establish that DCRSEs may be used to form asymptotically valid confidence intervals and p -values under a normal approximation. However, DCRSEs will tend to have more sampling variability than will conventional estimators that impose more structure (e.g., standard CRSEs), meaning that they may be unreliable for inference in small samples. Although simple corrections exist (cf. Bergé 2021; Cameron *et al.* 2011), theory and further refinements for small samples (e.g., Imbens and Kolesar 2016; Pustejovsky and Tipton 2018) remain topics of ongoing inquiry for multi-way clustering problems, including the dyadic clustering problem.

Note that the above approach and its implementation only correct for interdependence between shared countries. DCRSEs are *not* sufficient—although still improve over the naive approach or clustering by repeated dyad approach—when there are interdependencies throughout the entire system (i.e., across non-incident dyads). Two common examples of such systemic interdependencies in IR are alliance formation and multilateral trade deals. In deciding to form an alliance, friendly countries i and j may be influenced by a previous alliance formed by countries i and k (i.e., the ij alliance is more attractive now that ik are also allied). However, the ij alliance could also be influenced by an alliance between unfriendly countries a and b . DCRSEs do not capture the ij – ab interdependence as there are no common dyad members. Likewise, a multilateral trade deal may also be influenced by multiple pairs of relationships that do not necessarily share members.⁴ Aronow *et al.* (2018) developed conservative estimators for the variance of least squares estimators in such cases where there is further dependence—and in such cases we expect the variance of least squares estimates to be larger than those that only take into account dyadic clustering. There is therefore still a need to understand the data-generating process and underlying theory of an empirical setting prior choosing an estimation strategy.⁵

²This “multi-way decomposition” approach was first introduced by Cameron, Gelbach, and Miller (2011), and is based on the realization that each dyad member “is the basis of its own cluster that intersects with other units’ clusters,” and that each of those dyad-member-specific clusters can be adjusted for using conventional cluster-robust variance estimators (Aronow *et al.* 2015). DCRSEs are mathematically equivalent to multi-way clustering on (undirected) dyad-member-specific clusters, though the large number of such multi-way clusters implies important analytic and computational distinctions.

³To account for rare instances where negative parameter estimate variances arise due to non-positive semi-definite variance-covariance matrices, an “eigendecomposition of the estimated variance matrix . . . [that] converts any negative eigenvalue(s) to zero” and then reassembles the matrix to “force” positive semi-definiteness is performed, in line with suggestions from Cameron *et al.* (2011) and Cameron and Miller (2014).

⁴See Poast (2016) for a more thorough discussion.

⁵Political scientists also employ network analysis approaches that model dependence beyond incident dyads. For example, Duque (2018) uses a (temporal) exponential random graph model (ERGM) to examine the determinants of embassy formation; Schoeneman, Zhu, and Desmarais (2022) use an ERGM approach to investigate FDI networks; Cho (2023) uses ERGM to examine legislative co-sponsorship networks; Kinne and Bunte (2020) use a stochastic actor oriented model to study how defense and economic cooperation are related; and Dorff, Gallop, and Minhas (2020) use latent space methods (additive and multiplicative effects) to predict conflict between groups. Some of these approaches pertain to network formation models—not models of traits conditional on networks—and therefore do not relate to error terms in the context of dyadic networks. Other approaches do pertain to modeling traits and outcomes conditional on networks, and researchers may have interest in applying or adapting these methods to their own use cases when appropriate.

4. Reanalyzing Previous Studies

To study the consequences of failing to account for dyadic clustering *in practice*, we reanalyze recent, prominent studies from the international relations literature by applying DCRSEs to estimates in empirical contexts where DCRSEs are uniquely suited to handle dyadic clustering.⁶

Specifically, we reanalyze all empirical articles published in *International Organization* over a period of 6 years (from January 2014 to January 2020) that feature dyadic data. Studies were discovered by performing a Google Scholar search of all publications mentioning any form of the word “dyad” in this period. Specifically, a search query specified the publication name “International Organization” and keywords “dyadic OR dyad OR dyads.” This process returned 70 candidate studies for reanalysis.

Each study was then assessed to determine its susceptibility to dyadic clustering. Studies were excluded from reanalysis for three primary reasons: (1) the study did not actually analyze dyadic data; (2) dyadic observations primarily featured nested relationships between dyad members, for which single- or multi-way clustering of standard errors is sufficient⁷; or (3) the dyadic observations featured a common dyad member across all observations, for example, a nominally dyadic dataset consisting entirely of U.S. bilateral trade flows. There were 22 studies not excluded by these criteria (see Supplementary Table A.2 for a list of included studies). For each of these studies, replication data were either publicly available or provided by the authors.

For each eligible study, models that featured a key explanatory variable (KEV) (Lall 2016) fit to dyadic data were identified and reanalyzed. For this reanalysis, KEVs are defined as independent variables whose parameter estimates are directly referenced in the study, or otherwise clearly pertain to the study’s stated hypotheses. Control variables are not considered to be KEVs, even if discussed or directly referenced in the study. Specifically, a model was selected for reanalysis if: (1) the model was dyadic; (2) a KEV appeared in the model; and (3) the model was not relegated to an analysis explicitly denoted as a robustness check, sensitivity analysis, or supplementary analysis, unless one of these analyses was the only dyadic analysis in the study.

The final analytic sample consisted of 691 KEVs across 174 models from 22 studies.⁸ While many studies clustered standard errors on repeated dyads, none utilized a non-parametric DCR variance estimation strategy in conducting primary analyses. Only 24 models across three studies did not use any sort of robust or clustered standard error.⁹ Three studies employ at least one model that has fixed effects for both members of the dyad under consideration.¹⁰ Of the 22 studies, 20 perform analysis on state (country) dyads and 2 perform analysis on international organization (IO) dyads.

All models were replicated and then re-estimated using the previously discussed DCR variance estimator formulated in Aronow *et al.* (2015) and implemented using an original suite of functions and commands for R and Stata, respectively. To ensure the comparability of results, the replications and reanalyses of selected models were conducted using the statistical software of origin. For the purposes of our reanalysis, dyads are assumed to be undirected, as modeling dyadic clustering based on directed dyads would require a stronger assumption about independence across observations.¹¹ Also, for the purposes of our primary reanalysis, there are no small-sample corrections made to our standard error

⁶See Carlson, Incerti, and Aronow (2023) for a complete replication package.

⁷For example, consider model parameter estimates derived from dyadic data on civil wars consisting of rebel-state dyads. Because rebel groups are commonly nested within states, correlations across observations in the data can be accounted for using only state clustered standard errors.

⁸The large number of KEVs are primarily driven by two outlier studies—Bermeo (2017) and Goemans and Schultz (2017)—which have 173 and 170 KEVs, respectively. We control for these outliers by weighting our analyses by the inverse of the total number of KEVs for a given study appearing in the analytic sample.

⁹See Supplementary Table A.3 for a summary of the results of DCR estimation for these models when repeated dyad cluster robust standard errors are added in.

¹⁰Supplementary Table A.4 shows the results of applying DCRSEs to the KEVs in these three papers.

¹¹Note that in the event that the dyadic information is indeed directional, we simply recover the directed dependence structure.

Table 5. Primary results, various levels of aggregation

Aggregation	Category	Studies	KEVs	SER ^a	Sig.→Sig. ^b	Sig.→Insig. ^b	Insig.→ Sig. ^b	Insig.→ Insig. ^b
Year	2014	4	58	1.94	0.22	0.50	0.03	0.26
	2015	5	122	1.48	0.62	0.11	0.03	0.24
	2016	5	97	2.26	0.53	0.26	0.00	0.22
	2017	3	349	1.96	0.56	0.15	0.03	0.26
	2018	2	24	1.36	0.61	0.11	0.06	0.22
	2019	1	15	1.21	0.20	0.00	0.00	0.80
	2020	2	26	1.05	0.17	0.19	0.28	0.37
Subfield	IOs/Int'l Law	5	91	2.19	0.32	0.34	0.04	0.29
	IPE	6	270	1.79	0.41	0.19	0.09	0.31
	Security	11	330	1.51	0.54	0.19	0.02	0.25
All	All	22	691	1.74	0.46	0.22	0.05	0.28

^a “SER” denotes an inverse “study frequency” weighted (ISFW) average of standard error ratios for a given level of aggregation.

^b “Sig.→Sig.,” “Sig.→Insig.,” “Insig.→Sig.,” and “Insig.→Insig.” denote ISFW proportions of *p*-values that change significance levels in these respective ways for a given level of aggregation.

estimates. Finite-sample corrections would inflate our standard error estimates to account for increased sampling variability, and thus might paint an overly pessimistic picture of the original literature.¹²

4.1. Results

To quantify the impact of neglecting dyadic clustering in prior empirical IR findings, we compare DCR re-estimated standard errors with non-DCR standard errors for all KEVs in all models. We compute a standard error ratio (SER) for all KEVs, which is the DCRSE divided by the standard error produced using the original variance estimation strategy.¹³ We also examine the precision of reanalyzed KEVs, with special attention paid to estimates that lose statistical significance at a conventional level (i.e., 5%).

The primary results of the reanalysis are presented as aggregated by year and subfield, as well as overall, in Table 5. The empirical distribution of SERs is presented in the histogram in Figure 1, and a breakdown of average SERs by study analyzed can be found in Supplementary Table A.2. The inverse “study frequency” weighted (ISFW) average of SERs across all KEVs from all studies is 1.74, and the ISFW proportion of KEVs that go from significant to insignificant at the 5% level across all studies is 0.22, where “study frequency” is the total number of KEVs for a given study appearing in the analytic sample.¹⁴ Due to the sampling variability of the standard error estimators, we see a small but not zero proportion (0.05) of KEV estimates change from statistical insignificance to significance at the 5% level.

We also examine which subfields suffer most from dyadic clustering. Table 5 shows that no subfield is immune to standard error inflation. All subfields in our sample have an average SER indicating standard

¹²In Supplementary Table A.5, we repeat our reanalysis with the small-sample correction based on guidance from, for example, Bergé (2021) and Cameron *et al.* (2011), and discover no substantive differences in findings with respect to our primary reanalysis.

¹³For example, if the original model clustered standard errors on repeated dyads, the SER is a ratio of DCR standard error to repeated dyad clustered standard error. The SER can therefore be interpreted as the inflation (or deflation) of the original standard error introduced by inter-dyad dependencies.

¹⁴Inverse “study frequency” weighting is used to ensure that multi-study averages of SERs or multi-study proportions of significant results are not dominated by studies that contribute more KEVs to the analytic sample. ISFW is equivalent to an average of averages across studies.

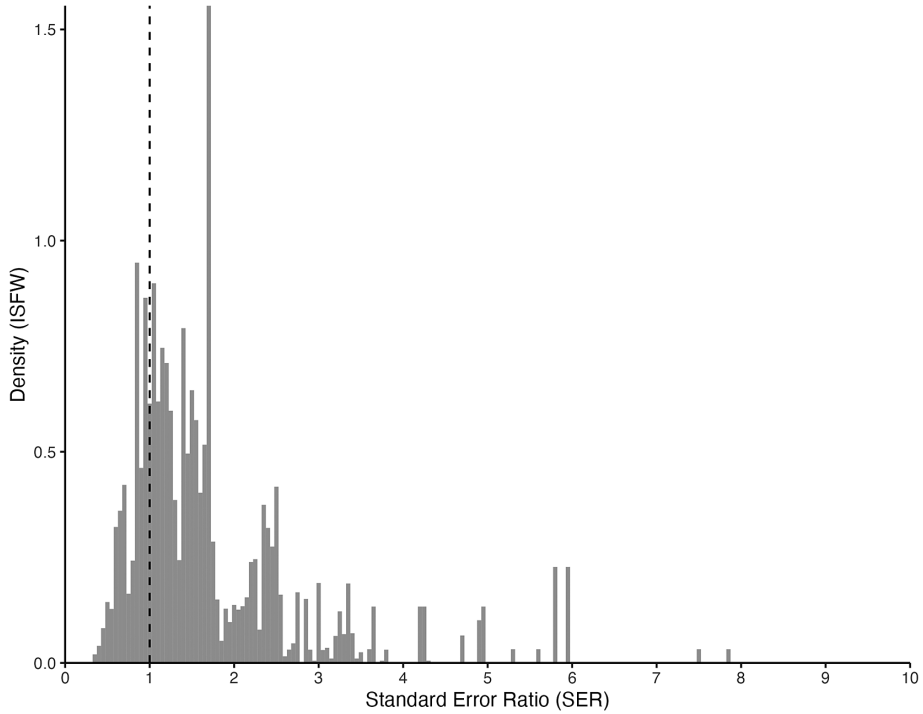


Figure 1. Histogram of key explanatory variable standard error ratios.
 Note: Key explanatory variables are independent variables whose parameter estimates are directly referenced in the study, or otherwise clearly pertain to the study’s stated hypotheses. Standard error ratios are the dyadic clustering robust standard errors divided by the standard error produced using the original variance estimation strategy.

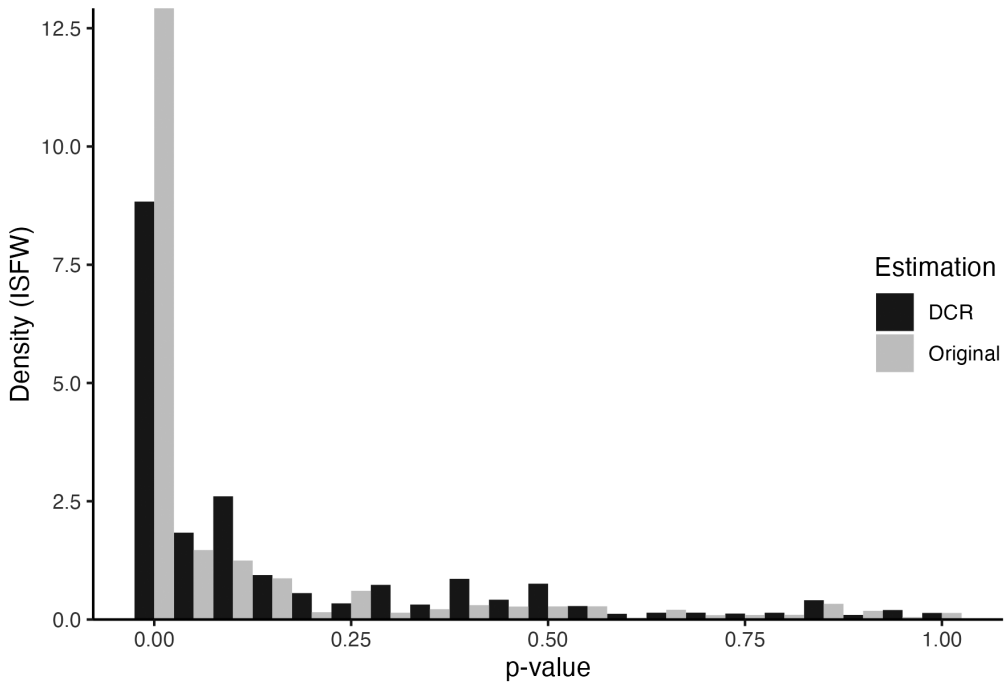


Figure 2. Histogram of *p*-values before and after reanalysis using dyadic clustering robust standard errors.

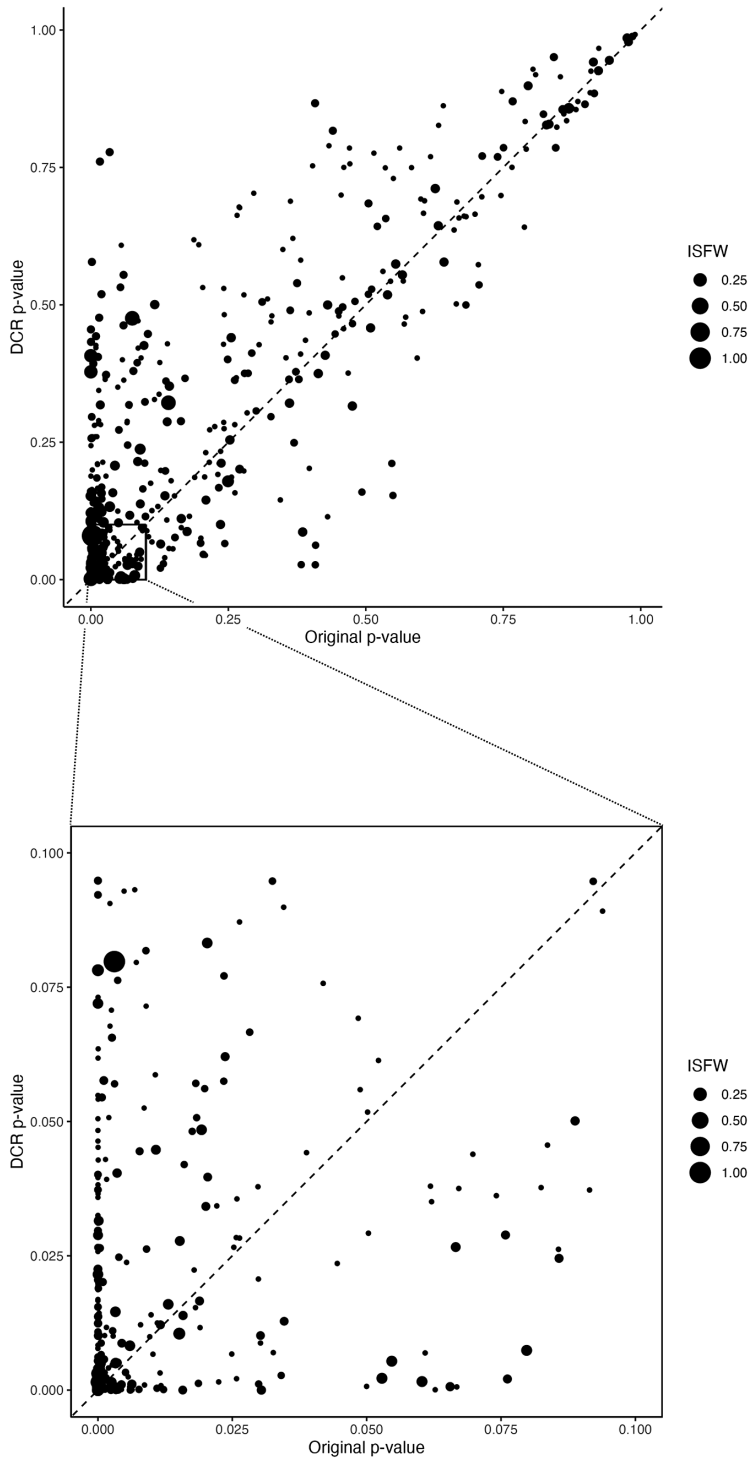


Figure 3. Scatter plots of p -values before and after reanalysis using dyadic clustering robust standard errors.
Note: The top panel depicts all p -values in the reanalysis. The bottom panel depicts p -values below 0.1. The area of each plotted data point is proportional to its inverse study frequency weight.

error inflation of more than 50%, and all subfields see 19% or more of all of their estimates change from statistically significant to insignificant at the 5% level. The “IOs and International Law” average SER indicates that DCRSEs are on average more than twice the size of originally reported standard errors, and that 34% of estimates become insignificant with DCRSEs.

For individual studies, the average SER ranges from 0.90 to 4.16, as seen in Supplementary Table A.2. On average, KEVs from 18 of 22 studies are less precise after accounting for dyadic clustering. Only 7 of 22 studies do not lose statistical significance in any KEV estimates due to DCRSE re-estimation. Three studies see half or more of their KEVs become insignificant at the 5% level upon reanalysis. Figure 2 shows how the empirical distribution of KEV p -values from all reanalyzed studies shifts due to the application of DCRSEs. Figure 3 visualizes how the precisions of individual KEV estimates change with the application of DCRSEs across all reanalyzed studies, with the area of each plotted data point being proportional to its inverse study frequency weight.

5. Conclusion

Though the need to account for the complex dependencies in dyadic data has been noted by previous researchers, these recommendations have not been commonly applied in practice. We investigated the consequences of failing to account for dyadic clustering in previous empirical research by reanalyzing all quantitative dyadic analyses in *International Organization* published in a 6-year window, thereby revealing a lower bound on the severity of the consequences of inadequate clustering practices in previous quantitative dyadic research.

We find that the standard errors associated with KEVs in reanalyzed studies are approximately half of what they would have been if calculated as DCRSEs, but that two-thirds of statistically significant KEVs remain statistically significant when using DCRSEs. Failure to compute DCRSEs therefore does not appear to have led to systematically false substantive conclusions in recent empirical IR literature, but can lead to a systematically large overestimation of the precision of estimates. In short, dyadic clustering matters, yet is not so severe as to make statistical inference using dyadic data infeasible.

However, solely accounting for dyadic clustering may not go far enough. For any of the studies we reanalyze, dependencies may exist in the data that extend across non-incident dyads, for example, due to network effects, in which case the problem may be even more severe than what our reanalysis suggests. That said, DCRSEs may be of particular interest to researchers because dyadic clustering—the clustering structure associated with dependence across dyads that share a member—is a feature of many dyadic datasets of interest to social scientists. Accordingly, we offer software in R and Stata to facilitate future analyses robust to dyadic clustering. These open-source packages implement DCRSEs for all the models in the reanalysis sample (and more), and mirror syntax familiar to users of R and Stata.

Acknowledgements. We would like to thank Austin Jang for excellent and extensive research assistance, as well as Jonathon Baron, Laurent Bergé, Forrest Crawford, Joshua Kalla, Winston Lin, Cleo O’Brien-Udry, Paul Goldsmith-Pinkham, Cyrus Samii, Beth Tipton, and three anonymous reviewers for helpful comments and conversations.

Supplementary Material. For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2023.26>.

Data Availability Statement. Replication code for this article can be accessed via Dataverse (Carlson *et al.* 2023). The statistical programming suite for DCR estimation is also available. To access the source code for the `dcr` command for Stata (version 15 or higher), clone its GitHub repository:

```
git clone git@github.com:jscarlson/stata-dcr.git
```

To access the `dcr` package for R, run:

```
devtools::install_github("jscarlson/dcr")
```

References

- Aronow, P. M., F. W. Crawford, and J. R. Zubizarreta. 2018. “Confidence Intervals for Linear Unbiased Estimators under Constrained Dependence.” *Electronic Journal of Statistics* 12 (2): 2238–2252.
- Aronow, P. M., and B. T. Miller. 2019. *Foundations of Agnostic Statistics*. Cambridge: Cambridge University Press.

- Aronow, P. M., C. Samii, and V. A. Assenova. 2015. "Cluster-Robust Variance Estimation for Dyadic Data." *Political Analysis* 23 (4): 564–577.
- Beck, N., and J. N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89 (3): 634–647.
- Bergé, L. 2021. *On Standard-Errors*. https://cran.r-project.org/web/packages/fixest/vignettes/standard_errors.html.
- Bermeo, S. B. 2017. "Aid Allocation and Targeted Development in an Increasingly Connected World." *International Organization* 71 (4): 735–766.
- Blackwell, M., and A. N. Glynn. 2018. "How to Make Causal Inferences with Time-Series Cross-Sectional Data under Selection on Observables." *American Political Science Review* 112 (4): 1067–1082.
- Bliss, H., and B. Russett. 1998. "Democratic Trading Partners: The Liberal Connection, 1962–1989." *The Journal of Politics* 60 (4): 1126–1147.
- Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2011. "Robust Inference with Multiway Clustering." *Journal of Business & Economic Statistics* 29 (2): 238–249.
- Cameron, A. C., and N. Golotvina. 2005. "Estimation of Country-Pair Data Models Controlling for Clustered Errors: With International Trade Applications." Technical report. Working Paper.
- Cameron, A. C., and D. L. Miller. 2014. "Robust Inference for Dyadic Data." Unpublished manuscript, University of California-Davis.
- Carlson, J., T. Incerti, and P. Aronow. 2023. "Replication Data for: Dyadic Clustering in International Relations. Harvard Dataverse, V1." <https://doi.org/10.7910/DVN/9I0LRQ>.
- Cho, A. 2023. "Climate Change Co-Sponsorship Networks in South Korea: Focusing on Formal and Informal Ties of Legislators." *The Journal of Legislative Studies* 29 (2): 291–311.
- Cranmer, S. J., and B. A. Desmarais. 2016. "A Critique of Dyadic Design." *International Studies Quarterly* 60 (2): 355–362.
- Dafoe, A. 2011. "Statistical Critiques of the Democratic Peace: Caveat Emptor." *American Journal of Political Science* 55 (2): 247–262.
- Davidson, R., and J. G. MacKinnon. 2004. *Econometric Theory and Methods*, Vol. 5. New York: Oxford University Press.
- Dixon, W. J., and B. E. Moon. 1993. "Political Similarity and American Foreign Trade Patterns." *Political Research Quarterly* 46 (1): 5–25.
- Dorff, C., M. Gallop, and S. Minhas. 2020. "Networks of Violence: Predicting Conflict in Nigeria." *The Journal of Politics* 82 (2): 476–493.
- Duque, M. G. 2018. "Recognizing International Status: A Relational Approach." *International Studies Quarterly* 62 (3): 577–592.
- Erikson, R. S., P. M. Pinto, and K. T. Rader. 2017. "Dyadic Analysis in International Relations: A Cautionary Tale." *Political Analysis* 22 (4): 457–463.
- Fafchamps, M., and F. Gubert. 2007. "The Formation of Risk Sharing Networks." *Journal of Development Economics* 83 (2): 326–350.
- Gartzke, E. 2007. "The Capitalist Peace." *American Journal of Political Science* 51 (1): 166–191.
- Gibler, D. M., and S. Wolford. 2006. "Alliances, Then Democracy: An Examination of the Relationship between Regime Type and Alliance Formation." *Journal of Conflict Resolution* 50 (1): 129–153.
- Goemans, H. E., and K. A. Schultz. 2017. "The Politics of Territorial Claims: A Geospatial Approach Applied to Africa." *International Organization* 71 (1): 31–64.
- Green, D. P., S. Y. Kim, and D. H. Yoon. 2001. "Dirty Pool." *International Organization* 55 (2): 441–468.
- Greene, W. H. 2002. *Econometric Analysis*. New York: Prentice Hall.
- Hays, J. C., A. Kachi, and R. J. Franzese, Jr. 2010. "A Spatial Model Incorporating Dynamic, Endogenous Network Interdependence: A Political Science Application." *Statistical Methodology* 7 (3): 406–428.
- Imai, K., and J. Lo. 2021. "Robustness of Empirical Evidence for the Democratic Peace: A Nonparametric Sensitivity Analysis." *International Organization* 75 (3): 901–919.
- Imbens, G. W., and M. Kolesar. 2016. "Robust Standard Errors in Small Samples: Some Practical Advice." *Review of Economics and Statistics* 98 (4): 701–712.
- Kinne, B. J., and J. B. Bunte. 2020. "Guns or Money? Defense Co-Operation and Bilateral Lending as Coevolving Networks." *British Journal of Political Science* 50 (3): 1067–1088.
- Lall, R. 2016. "How Multiple Imputation Makes a Difference." *Political Analysis* 24 (4): 414–433.
- Mansfield, E. D., H. V. Milner, and B. P. Rosendorff. 2000. "Free to Trade: Democracies, Autocracies, and International Trade." *American Political Science Review* 94 (2): 305–321.
- Maoz, Z., P. L. Johnson, J. Kaplan, F. Ogunkoya, and A. P. Shreve. 2019. "The Dyadic Militarized Interstate Disputes (MIDs) Dataset Version 3.0: Logic, Characteristics, and Comparisons to Alternative Datasets." *Journal of Conflict Resolution* 63 (3): 811–835.
- Neumayer, E., and T. Plümpner. 2010. "Spatial Effects in Dyadic Data." *International Organization* 64 (1): 145–166.
- Oneal, J. R., and B. Russett. 2001. "Clear and Clean: The Fixed Effects of the Liberal Peace." *International Organization* 55 (2): 469–485.

- Poast, P. 2016. "Dyads Are Dead, Long Live Dyads! The Limits of Dyadic Designs in International Relations Research." *International Studies Quarterly* 60 (2): 369–374.
- Pustejovsky, J. E., and E. Tipton. 2018. "Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models." *Journal of Business & Economic Statistics* 36 (4): 672–683.
- Schoeneman, J., B. Zhu, and B. A. Desmarais. 2022. "Complex Dependence in Foreign Direct Investment: Network Theory and Empirical Analysis." *Political Science Research and Methods* 10 (2): 243–259.
- Simon, M. W., and E. Gartzke. 1996. "Political System Similarity and the Choice of Allies: Do Democracies Flock Together, or Do Opposites Attract?" *Journal of Conflict Resolution* 40 (4): 617–635.
- Tabord-Meehan, M. 2019. "Inference with Dyadic Data: Asymptotic Behavior of the Dyadic-Robust *t*-Statistic." *Journal of Business & Economic Statistics* 37 (4): 671–680.