Advance Access publication July 29, 2015

Political Analysis (2015) 23:564–577
doi:10.1093/pan/mpv018

Cluster–Robust Variance Estimation for Dyadic Data

Peter M. Aronow
Department of Political Science, Yale University, 77 Prospect Street, New Haven, CT 06520
e-mail: peter.aronow@yale.edu

Cyrus Samii
Department of Politics, New York University, 19 West 4th Street, New York, NY 10012
e-mail: cds2083@nyu.edu (corresponding author)

Valentina A. Assenova
School of Management, Yale University, 165 Whitney Avenue, New Haven, CT 06520
e-mail: valentina.assenova@yale.edu

Edited by Jens Hainmueller

Dyadic data are common in the social sciences, although inference for such settings involves accounting for a complex clustering structure. Many analyses in the social sciences fail to account for the fact that multiple dyads share a member, and that errors are thus likely correlated across these dyads. We propose a non-parametric, sandwich-type robust variance estimator for linear regression to account for such clustering in dyadic data. We enumerate conditions for estimator consistency. We also extend our results to repeated and weighted observations, including directed dyads and longitudinal data, and provide an implementation for generalized linear models such as logistic regression. We examine empirical performance with simulations and an application to interstate disputes.

1 Introduction

Dyadic data are central in social science applications ranging from international relations to "speed dating."[1] A challenge with dyadic data is to account for the complex dependency structure that arises due to the connections between dyad members. For instance, in a study of international conflict, a change in leadership in one country may affect relations with all countries with which that country is paired in the data, thereby inducing a correlation between these dyadic observations. This generates a cluster of dependent observations associated with that country. As leadership changes occur in multiple countries, the correlations emanating from each of these countries will overlap into a web of interwoven clusters. We refer to such interwoven dependency in dyadic data as "dyadic clustering." By ignoring the dyadic clustering, the analysis would take the dyad-level changes emanating from a single leadership change as independently informative events, rather than a single, clustered event. An analysis that only accounts for correlations in repeated observations of dyads (whether by clustering standard errors or using dyad fixed effects) would fail to account for such inter-dyad correlation.

Statistical analysis of dyads typically estimates how dyad-level outcomes (e.g., whether there is open conflict between countries or the decision for one person to ask another on a date) relate to characteristics of the individual units as well as to the dyad as a whole (e.g., measures of proximity

Authors' note: The authors thank Neal Beck, Allison Carnegie, Dean Eckles, Donald Lee, Winston Lin, Kelly Rader, Olav Sorenson, the *Political Analysis* editors, and two reviewers for helpful comments. They thank Jonathan Baron and Lauren Pinson for research assistance. Supplementary materials for this article are available on the *Political Analysis* Web site. Replication materials are available on the *Political Analysis* Dataverse (https://dataverse.harvard.edu/dataverse/pan).

[1]For example, in the past five years the *American Political Science Review*, *American Journal of Political Science*, and *International Organization* have together published sixty-two papers that include dyadic analyses.

© The Author 2015. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oup.com

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

between units). The usual approach is to regress the dyad-level outcome on unit- and dyad-level predictors. Due to dyadic clustering, the observations contributing to such an analysis are not independent. Failure to account for dyadic clustering may result in significance tests or confidence intervals with sizes that are far too small relative to the true distribution of the parameters of interest.

We establish sufficient conditions for the consistency of a non-parametric sandwich estimator (Huber 1967; White 1980a) for the variance of regression coefficients under dyadic clustering. Cluster–robust sandwich estimators are common for addressing dependent data (Liang and Zeger 1986; Angrist and Pischke 2009, chap. 8). Cameron, Gelbach, and Miller (2011) provide a sandwich estimator for "multi-way" clustering, accounting, for example, for clustering between people by geographic location and age category. We extend these methods to dyadic clustering, accounting for the fact that dyadic clustering does not decompose neatly into a few crosscutting and disjoint groups of units; rather, each *unit* is the basis of its own cluster that intersects with other units' clusters. Fafchamps and Gubert (2007, Equation (2.5)) propose a sandwich estimator for dyadic clustering that is very similar to what we propose below. Their derivation is constructed through analogy to the results of Conley (1999). However, neither paper establishes conditions for consistency under dyadic clustering. We establish such consistency conditions. We also provide extensions to the longitudinal case and generalized linear models such as logistic regression. We evaluate performance with simulations and a reanalysis of a classic study on interstate disputes (Russett and Oneal 2001). The appendix generalizes to weighted data and generalized linear models, and the supplementary information provides another illustration with a speed dating experiment (Fisman et al. 2006).

Current statistical approaches to handling dyadic clustering include the use of parametric restrictions in mixed-effects models (Hoff 2005; Kenny, Kashy, and Cook 2006; Gelman and Hill 2007), the spatial error-lag model (Beck, Gleditsch, and Beardsley 2006), or permutation inference for testing against sharp null hypotheses (Erikson, Pinto, and Rader 2014). These approaches have important limitations that our approach overcomes. First, mixed-effects models and spatial lag models impose a parametric structure to address the clustering problem. This makes them sensitive to misspecification of the conditional mean (i.e., deviations between the data and the linearity assumption). Our approach is robust in that it provides asymptotically valid standard errors even under such misspecification. This is valuable in itself to the extent that models are typically approximations (Hubbard et al. 2010; Buja et al. 2014), although readers should not take this to mean that they are free from the obligation to fit as good an approximation as possible. It is also valuable in providing a reliable benchmark to use in evaluating model specification (White 1980a, 1981; King and Roberts 2015). Second, Erikson, Pinto, and Rader's (2014) solution of non-parametric randomization inference does not provide a procedure for obtaining valid confidence intervals, while our procedure does. Third, all three of these alternatives require considerable computational power, possibly exceeding available computational resources even for data analyses that are common in international relations or network analysis. We demonstrate below that a mixed-effects or spatial error-lag approach is infeasible in a typical international relations example. In contrast, our proposed estimator is easy to compute. The variance estimation methods we develop are a natural complement to non- and semi-parametric approaches to regression with dyadic data (Green, Kim, and Yoon 2001).

2 Setting

We work within the "agnostic" (Angrist and Imbens 2002; Lin 2013) or "assumption lean" (Buja et al. 2014) regression framework developed by Angrist and Pischke (2009), Goldberger (1991), and White (1980b, 1982).[2] We begin with a cross-section of undirected dyads and derive the basic

[2]Our results would hold under a regression model approach along the lines of Greene (2008) or Davidson and MacKinnon (2004). The value of using the "agnostic regression" approach is that the derivation is robust to violations of the assumption that the regression specification is correct for the conditional expectation. Nevertheless, this should not be taken as an endorsement for ignoring misspecification or assuming that linear approximations with robust standard errors can cure misspecification problems (King and Roberts 2015). This article leaves aside issues of causal identification and focuses only on estimation and statistical inference. Nonetheless, the issues discussed here would be relevant for studies that attempt to estimate causal effects with dyadic data.

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

convergence results for this case. Below we extend these results to repeated dyads, which covers directed dyads and longitudinal data. Proofs are in the Appendix.

Begin with a large population from which we take a random (i.i.d.) sample of units, with the sample members indexed by $i = 1, ..., N$ and grouped into $D = \binom{N}{2} = N(N-1)/2$ dyads. Pairs of unit indices within each dyad map to dyadic indices, $1, ..., D$, through the index function $d(i, j) = d(j, i)$, with an inverse correspondence $s(d(i, j)) = \{i, j\}$, and we assume no $d(i, i)$ type dyads. Consider a linear regression of $Y_{d(i,j)}$ on a k-length column vector of regressors $X_{d(i,j)}$ (which includes the constant):

$$Y_{d(i,j)} = X'_{d(i,j)}\beta + \epsilon_{d(i,j)},$$

where β is the slope that we would obtain if we could fit this model to the entire population, allowing for possible non-linearity in the true relationship between $\mathrm{E}[Y_{d(i,j)}|X_{d(i,j)}]$ and $X_{d(i,j)}$, and $\epsilon_{d(i,j)}$ is the corresponding population residual.[3] To lighten notation, for the remainder of the discussion we use $d = d(i, j)$.

Define the sample data as $\mathbf{X} = (X_1 ... X_D)'$ and $Y = (Y_1 ... Y_D)'$. The ordinary least squares (OLS) estimator is

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y,$$

with residual $e_d = Y_d - X'_d\hat{\beta}$. Since the values of Y_d and X_d are determined by the characteristics of the units i and j, ϵ_d and $\epsilon_{d'}$ for dyads containing either unit i or j are allowed to be correlated by construction. However, by random sampling of units, $\mathrm{Cov}(\epsilon_d, \epsilon_{d'}) = 0$ for all dyads that do not share a member. The number of pairwise dependencies between i and j for which $\mathrm{Cov}(\epsilon_d, \epsilon_{d'}) \neq 0$ is $O(N^3)$. Let the support for X_d and Y_d be bounded in a finite interval and assume the usual rank conditions on $\mathbf{X}$ (Chamberlain 1982; White 1984).

Lemma 1. *Suppose data and $\hat{\beta}$ as defined above. As $N \to \infty$, the asymptotic distribution of $\sqrt{N}(\hat{\beta} - \beta)$ has mean zero and variance*

$$V = \frac{N}{D^2}\mathrm{E}[X_d X'_d]^{-1}\mathrm{Var}\left[\sum_{d=1}^{D} X_d\epsilon_d\right]\mathrm{E}[X_d X'_d]^{-1}, \tag{1}$$

with

$$\mathrm{Var}\left[\sum_{d=1}^{D} X_d\epsilon_d\right] = \sum_{d=1}^{D}\left(\underbrace{\mathrm{E}\left\{X_d X'_d \mathrm{Var}[\epsilon_d|\mathbf{X}]\right\}}_{A} + \underbrace{\sum_{d' \in \mathscr{S}(d)}\mathrm{E}\left\{X_d X'_{d'}\mathrm{Cov}[\epsilon_d, \epsilon_{d'}|\mathbf{X}]\right\}}_{B}\right), \tag{2}$$

where $\mathscr{S}(d) = \{d' \neq d\} \cap \{d' : s(d) \cap s(d') \neq \emptyset\}$, the set of dyads other than d that share a member from d.

A is the dyad-specific contribution to the variance, and B is the contribution due to inclusion of common units in multiple dyads. Note that these features of the distribution of $\sqrt{N}(\hat{\beta} - \beta)$ establish the consistency of $\hat{\beta}$ as well, which is not surprising given standard results for the consistency of OLS coefficients on dependent data (White 1984).

3 Identification and Estimation

Given data as defined above, we examine the properties of a plug-in variance estimator that is analogous to the sandwich estimators defined for heteroskedastic or clustered data (White 1980a;

[3] If $\mathrm{E}[Y_{d(i,j)}|X_{d(i,j)}]$ is non-linear in $X_{d(i,j)}$, the population residual will itself vary in $X_{d(i,j)}$, which undermines inference that assumes that $\epsilon_{d(i,j)}$ is independent of $X_{d(i,j)}$. The agnostic regression approach avoids such an assumption.

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

Liang and Zeger 1986; Arellano 1987; Cameron, Gelbach, and Miller 2011). We establish sufficient conditions for consistency of the plug-in variance estimator. We consider the cross-sectional case and repeated observations case. The Appendix also contains extensions to weighted observations and generalized linear models.

Proposition 1. *Define the variance estimator*

$$\hat{V} = (\mathbf{X}'\mathbf{X})^{-1}\sum_{d=1}^{D}\left(X_d X_d' e_d^2 + \sum_{d' \in \mathscr{S}(d)} X_d X_{d'}' e_d e_{d'}\right)(\mathbf{X}'\mathbf{X})^{-1}. \tag{3}$$

Under the conditions of Lemma 1 and if X_{d} *and* e_d *have bounded support with nonzero second moments and cross-moments, then as* $N \to \infty$,

$$N\hat{V} - V \xrightarrow{p} 0.$$

The proposition indicates that $\hat{V}$ provides a consistent estimator to characterize the true asymptotic sampling distribution of the regression coefficients, $\hat{\beta}$. Standard error estimates for $\hat{\beta}$ are obtained from the square roots of the diagonals of $\hat{V}$. The assumption of bounded support for X_d and e_d merely rules out situations that are unlikely to arise in real-world data anyway (e.g., where the mass of the data pushes out toward infinity as N grows).

Repeated dyad observations are common in applied settings. For example, the data may include multiple observations for dyads over time. Dyadic panels are very common in studies of international relations. Or, if the dyadic information is directional, then the data will contain two observations for each dyad, with one observation capturing outcomes that go in the i to j direction, and the other capturing outcomes that go in the j to i direction. This is conceptually distinct from repeated observations over time. But if there are dyadic dependences for both senders and receivers of a directed dyad, then the dependence structure for a pair of directed i–j dyads will be the same as if we had repeated observations of an undirected i–j dyad. The results above translate straightforwardly to the repeated dyads setting. Formally, suppose that for each dyad d observations are indexed by $t = 1, ..., T(d)$, where the $T(d)$ values are fixed and finite. Let (Y_{dt}, X_{dt}') represent the data for observation t from dyad d,

$$Y_d = \begin{pmatrix} Y_{d1} \\ \vdots \\ Y_{dT(d)} \end{pmatrix} \quad \text{and} \quad X_d = \begin{pmatrix} X_{d1}' \\ \vdots \\ X_{dT(d)}' \end{pmatrix},$$

and $Y = (Y_1' \ldots Y_d' \ldots Y_D')'$ and $\mathbf{X} = (X_1' \ldots X_d' \ldots X_D')'$, that is, the stacked Y_d vectors and X_d matrices, and let β_r and $\hat{\beta}_r$ denote, respectively, the population slope and OLS estimator as defined above but now applied to the repeated dyads data.

Corollary 1. *For the repeated dyads case, assume the same conditions as in Proposition 1 and consider the following variance estimator*:

$$\hat{V}_r = (\mathbf{X}'\mathbf{X})^{-1}\sum_{d=1}^{D}\left(X_d' e_d e_d' X_d + \sum_{d' \in \mathscr{S}(d)} X_d' e_d e_{d'}' X_{d'}\right)(\mathbf{X}'\mathbf{X})^{-1}.$$

Then, as $N \to \infty$,

$$N\mathrm{Var}[\hat{\beta}_r - \beta_r] - V_r \xrightarrow{p} 0,$$

and

$$N\hat{V}_r - V_r \xrightarrow{p} 0,$$

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

where

$$V_r = \frac{N}{\left[\sum_{d=1}^{D} T(d)\right]^2} \mathrm{E}[X_d X_d']^{-1} \mathrm{E}\left(\mathrm{E}[X_d' \mathrm{Cov}[\epsilon_d|\mathbf{X}] X_d] + \sum_{d' \in \mathscr{S}(d)} \mathrm{E}[X_d' \mathrm{Cov}[\epsilon_d, \epsilon_{d'}|\mathbf{X}] X_{d'}]\right) \mathrm{E}[X_d X_d']^{-1}.$$

Some remarks are in order when it comes to the repeated observations case. First, our analysis of the repeated observations case takes the $T(d)$ values to be fixed and finite and demonstrates consistency as N grows. In a cross-national study, this would mean that one should pay the most attention to the number of countries, rather than the amount of time, that are in the data. Under fixed N and growing $T(d)$ further assumptions would be required for consistency, such as serial correlation of fixed order or, more generally, strong mixing over time. Second, with repeated observations, a common strategy for identifying effects is to use dyad-specific fixed effects (Green, Kim, and Yoon 2001). With fixed and finite $T(d)$, this does not introduce any new complications. As in Arellano (1987) for the case of serial correlation with fixed effects, the results for dyadic clustering translate directly to centered data, and dyadic fixed-effects regression amounts to centering the data on dyad-specific means. With fixed N and growing $T(d)$, however, additional assumptions are needed for consistency (Hansen 2007; Stock and Watson 2008).

For efficient computation, we follow Cameron, Gelbach, and Miller (2011) to perform a "multiway decomposition" of the dyadic clustering structure:

Proposition 2. *We have the following algebraic equivalence*:

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

where

$$\hat{V}_{C,i} = (\mathbf{X}'\mathbf{X})^{-1} \hat{\Sigma}_{C,i} (\mathbf{X}'\mathbf{X})^{-1},$$
$$\hat{V}_D = (\mathbf{X}'\mathbf{X})^{-1} \hat{\Sigma}_D (\mathbf{X}'\mathbf{X})^{-1},$$
$$\hat{V}_0 = (\mathbf{X}'\mathbf{X})^{-1} \hat{\Sigma}_0 (\mathbf{X}'\mathbf{X})^{-1},$$

and

$$\hat{\Sigma}_{C,i} = \sum_{j \neq i} \sum_{k \neq i} X_{d(i,j)}' e_{d(i,j)} e_{d(i,k)}' X_{d(i,k)} + \sum_{j,k \neq i} \sum_{t=1}^{T(d(j,k))} X_{d(j,k)t} X_{d(j,k)t}' e_{d(j,k)t}^2,$$

$$\hat{\Sigma}_D = \sum_{d=1}^{D} X_d' e_d e_d' X_d, \text{ and } \hat{\Sigma}_0 = \sum_{d=1}^{D} \sum_{t=1}^{T(d)} X_{dt} X_{dt}' e_{dt}^2.$$

$\hat{V}_{C,i}$ is the usual asymptotically consistent cluster–robust variance estimator (with no degrees-of-freedom adjustment) that clusters all dyads containing unit i and assumes all other observations to be independent. $\hat{V}_D$ is the same cluster robust estimator but clustering all repeated dyad observations. $\hat{V}_0$ is the usual asymptotically consistent heteroskedasticity robust (HC) variance estimator that assumes all observations (even within repeated dyad groupings) are independent. To understand this decomposition, note that dyadic clustering involves clustering on each of N units. But in summing the contributions from unit-specific clusters (the $\hat{V}_{C,i}$s), we double count the dyad contributions ($\hat{V}_D$) and add in the independent contributions (V_0) N times. We can correct for this by subtracting $\hat{V}_D$, which also removes the V_0 component, and then subtracting $N-2$ of the V_0 components. (The cross-section cases simply set $\hat{V}_D = \hat{V}_0$.) This decomposition shows that one can compute the dyadic cluster robust estimator using readily available robust standard error commands.[4]

[4]Code for implementation in R is available from the replication archive posted to the Dataverse page references in the acknowledgments note.

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

Our dyadic cluster robust variance estimator allows one to perform valid inference under less restrictive dependence assumptions that are used to identify random effects or spatial error lag models. Our multi-way decomposition also shows that it is computationally much simpler than those alternatives. The latter point is relevant when one has many units or time periods, in which case random effects and spatial error lag models, despite their restrictions, may be infeasible with current computational resources given the challenges of evaluating a likelihood with as many dependencies as emerge in a dyadic analysis.

4 Simulation Evidence

We use Monte Carlo simulations to evaluate the finite sample properties of the proposed estimators under the cross-sectional and repeated dyads settings. We suppose that population values obey the following:

$$Y_{d(i,j)t} = \beta_0 + \beta_1 |X_i - X_j| + \underbrace{\alpha_i + \alpha_j + \nu_{d(i,j)t}}_{\epsilon_{d(i,j)t}}, \tag{4}$$

where $Y_{d(i,j)t}$ is the tth observed outcome for the dyad that includes units i and j, X_i, X_j, α_i, α_j, and $u_{d(i,j)t}$ are independent draws from standard normal distributions, and the compound error, $\epsilon_{d(i,j)t} = \alpha_i + \alpha_j + \nu_{d(i,j)t}$, is unobserved. In the cross-sectional case, we only observe one outcome per dyad, so $t = 1$ for all observations (i.e., the t subscript is extraneous for the cross-sectional case). In the repeated observations case, we have two observations per dyad, so $t = 1, 2$ for all dyads. We fix $\beta_0 = 0$ and $\beta_1 = 1$. We use OLS to estimate $\hat{\beta}_0$ and $\hat{\beta}_1$. In the supplementary information we show results for mixed-effects models as well. The fact that X_i and α_i are constant across dyads that include unit i (same for j) implies non-zero intra-class correlation in both X and ϵ among sets of dependent dyads, in which case ignoring the dependence structure will tend to understate the variability in $\hat{\beta}_0$ and $\hat{\beta}_1$. This is the dyadic version of Moulton's (1986) problem.

Results from 500 simulation runs are shown in Fig. 1. The x-axis shows the sample size, where we show results for samples with 20, 50, 100, and 150 units (implying 190, 1225, 4950, and 11,175 dyads, respectively). The y-axis is on the scale of the standard error of the coefficients. The black diamonds plot the empirical standard errors for each of the sample sizes (i.e., the standard deviation of the simulation replicates of the coefficient estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$). The black box plots show the distribution of our proposed standard error estimator. The gray box plots show, in the top figures, the distribution of the "HC2" HC variance estimator, which does not account for either dyadic or repeat observation clustering (White 1980a; MacKinnon and White 1985), and for the bottom figures, the "cluster robust" analog of White's (1980a) estimator accounting for dependence in repeated dyad observations (Liang and Zeger 1986; Arellano 1987).

Convergence of our proposed standard errors based on $\hat{V}$ and $\hat{V}_r$ (black box plots) to the true standard error (black diamonds) is quick. The alternative estimators, which do not account for dyadic clustering, grossly understate the variability in the coefficient estimates. These results suggest that in finite samples from data-generating processes that resemble our simulations, the proposed estimator is quite reliable for inference on OLS coefficients so long as the sample size is on the order of 50–100 units. Most applications in political science operate with samples of at least this size. Changes to the shape of the error distributions (e.g., uniform or bimodal) yield the same results.[5]

In the supplementary information, we demonstrate comparisons between mixed-effects models and our "robust" estimator.[6] Specifically, we compare OLS with dyadic cluster robust standard errors to a two-way random effects model that incorporates normal random effects for each of the N units (Gelman and Hill 2007). The data-generating process captured above (i.e., expression 4

[5]The simulation code, which allows one to choose arbitrary data distributions, is available from the replication archive posted to the Dataverse page referenced in the acknowledgements note.

[6]We sought also to include a comparison to spatial error lag models, but the models failed to converge with $N \geq 100$. We note that in the original Beck, Gleditsch, and Beardsley (2006) paper, they limited their analysis of dyadic data to small cross-sections. Our experience suggests that it may be infeasible to fit such models with substantially larger data sets.

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

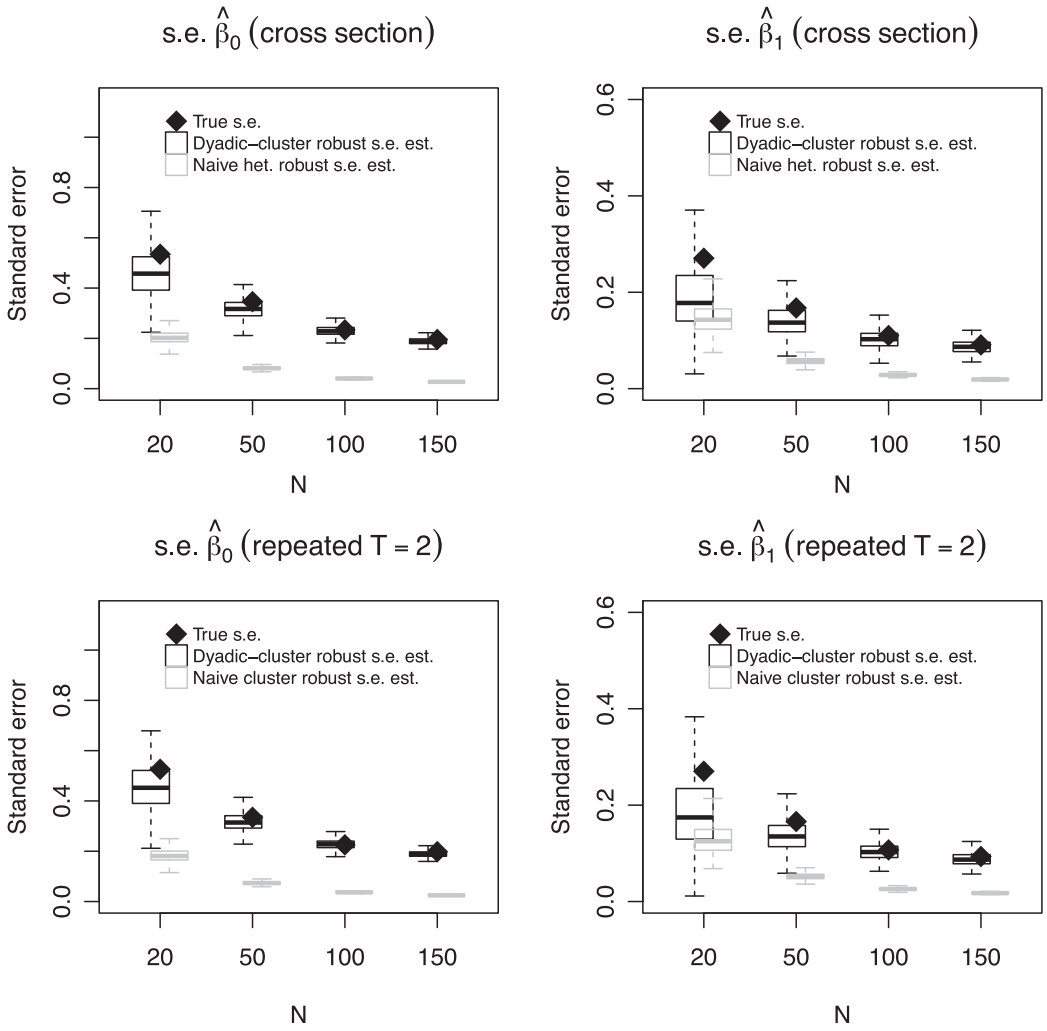


Fig. 1 Convergence of the standard error estimator as N goes from 20 to 150 (implying number of dyads goes from 190 to 11,175) for a cross-section of dyads (top) and repeated dyad observations ($T = 2$ for all dyads). Black box plots show the distribution of standard error estimates using the proposed dyadic cluster robust estimator. In the top two figures, gray box plots show standard error estimates from a HC estimator, while in the bottom two figures, gray box plots show standard error estimates from a "naive" cluster–robust estimator that clusters by dyad over repeated observations (but does not cluster by units across dyads).

with normal data) corresponds precisely to the assumptions of a two-way normal random effects model, and so it is no surprise that we find this model to be more efficient than OLS and to produce consistent standard errors. However, as discussed above, such models are sensitive to misspecification. For example, in fitting a linear approximation model to mildly non-linear data, the dyadic cluster robust standard errors continue to be consistent for the coefficients of the linear approximation. The degree of misspecification in this case is not extreme (as illustrated in the Appendix) and resembles the kind of approximations that are common in applied work (Buja et al. 2014). Nonetheless, the random effects standard errors are inconsistent and substantially anti-conservative. Such sensitivity to misspecification is one problem for the random effects model. Another problem arises when there is unobserved country-level heterogeneity that could confound estimates of the coefficients of interest. In such situations, our dyadic cluster robust estimator would be a natural complement to a fixed effects analysis, analogous to what Arellano (1987) proposes for non-dyadic repeated observation settings. The third problem arises in computational feasibility, an issue that arises in our application, to which we now turn.

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

5 International Militarized Disputes Application

Our application is based on the classic study by Russett and Oneal (2001) on the determinants of international militarized conflicts. They study how the likelihood of a militarized dispute between states in a dyad relates to various dyad-level attributes, including an indicator of whether the two states are formal allies, the log of the ratio of an index of military capabilities, the lowest score in the dyad on a democracy index, the ratio of the value of trade and the larger GDP of the two states, an indicator of whether both states are members of a common international organization, an indicator of whether the states are geographically noncontiguous, the log of the geographic (Euclidean) distance between the two states' capitals, and an indicator of whether both states are "minor powers" in the international system. Dyadic clustering could arise in many ways with these data, for example if a country entered into an alliance, thereby changing the joint alliance indicators, or if the military capabilities of a country changed, thereby changing the power ratios.

We replicate Russett and Oneal's primary analysis as reported in their Table A5.1. They use annual data on 146 states in the international system paired into 1178 dyads (out of 10,585 possible) and observed for as few as one and as many as 90 years between 1885 and 1991, for a total of 39,988 observations. In their original analysis, Russett and Oneal fit a GEE model assuming AR(1) errors within dyads over time (Zorn 2001). Table 1 shows the results of our reanalysis. Columns (1) and (2) replicate the original published results. Columns (3)–(6) show coefficients and various standard error estimates for a simple (pooled) logistic regression. There is little difference in the coefficient estimates from the GEE-AR(1) model when compared with the simple logistic regression, so we focus on the standard error estimates. Column (4) contains estimates that account only for dyad-year heteroskedasticity, Column (5) also accounts for arbitrary dependence over time for each dyad, and then Column (6) also accounts for dyadic clustering. Accounting for the dyadic and repeated-observation clustering results in standard error estimates that are sometimes an order of magnitude larger than what we obtain when we ignore all clustering and also considerably larger than what one would estimate were one to account for repeated dyads clustering but ignore the inter-dyadic clustering. The latter result is also relevant when comparing the standard errors from the original GEE-AR(1) model, as they resemble the estimates in Column (5). In their original analysis, Russett and Oneal found that all of the predictors had a statistically significant relationship to the likelihood of militarized conflict. But when one takes into account dyadic clustering, the coefficient for international organizations would fail to pass a conventional significance test (e.g., $t > 1.64$).

We attempted also to study how inferences might change with a random effects approach and a spatial error lag approach. The appropriate random effects model would be a three-way random effects model to account for each state in a dyad as well as the dyad as a whole over time. We attempted to fit three-way random effects logit models using the `lmer` package in `R` and the `melogit` function in Stata. We posted the jobs to a university cluster with 24 Intel Xeon cores and 48 gigabytes of RAM per node. In neither case did the random effects models converge. In fact, in neither case could we obtain estimates of any form after days on the cluster. For the spatial error lag model, we sought to implement both the spatial error lag model in the `spdep` package in `R` and the two-step estimator of Neumayer and Pluemper (2010) in Stata. The latter take as an input dyadic dependence information that can be produced using the `spundir` package in Stata. The spatial error lag model in the `spdep` failed to produce estimates, and the `spundir` package failed to resolve within a reasonable time (a week) while running on the same cluster.[7] This brings to the foreground the issue of computational complexity, by which our cluster robust estimator is much less demanding.

[7] See also footnote 6. The computational complexity is due to the fact that the model attempts to account for dependence across all countries in the data set (rather than small subsets, as would be the case in a typical spatial analysis) and the fact that in this application the set of cross-section units changes over time and therefore does not allow for a stable adjacency matrix over time.

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

Table 1 Estimates and robust standard error estimates from logistic regression of an indicator for militarized conflict between dyads on dyad-level predictors, 1885–1991, based on Russett and Oneal (2001, Table A5.1)

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
	Original GEE estimates		*Simple logistic regression estimates*			
Predictor of militarized conflict	*Coefficients*	*S.E.*	*Coefficients*	*Het. Robust S.E.*	*Naive cluster Robust S.E.*	*Dyadic cluster Robust S.E.*
Alliances	−0.539	0.159	−0.595	0.069	0.175	0.265
Power ratio	−0.318	0.043	−0.328	0.020	0.047	0.070
Lower democracy	−0.061	0.009	−0.064	0.005	0.010	0.015
Lower dependence	−52.924	13.405	−67.668	10.734	17.560	24.749
International organizations	−0.013	0.004	−0.011	0.002	0.005	0.008
Non-contiguity	−0.989	0.168	−1.048	0.074	0.181	0.185
Log distance	−0.376	0.065	−0.375	0.026	0.068	0.102
Only minor powers	−0.647	0.178	−0.618	0.078	0.188	0.344
Constant	−0.128	0.536	−0.181	0.211	0.562	0.840

Note: $N = 39{,}988$ dyad–year observations for 1,178 dyads and 146 units.

6 Conclusion

We have established convergence properties for a non-parametric variance estimator for regression coefficients that accounts for dyadic clustering. The estimator applies no restrictions on the dependency structure beyond the dyadic clustering assumption. The estimator is robust to the regression model being misspecified for the conditional mean. Such robustness is important because regression analysis typically relies on linear (in coefficients) approximations of unknown conditional mean functions (Angrist and Pischke 2009; Hubbard et al. 2010; Buja et al. 2014). Of course our analysis in no way excuses analysts from their responsibility to obtain as good an approximation as possible, nor does a robust fix for standard errors also solve the problem of finding a good approximation. Given a reasonable approximation for the conditional mean, the methods we have developed here allow for more accurate asymptotic statistical significance tests and confidence intervals.

The estimator is consistent in the number of units that form the basis of dyadic pairs. Simulations show that the estimator approaches the true standard error with modestly sized samples from a reasonable data-generating process. Applications show that inferences can be seriously anti-conservative if one fails to account for the dyadic clustering. This estimator is a natural complement to the non-parametric and semi-parametric regression analyses that are increasingly common in the social sciences (Angrist and Pischke 2009). Given that we can express the estimator as the sum of simpler and easy-to-compute robust variance estimators, it could be applied to any estimator for which a cluster–robust option is available.

A cost of the robust approach is efficiency. Our simulations show that a two-way (for cross-sectional data) random effects model can be considerably more efficient and provide reliable inference when the conditional mean is correctly specified. Our robust estimator could provide the basis of a test for misspecification (White 1980a, 1981; King and Roberts 2015). If there is little evidence of misspecification, the random effects estimator would be a reasonable choice given its efficiency. However, problems of computational non-convergence may make the random effects estimator infeasible—we encountered this in our application when we tried to use a three-way random effects model for dyadic panel data.

Of course, accounting for dynamic dyadic clustering may fail to fully account for all relevant dependencies in the data. For example, units may exhibit higher-order network effects: for example, a shock to unit A may affect unit C through connections that run via a third unit, B. In such cases, the methods developed here will likely yield standard errors that are too small, although they should still outperform methods that fail even to account for dyadic clustering.

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

Appendix A: Proof of Lemma 1

Proof. We have

$$\sqrt{N}(\hat{\beta}-\beta)=\left[\frac{1}{D}\sum_{d=1}^{D}X_dX_d'\right]^{-1}\frac{\sqrt{N}}{D}\sum_{d=1}^{D}X_d\epsilon_d.$$

Take $[(1/D)\sum_{d=1}^{D}X_dX_d']^{-1}$ first. By continuity of the inverse, the finite support of X_d and e_d, $\mathbf{X}$ full rank, and the fact that the sum of dependent terms in this sum is $O(D^{3/2})$ (Lehmann 1999, Equation (2.8.4)), we have that

$$\left[\frac{1}{D}\sum_{d=1}^{D}X_dX_d'\right]^{-1}\xrightarrow{p}\mathrm{E}\left[X_dX_d'\right]^{-1}.$$

By Slutsky's theorem, $\sqrt{N}(\hat{\beta}-\beta)$ has the same asymptotic distribution as $\mathrm{E}\left[X_dX_d'\right]^{-1}\frac{\sqrt{N}}{D}\sum_{d=1}^{D}X_d\epsilon_d$, which has mean zero and variance,

$$V=\frac{N}{D^2}\mathrm{E}\left[X_dX_d'\right]^{-1}\mathrm{Var}\left[\sum_{d=1}^{D}X_d\epsilon_d\right]\mathrm{E}\left[X_dX_d'\right]^{-1}. \tag{5}$$

Then,

$$\begin{aligned}\mathrm{Var}\left[\sum_{d=1}^{D}X_d\epsilon_d\right]&=\sum_{d=1}^{D}\sum_{d'=1}^{D}\mathrm{Cov}\left[X_d\epsilon_d,X_{d'}\epsilon_{d'}\right]\\&=\sum_{d=1}^{D}\sum_{d'=1}^{D}\mathrm{E}\left\{X_dX_{d'}'\mathrm{Cov}\left[\epsilon_d,\epsilon_{d'}|\mathbf{X}\right]\right\}\\&=\sum_{d=1}^{D}\left(\underbrace{\mathrm{E}\left\{X_dX_d'\mathrm{Var}\left[\epsilon_d|\mathbf{X}\right]\right\}}_{A}+\underbrace{\sum_{d'\in\mathscr{S}(d)}\mathrm{E}\left\{X_dX_{d'}'\mathrm{Cov}\left[\epsilon_d,\epsilon_{d'}|\mathbf{X}\right]\right\}}_{B},\right)\end{aligned} \tag{6}$$

where $\mathscr{S}(d)=\{d'\neq d\}\cap\{d':s(d)\cap s(d')\neq\emptyset\}$, the set of dyads other than d that share a member from d. □

Appendix B: Proof of Proposition 1

Proof. By Chebyshev's inequality, two conditions are sufficient for $N\hat{V}-V\xrightarrow{p}0$ as $N\to\infty$: (i) $\mathrm{E}\left[\hat{V}-V\right]\to 0$ and (ii) $\mathrm{Var}\left[N\hat{V}-V\right]=\mathrm{Var}\left[N\hat{V}\right]\to 0$ (Lehmann 1999, Theorem 2.1.1).

For (i), consider the interior (meat) of $\hat{V}$:

$$\hat{\Sigma}=\sum_{d=1}^{D}\left(X_dX_d'e_d^2+\sum_{d'\in\mathscr{S}(d)}X_dX_{d'}'e_de_{d'}\right)$$

and the corresponding term for V:

$$\Sigma=\sum_{d=1}^{D}\left(\mathrm{E}\left\{X_dX_d'\mathrm{Var}\left[\epsilon_d|\mathbf{X}\right]\right\}+\sum_{d'\in\mathscr{S}(d)}\mathrm{E}\left\{X_dX_{d'}'\mathrm{Cov}\left[\epsilon_d,\epsilon_{d'}|\mathbf{X}\right]\right\}\right).$$

We have

$$\mathrm{E}\left[\hat{\Sigma}\right]=\sum_{d=1}^{D}\left(\mathrm{E}\left\{X_dX_d'\mathrm{E}\left[e_d^2|\mathbf{X}\right]\right\}+\sum_{d'\in\mathscr{S}(d)}\mathrm{E}\left\{X_dX_{d'}'\mathrm{E}\left[e_de_{d'}|\mathbf{X}\right]\right\}\right)\to\Sigma$$

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

by consistency of $\hat{\beta}$, an implication of Lemma 1. Lemma 1 also established that $(\mathbf{X}'\mathbf{X}/D)^{-1} \xrightarrow{p} \mathrm{E}[X_d X_d']^{-1}$, while boundedness of X_d implies uniform boundedness of $\mathrm{E}[X_d X_d']^{-1}$. Taking these elements together, uniform continuity implies $\mathrm{E}[\hat{V} - V] \to 0$ (cf. White 1980a, Theorem 1).

For (ii), we follow the strategy of Lehmann (1999, Section 2.8), which requires showing that growth of covariance contributions due to dependent units is bounded in N. Begin by rewriting $N\hat{V}$ as

$$N\hat{V} = N(\mathbf{X}'\mathbf{X}/D)^{-1}\left(D^{-2}\sum_{d}\sum_{d'} I_{d,d'} X_d X_{d'}' e_d e_{d'}\right)(\mathbf{X}'\mathbf{X}/D)^{-1},$$

where $I_{d,d'}$ is an indicator function denoting a pair of dyads d and d' that share a member, as defined above. We have established that $(\mathbf{X}'\mathbf{X}/D)^{-1}$ converges in probability to a finite limit, $E[X_d X_d']^{-1}$, so we can treat this as $O(1)$. Given bounded support, $\mathrm{Var}[N\hat{V}]$ is of the same order as

$$\begin{aligned}\Omega &= N^2 D^{-4}\mathrm{Var}\left[\sum_{d}\sum_{d'} I_{d,d'} X_d X_{d'}' e_d e_{d'}\right] \\ &= N^2 D^{-4}\sum_{d}\sum_{d'}\sum_{d''}\sum_{d'''} \mathrm{Cov}\left(I_{d,d'} X_d X_{d'}' e_d e_{d'}, I_{d'',d'''} X_{d''} X_{d'''}' e_{d''} e_{d'''}\right).\end{aligned}$$

To show $\mathrm{Var}[N\hat{V}] \to 0$, it is sufficient to establish that Ω is at most $O(N^{-1})$. Consider quadruples of dyads indexed by d, d', d'', d''' such that $d \neq d' \neq d'' \neq d'''$. Suppose there are Q quadruples and define $p_Q(A) = |A|/Q$. The covariance terms take positive values when $I_{d,d'} = I_{d'',d'''} = 1$, which occurs at rate $O(N^{-2})$ (i.e., when we have pairs of terms that are actually summed into $\hat{V}$), and we have dependence across (d, d') and (d'', d'''). Then, the six distinct ways that such dependence can occur and the associated proportion of quadruples for which these occur for a sample of size N are as follows:

$$p_Q(\{d \neq d''\} \cap \{d'' : s(d) \cap s(d'') \neq \emptyset\}) = O(N^{-1}),$$

$$p_Q(\{d \neq d'''\} \cap \{d''' : s(d) \cap s(d''') \neq \emptyset\}) = O(N^{-1}),$$

$$p_Q(\{d \neq d'' \neq d'''\} \cap \{d'', d''' : s(d) \cap s(d'') \neq \emptyset \cap s(d) \cap s(d''') \neq \emptyset\}) = O(N^{-2}),$$

$$p_Q(\{d' \neq d''\} \cap \{d'' : s(d') \cap s(d'') \neq \emptyset\}) = O(N^{-1}),$$

$$p_Q(\{d' \neq d'''\} \cap \{d''' : s(d') \cap s(d''') \neq \emptyset\}) = O(N^{-1}), \text{ and}$$

$$p_Q(\{d' \neq d'' \neq d'''\} \cap \{d'', d''' : s(d') \cap s(d'') \neq \emptyset \cap s(d') \cap s(d''') \neq \emptyset\}) = O(N^{-2}).$$

Therefore, the proportion of quadruples yielding a positive covariance is the proportion of cases contributing non-zero values to the sum Ω. This proportion is bounded by the proportion of cases where the indicator functions are both one and the order of the proportion of dependent quadruples from across the six cases characterized above, that is, $O(N^{-2}N^{-1}) = O(N^{-3})$. Then, Var $[N\hat{V}]$ is at most $O(N^2 D^{-4} D^4 N^{-3}) = O(N^{-1}) \to 0$. □

Appendix C: Proof of Proposition 2

Proof. Consider the interior (meat) of $\hat{V}_r$:

$$\hat{\Sigma}_r = \sum_{d=1}^{D}\left(X_d' e_d e_d' X_d + \sum_{d' \in \mathscr{S}(d)} X_d' e_d e_{d'}' X_{d'}\right) = \sum_{d=1}^{D} X_d' e_d e_d' X_d + \sum_{d=1}^{D}\sum_{d' \in \mathscr{S}(d)} X_d' e_d e_{d'}' X_{d'}.$$

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

Taking the second component,

$$\sum_{d=1}^{D}\sum_{d'\in\mathscr{S}(d)} X'_d e_d e'_{d'} X_{d'} = \sum_{i=1}^{N} C_i, \text{ with } C_i = \sum_{j\neq i}\sum_{k\neq i,j} X'_{d(i,j)} e_{d(i,j)} e'_{d(i,k)} X_{d(i,k)}.$$

Define

$$D_i = \sum_{j\neq i} X'_{d(i,j)} e_{d(i,j)} e'_{d(i,j)} X_{d(i,j)} \text{ and } E_i = \sum_{j,k\neq i}\sum_{t=1}^{T(d(j,k))} X_{d(j,k)t} X'_{d(j,k)t} e^2_{d(j,k)t}.$$

Then, working with $\hat{\Sigma}_{C,i}$, $\hat{\Sigma}_D$, and $\hat{\Sigma}_0$ as defined in the proposition,

$$\hat{\Sigma}_{C,i} = C_i + D_i + E_i, \ \ 2\hat{\Sigma}_D = \sum_{i=1}^{N} D_i, \ \text{ and } \ (N-2)\hat{\Sigma}_0 = \sum_{i=1}^{N} E_i.$$

Then,

$$\begin{aligned}\hat{\Sigma}_r &= \hat{\Sigma}_D + \sum_{i=1}^{N} C_i \\ &= \hat{\Sigma}_D + \sum_{i=1}^{N} [C_i + D_i + E_i - D_i - E_i] \\ &= \sum_{i=1}^{N} \hat{\Sigma}_{C,i} - \hat{\Sigma}_D - (N-2)\hat{\Sigma}_0,\end{aligned}$$

and by linear distributivity for $(\mathbf{X}'\mathbf{X})^{-1}$

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

with $\hat{V}_{C,i}$, $\hat{V}_D$, and $\hat{V}_0$ as defined in the proposition. □

Appendix D: Weighted Observations

Weighting is a common way to adjust for unequal probability sampling of dyadic interactions, among other applications. The extension to the weighted case is straightforward. Assume weighted directed dyad observations with weights finite and fixed; denote the weight for dyad d as w_d. Define the sample data as $\mathbf{X} = (X'_1 \ldots X'_D)'$ and $Y = (Y_1 \ldots Y_D)'$ and the matrix of weights as $\mathbf{W} = \text{diag}\{w_1, \ \ldots, w_D\}$. Then, the weighted least squares estimator is

$$\hat{\beta}_w = (\mathbf{X}' W X)^{-1} \mathbf{X}' \mathbf{W} Y.$$

Corollary 2. *For the weighted dyads case, assume the same conditions as in Proposition 1 and consider the following variance estimator*:

$$\hat{V}_w = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \sum_{d=1}^{D} \left(X_d X'_d w_d^2 e_d^2 + \sum_{d'\in\mathscr{S}(d)} X_d X'_{d'} w_d w_{d'} e_d e_{d'} \right) (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}. \tag{7}$$

Then, as $N \to \infty$,

$$N\text{Var}\,[\hat{\beta}_w - \beta_w] - V_w \xrightarrow{p} 0,$$

and

$$N\hat{V}_w - V_w \xrightarrow{p} 0,$$

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

where

$$V_w = \frac{N}{D^2}\mathrm{E}\left[w_d X_d' X_d\right]^{-1}\mathrm{E}\left(\mathrm{E}\left[X_d X_d' w_d^2 \mathrm{Var}\left[\epsilon_d|\mathbf{X}\right]\right] + \sum_{d' \in \mathscr{S}(d)} \mathrm{E}\left[X_d X_{d'}' w_d w_{d'} \mathrm{Cov}\left[\epsilon_d, \epsilon_{d'}|\mathbf{X}\right]\right]\right)\mathrm{E}\left[w_d X_d X_d'\right]^{-1}.$$

Appendix E: Generalized Linear Models

An implementation for generalized linear models follows the usual M-estimation results (Stefanski and Boos 2002; Wooldridge 2010, chap. 12). Given an estimating equation, $\psi(D;\theta)$, on data D, and with parameters θ and parameter estimates $\hat{\theta}$, the sandwich approximation for the variance is $\mathbf{A}^{-1}\mathbf{B}\{\mathbf{A}^{-1}\}'$, where

$$\mathbf{A} = \mathrm{E}\left[-\frac{\partial}{\partial\theta'}\psi(D;\hat{\theta})\right] \text{ and } \mathbf{B} = \mathrm{E}\,\psi(D;\hat{\theta})\psi(D_i;\hat{\theta})'.$$

For logistic regression, the estimating equation is

$$\psi((\mathbf{X}_d, Y_d);\hat{\beta}) = \mathbf{X}_d(Y_d - p_d),$$

where $p_d = \mathrm{expit}(\mathbf{X}_d\hat{\beta})$, the predicted probability for dyad d. The plug-in variance estimator for logistic regression coefficients is given by

$$\hat{V}_l = (\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}\sum_{d=1}^{D}\left(X_d X_d' r_d^2 + \sum_{d' \in \mathscr{S}(d)} X_d X_{d'}' r_d r_{d'}\right)(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1},$$

where

$$\mathbf{M} = \mathrm{diag}(p_d(1-p_d))_{D\times D} \text{ and } r_d = Y_d - p_d.$$

The extensions to repeated and weighted observations follow analogously.

Conflict of interest statement. None declared.

References

Angrist, Joshua D., and Guido W. Imbens. 2002. Comment on "Covariance adjustment in randomized experiments and observational studies" by Paul R. Rosenbaum. *Statistical Science* 17(3):304–7.

Angrist, Joshua D., and Jorn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Arellano, Manuel. 1987. Computing robust standard errors for within-group estimators. *Oxford Bulletin of Economics and Statistics* 49(4):431–34.

Beck, Nathanial, Kristian Skrede Gleditsch, and Kyle Beardsley. 2006. Space is more than geography: Using spatial ecometrics in the study of political economy. *International Studies Quarterly* 50:27–44.

BujaAndreas, Richard Berk, Lawrence Brown, Edward George, Emil Pitkin, Mikhail Traskin, Linda Zhao, and Kai Zhang. 2014. Models as approximations: A conspiracy of random predictors and model violations against classical inference in regression. Manuscript, Wharton School, University of Pennsylvania, Philadelphia.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. Robust inference with multi-way clustering. *Journal of Business and Economic Statistics* 29(2):238–49.

Chamberlain, Gary. 1982. Multivariate regression models for panel data. *Journal of Econometrics* 18(1):5–46.

Conley, Timothy G. 1999. GMM estimation with cross-sectional dependence. *Journal of Econometrics* 92:1–45.

Davidson, Russell, and James G. MacKinnon. 2004. *Econometric theory and methods*. Oxford: Oxford University Press.

Erikson, R. S., P. M. Pinto, and K. T. Rader. 2014. Dyadic analysis in international relations: A cautionary tale. *Political Analysis* 22(4):457–63.

Fafchamps, Marcel, and Flore Gubert. 2007. The formation of risk sharing networks. *Journal of Development Economics* 83:326–50.

Fisman, Raymond, Sheena S. Iyengar, Emik Kamenica, and Itamar Simonson. 2006. Gender differences in mate selection: Evidence from a speed dating experiment. *Quarterly Journal of Economics* 121:673–97.

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press

Gelman, Andrew, and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
Goldberger, Arthur S. 1991. *A course in econometrics*. Cambridge, MA: Harvard University Press.
Green, Donald P., Soo Yeon Kim, and David H. Yoon. 2001. Dirty pool. *International Organization* 55(2):441–68.
Greene, William H. 2008. *Econometric analysis*. 6th ed. Upper Saddle River, NJ: Pearson.
Hansen, Christian B. 2007. Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics* 141:597–620.
Hoff, Peter D. 2005. Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association* 100(469):286–95.
Hubbard, Alan E., Jennifer Ahern, Nancy L. Fliescher, Mark Van Der Laan, Sheri A. Lippman, Nicholas Jewell, Tim Bruckner, and William A. Satariano. 2010. To GEE or not to GEE: Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology* 21(4):467–74.
Huber, Peter J. 1967. *The behavior of maximum likelihood estimates under nonstandard conditions*. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1, pp. 221–33. Berkeley, CA: University of California Press.
Kenny, David A., Deborah A. Kashy, and William L. Cook. 2006. *Dyadic data analysis*. New York: Guilford Press.
King, Gary, and Margaret E. Roberts. 2015. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It." *Political Analysis* 23(2):159–79.
Lehmann, Erich L. 1999. *Elements of large sample theory*. New York: Springer-Verlag.
Liang, Kung-Yee, and Scott L. Zeger. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73(1):13–22.
Lin, Winston. 2013. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics* 7(1):295–318.
MacKinnon, James G., and Halbert White. 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29(3):305–25.
Moulton, Brent R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32:385–97.
Neumayer, Eric, and Thomas Pluemper. 2010. Spatial effects in dyadic data. *International Organization* 64(1):145–65.
Russett, Bruce M., and John R. Oneal. 2001. *Triangulating peace: Democracy, interdependence, and international organizations*. New York: Norton.
Samii, Cyrus. 2015. Cluster-Robust Variance Estimation for Dyadic Data. http://dx.doi.org/10.7910/DVN/OMJYE5, Harvard Dataverse, V1 [UNF:6:WJJ3ZmDS7COvpy1kwztcMQ==].
Stefanski, Leonard A., and Dennis D. Boos. 2002. The calculus of M-estimation. *American Statistician* 56(1):29–38.
Stock, James H., and Mark W. Watson. 2008. Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica* 76(1):155–74.
White, Halbert. 1980a. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4):817–38.
———. 1980b. Using least squares to approximate unknown regression functions. *International Economic Review* 21(1):149–70.
———. 1981. Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* 76(374):419–33.
———. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–25.
———. 1984. *Asymptotic theory of econometricians*. New York: Academic Press.
Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.
Zorn, Christopher. 2001. Generalized estimating equation models for correlated data: A review with applications. *American Journal of Political Science* 45:470–90.

https://doi.org/10.1093/pan/mpv018 Published online by Cambridge University Press