Clustered Errors in Stata

Austin Nichols and Mark Schaffer

10 Sept 2007

Clustered Errors

Suppose we have a regression model like

$$Y_{it} = X_{it}\beta + u_i + e_{it}$$

where the u_i can be interpreted as individual-level fixed effects or errors. The t index brings to mind panel data, with multiple observations on people or firms over time, but in fact the t index can represent any arbitrary index for observations grouped along two dimensions.

The usual assumption is that e_{it} is **iid** (independently and identically distributed) but this is clearly violated in many cases. A natural generalization is to assume "clustered errors" i.e. that observations within group i are correlated in some unknown way, inducing correlation in e_{it} within i, but that groups i and j do not have correlated errors.

In the presence of clustered errors, OLS estimates are still unbiased but standard errors may be quite wrong, leading to incorrect inference in a surprisingly high proportion of finite samples.

Types of Clustering—Serial Corr. and Cluster Sampling

The notation above naturally brings to mind a paradigmatic case of clustering: a panel model with group-level shocks (u_i) and serial correlation in errors (e_{it}), in which case i indexes panel and t indexes time of observation. This type of clustering could also arise from a survey in which blocks of observations are selected randomly, but there is no reason to suppose that observations within block have uncorrelated errors. For example, consider a random sample of schools that contain students whose response to some policy X might be correlated (in which case i indexes school and t indexes student within school).

Another leading example is where e_{it} is **iid** but the u_i are unmodeled (perhaps because some variable of interest in X does not vary across both i and t). Here, the group-level innovations drive the clustering of errors $\nu_{it} = u_i + e_{it}$.

Types of Clustering—Induced Group-Level Shocks

Another similar type occurs where observations are randomly sampled, but the explanatory variable X is measured at a higher level (see Moulton 1990; Bertrand, Duflo, and Mullainathan 2004). For example, students might be randomly sampled to model test scores as a function of school characteristics, but this will result in clustered errors at the school level. If students were randomly sampled to model test scores as a function of classes taken (measured at the individual level, not the school level), but classes taken and their effects on test scores are correlated within school, this may also induce clustering of errors at the higher level (school in the example). Any measurement error or misspecification in these types of models will also naturally "induce" group-level shocks u_i and correlation in errors e_{it} even if they are absent in the "true" model.

FE and Clusters

The model as written

$$Y_{it} = X_{it}\beta + u_i + e_{it}$$

naturally raises the question of how "fixed effects" (FE) relate to "clustered errors." In short, if the e_{it} are **iid**, the u_i if unmodeled produce clustering of the aggregate error $\nu_{it} = u_i + e_{it}$.

FE is an example of modeling "error components", where the structure is very special, i.e., every observation within the group is equally-well correlated with every other observation. It's from the **GLS school**, so to speak. Partial out the fixed effects and you're left with a homoskedastic idiosyncratic error—use **xtreg,fe** with the classical VCV.

"Clustered errors" is an example of Eicker-Huber-White-robust treatment of errors, i.e., make as few assumptions as possible. We keep the assumption of zero correlation across groups as with fixed effects, but allow the within-group correlation to be anything at all—use **regress** with **cluster()**.

Combining FE and Clusters

If the model is overidentified, clustered errors can be used with two-step GMM or CUE estimation to get coefficient estimates that are efficient as well as robust to this arbitrary within-group correlation—use **ivreg2** with the **cluster(v) gmm2s** option set.

Finally, the "GLS" and "robust" approaches can be combined. Partial-out the fixed effects, and then use cluster-robust to address any remaining within-group correlation—use **xtreg,fe** with cluster().

First-differencing (FD) can be similarly motivated: FD to get rid of the fixed effects, and then use cluster-robust errors to mop up the remaining and/or introduced serial correlation.

For some reason, combining the GLS and robust approaches is absolutely standard in the panel/serial correlation literature, and almost completely ignored in cross-section/heteroskedasticity practice. It's perfectly reasonable to do feasible GLS on a cross-section to get improvements in efficiency and then use robust SEs to address any remaining heteroskedasticity, but nobody seems to do this (GLS is too old-fashioned, perhaps?).

Number of Clusters

The cluster-robust standard error estimator converges to the true standard error as the number of clusters M approaches infinity, not the number of observations N.

Kézdi (2004) shows that 50 clusters (with roughly equal cluster sizes) is often close enough to infinity for accurate inference, and further that, even in the absence of clustering, there is little to no cost of using the CRSE estimator, as long as the number of clusters is large.

With a small number of clusters ($M << 50$), or very unbalanced cluster sizes, the cure can be worse than the disease, i.e. inference using the cluster-robust estimator may be incorrect more often than when using the

Rank of VCV

The rank of the variance-covariance matrix produced by the cluster-robust estimator has rank no greater than the number of clusters M, which means that at most M linear constraints can appear in a hypothesis test (so we can test for joint significance of at most M coefficients).

In a fixed-effect model, where there are a large number of parameters, this often means that test of overall model significance is feasible. However, testing fewer than M linear constraints is perfectly feasible in these models, though when fixed effects and clustering are specified at the same level, tests that involve the fixed effects themselves are inadvisable (the standard errors on fixed effects are likely to be substantially underestimated, though this will not affect the other variance estimates in general).

Estimates and their VCV

Note that the heteroskedasticity-robust and cluster-robust estimators for standard errors have no impact whatsoever on point estimates.

One **could** use information about the within-cluster correlation of errors to obtain more efficient estimates in many cases (see e.g. Diggle et al. 2002). There are also a variety of multi-level methods of parametrizing the distribution of errors to obtain more efficient estimates (using e.g. **xtmixed** and other model types—see Rabe-Hesketh and Skrondal 2005 for more). We will focus however on models where the **point estimates are unchanged** and only the estimated variance of our point estimates is affected by changing assumptions about errors.

In addition to improving the efficiency of the point estimates in regressions, modeling intra-cluster correlations can also result in improvements in meta-analysis, both in correctly modeling the variance of individual estimates and computing effect sizes. See Hedges (2006) for details.

Two Families of Sandwich Estimators

The OLS estimator of the Var-Cov matrix is: $\hat{V}_O = q\hat{\mathrm{V}} = q\left(\mathrm{X}'\mathrm{X}\right)^{-1}$ (where for `regress`, q is just the residual variance estimate $s^2 = \frac{1}{N-k}\sum_{j=1}^{N}\hat{e}_i^2$). The heteroskedasticity-robust estimator is:

$$\hat{V}_H = q_c\hat{\mathrm{V}}\left(\sum_{j=1}^{N} w_j\varphi_j' w_j\varphi_j\right)\hat{\mathrm{V}}$$

where the φ_j are observation-level contributions to $\partial \ln L/\partial\beta$ and the w_j are observation-level weights. For `regress`, the φ_j are just $\hat{e}_j\mathrm{x}_j$. The *VDV* structure explains the common name "sandwich estimator" though the cluster-robust estimator is also a sandwich estimator:

$$\hat{V}_C = q_c\hat{\mathrm{V}}\left(\sum_{j=1}^{M}\left(\varphi_j^G\right)'\varphi_j^G\right)\hat{\mathrm{V}}$$

where now the φ_j^G are within-cluster weighted sums of observation-level contributions to $\partial \ln L/\partial\beta$, and there are M clusters.

Sandwich Estimators and Other Robustifications

Eicker (1967) and Huber (1967) introduced these sandwich estimators, but White (1980; 1982), Liang and Zeger (1986), Arellano (1987), Newey and West (1987), Froot (1989), Gail, Tan, and Piantodosi (1988), Kent (1982), Royall (1986), and Lin and Wei (1989), Rogers (1993), Williams (2000), and others explicated and extended aspects of the method in a non-survey context, so these are often cited as sources in specific applications. In the context of clustering induced by survey design, Kish and Frankel (1974), Fuller (1975), and Binder (1983), and Binder and Patak (1994), also derived results on cluster-robust estimators with broad applicability.

Baum, Schaffer, and Stillman (2003; 2007) describe a variety of SE estimators (all calculated by their **ivreg2** program for Stata, with or without instrumental variables) robust to various other violations of the **iid** error assumptions, including heteroskedasticity-and-autocorrelation-robust (HAC-robust) estimators.

Stock and Watson (2006) point out that with fixed effects, both the standard heteroskedasticity-robust and HAC-robust covariance estimators are inconsistent for T fixed and $T > 2$, but the cluster-robust estimator does not suffer from this problem. One of their conclusions is that if serial correlation is expected, the cluster-robust estimator is the preferred choice.

Finite-Sample Adjustments

The finite-sample adjustment q_c takes two forms (not including $q_c = 1$, another possibility), depending on the model used:

$$q_c = \frac{N-1}{N-k}\frac{M}{M-1}$$

where M is the number of clusters and N the number of observations, or

$$q_c = \frac{M}{M-1}$$

The Methods and Formulas of [R] `regress` calls these the regression-like formula and the asymptotic-like formula respectively. Fuller et al. (1986) and Mackinnon and White (1985) discuss finite-sample adjustments in more detail.

The Nature of the CRSE Correction

The heteroskedasticity-robust SE estimator scales not by the sum of squared residuals, but by the sum of "squared" products of residuals and the X variables, and the CRSE estimator further sums the products within cluster (if the products are negatively correlated within cluster, the CRSE will be smaller than the HRSE, and if positively correlated, larger). If the traditional OLS model is true, the residuals should, of course, be uncorrelated with the X variables, but this is rarely the case in practice.

The correlation may arise not from correlations in the residuals within a correctly specified model, but from specification error (such as omitted variables), so one should always be alert to that possibility.

Misspecification and the CRSE Correction

As Sribney (1998) points out: When CRSE estimates are smaller than standard SE estimates,

> *since what you are seeing is an effect due to (negative) correlation of residuals, it is important to make sure that the model is reasonably specified and that it includes suitable within-cluster predictors. With the right predictors, the correlation of residuals could disappear, and certainly this would be a better model.*
>
> *...suppose that you measured the number of times each month that individuals took out the garbage, with the data clustered by household. There should be a strong negative correlation here. Adding a gender predictor to the model should reduce the residual correlations.*

The CRSE will do nothing about bias in $\hat{\beta}$ when $E(X'e) \neq 0$.

Approximating the CRSE Correction

As Cameron, Gelbach, and Miller (2006a, p.5) note, if the primary source of clustering is due to group-level common shocks, a useful approximation is that for the jth regressor the default OLS variance estimate based on $s^2(X'X)^{-1}$ should be inflated by a factor of

$$1 + \rho_{\mathrm{e}}\rho_{\mathrm{x}_j}(\bar{N}_g - 1)$$

where ρ_{x_j} is the intra-cluster correlation of x_j, ρ_{e} is the intra-cluster correlation of residuals, and $\bar{N}_g$ is the average cluster size; in many settings the adjustment factor can be large even if ρ_{e} is small.

This approximation is closely related to the approximation given in Kish (1965, p.162) for the estimation of means in clustered data: he recommends inflating the variance estimate for the mean by a factor (or the SE by the square root of the factor):

$$1 + r(\bar{N}_g - 1)$$

where r is the measure of intraclass correlation (ICC) known as roh [not rho]. The approximation for regression with group-level common shocks is quite similar, with the adjustment that we now want the mean of y conditional on X.

Non-Nested and Nested Clusters

An extension of the basic one-dimensional case is to multiple levels of clustering. For example, errors may be clustered by country and by city, or errors may be clustered by country and by year. In the first case, the levels of clustering are nested, but in the second case, the clustering is along two dimensions and observations in each cluster along one dimension may appear in multiple clusters along the other. The latter case of non-nested clusters is discussed by Cameron, Gelbach, and Miller (2006a), who provide Stata code for estimating cluster-robust standard errors in this case.

To estimate cluster-robust standard errors in the presence of nested multi-level clustering, one can use the **svy** suite of commands.

Nested Clusters Using **svy**

It is straightforward to compute cluster-robust estimates for multi-level clustering with nested clusters using

```
svyset clevel1 || clevel2
```

(pweights are easily added as well) and then any command that allows the **svy:** prefix. In general, however, the correction at the highest level is the important one. Specifying clustering at the classroom level and clustering at the school level is unlikely to result in any substantive differences in inference relative to merely specifying clustering at the school level.

This argues for always specifying clustering at the highest of all nested levels at which intra-cluster correlation in errors may be a problem, but there is a tradeoff: at higher levels the number of clusters will be smaller, so the asymptotic results for the estimator are less likely to hold.

Clustering Using **suest**

The **suest** command also implements a cluster correction, for T_G equations with M observations each corresponding to M clusters (both assume M going off to infinity). In **suest**, 5 equations of 10 observations each is not equivalent to 50 observations, it's 10 "super-observations."

The intuition behind **suest** is that it clusters by observational unit across equations. The **suest** clusters are the outer products of the errors for an observational unit, where "errors" means the vector of errors across equations.

Again, we have the contrast between the GLS school and the robust school. In the GLS school, you run **sureg** and estimate the cross-equation covariances. In the robust school, you run **suest** and come up with a Var-Cov matrix that is robust to arbitrary cross-equation correlation.

Problems with Cluster-Robust SE's

Why specify **cluster** (or use **svy**)?

- If the assumptions are satisfied, and errors are clustered, you'll get much better SE estimates.
- If the assumptions are satisfied, and errors aren't clustered, you'll get roughly the same SE estimates as if you had not specified **cluster** (i.e. no cost of robustness).

Why not always specify **cluster** (or use **svy**)?

- Convergence
- Bias
- Correlation across clusters
- Degrees of freedom

Speed of Convergence

The CRSE is asymptotic in the number of clusters M. If M is small, there is no guarantee that the cluster-robust estimator will improve your inference—the cluster-robust estimator may make matters worse.

Kézdi (2004) shows that 50 clusters is often close enough to infinity for accurate inference, but these are simulations for a specific type of model. You may want to do simulations for a model that fits your specific application if you are worried about the convergence of the cluster-robust estimator, and what it implies for the reliability of your inferences.

Downward Bias

Rogers (1993) argues that "if no cluster is larger than 5 percent or so of the total sample, the standard errors will not be too far off because each term will be off by less than 1 in 400." This implies that CRSE's with 20 equal-sized clusters would suffer from a very small bias.

With finite M, the cluster-robust estimator produces estimates of standard errors that are too small on average (i.e. they are biased downward). With M much less than 50, the bias can be substantial, particularly with $M < 10$. Cameron, Gelbach, and Miller (2006b) report that a "wild bootstrap" cluster-robust estimator performs well when $M < 50$. See also Wooldridge (2003) for more discussion and suggestions.

Degrees of freedom

Since the rank of the VCV matrix produced by the CRSE is no greater than the number of clusters M you may not be able to test as many parameters as desired. For example, you could not cluster at the panel level and test for panel-specific intercepts and trends, since you would have at least twice as many parameters as degrees of freedom.

Given the limits on the number of parameters that may be tested in theory, even asymptotically, one might be worried about the small-sample properties of tests that involve nearly as many constraints as M. We will present simulations for certain cases.

A test for clustering

If you're worried about potential problems when using CRSE estimates, you'd like to test for the presence of clustering, to see whether you really need to adjust for clustering. Kézdi (2007) provides a test for clustering in the spirit of the White (1980) test for heteroskedasticity (see `hettest`, `whitetst`, `ivhettest` in Stata)

Two programs `cltest` and `xtcltest` (to be available from SSC soon) implement the Kézdi (2007) test in Stata, when run after `reg` and `xtreg` estimation commands, respectively.

Balanced Panels, Equal Cluster Sizes, OLS-SE

Suppose we have the error components model

$$Y_{it} = X_{it}\beta + u_i + e_{it}$$

with $\beta_1 = 1$ and we have 50 balanced clusters, and 20 observations per cluster. Let the share of error variance due to the within-cluster component vary from 0 to 1 (across rows) and the share of within-cluster variation in regressors vary from 0 to 1 (across columns), and test $H_0 : \beta_1 = 1$ with $\alpha = 0.05$:

Rejection rates, nominal 5 percent level, OLS-SE

	0	25	50	75	100
0	.048	.043	.049	.048	.065625
25	.054	.057	.113	.157	.3052959
50	.052	.153	.312	.455	.6832814
75	.054	.209	.468	.679	.876161
100	.056	.241	.503	.716	

Balanced Panels, Equal Cluster Sizes, HRSE

Rejection rates, nominal 5 percent level, Het-Robust SE

	0	25	50	75	100
0	.049	.045	.05	.049	.0708333
25	.051	.057	.112	.154	.3094496
50	.054	.154	.321	.459	.6874351
75	.053	.202	.475	.679	.877193
100	.056	.242	.503	.715	

Balanced Panels, Equal Cluster Sizes, CRSE

Rejection rates, nominal 5 percent level, Clust-Robust SE

	0	25	50	75	100
0	.054	.039	.06	.09	
25	.053	.046	.107	.196	
50	.052	.07	.139	.335	
75	.056	.08	.179	.425	
100	.054	.078	.189	.434	

Balanced Panels, Equal Cluster Sizes, FECRSE

Rejection rates, nominal 5 percent level, FE and Clust-Robust SE

	0	25	50	75	100
0	.061	.038	.055	.055	
25	.054	.04	.044	.042	
50	.057	.054	.053	.062	
75	.056	.047	.044	.058	
100	.046	.047	.052	.042	

Unbalanced Panels and Unequal Cluster Sizes, OLS-SE

Now suppose we have 50 clusters and 1000 observations again, but 10 observations per cluster in 49 clusters and one cluster with 510 obs:

Rejection rates, nominal 5 percent level, OLS-SE

	0	25	50	75	100
0	.047	.056	.053	.058	.0679916
25	.047	.071	.073	.1	.1753112
50	.05	.171	.223	.347	.5658996
75	.04	.221	.41	.589	.8569948
100	.044	.27	.452	.677	

Unbalanced Panels and Unequal Cluster Sizes, HRSE

	0	25	50	75	100
0	.045	.053	.05	.059	.0700837
25	.048	.069	.077	.098	.1991701
50	.05	.166	.207	.34	.5774059
75	.047	.216	.388	.569	.8632124
100	.045	.271	.436	.654	

Unbalanced Panels and Unequal Cluster Sizes, CRSE

	0	25	50	75	100
0	.113	.104	.106	.123	
25	.105	.104	.095	.166	
50	.071	.133	.106	.253	
75	.031	.111	.096	.297	
100	.024	.116	.092	.299	

Unbalanced Panels and Unequal Cluster Sizes, FECRSE

	0	25	50	75	100
0	.119	.112	.115	.127	
25	.134	.123	.097	.111	
50	.106	.113	.103	.129	
75	.118	.118	.123	.126	
100	.088	.11	.078	.084	

Testing the Limits of df

Kézdi (2004) and our own simulations tell us that the CRSE performs extremely well in relation to the HRSE or OLS SE estimators with respect to inference on a single parameter, as long as we have at least 50 clusters.

However, we know that we cannot test more than M coefficients. It makes sense to question how well the CRSE estimator performs when testing $M-2$ or $M-1$ coefficients.

Preliminary simulations show that the rejection rate rises from 5 percent to 100 percent as the number of coefficients increases from 1 to M. This needs further investigation.

Comparisons to a Parametric Correction

Suppose we have autocorrelated errors in a panel model:

$$Y_{it} = X_{it}\beta + u_i + e_{it}$$

with

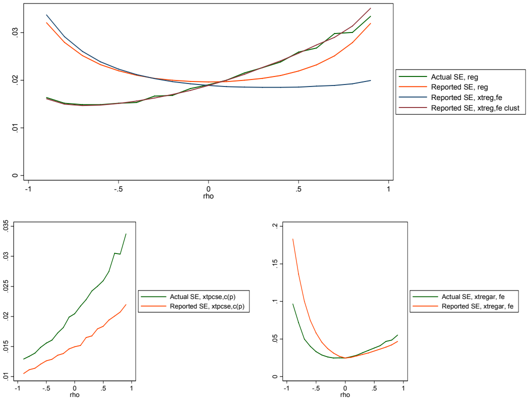
$$e_{it} = \rho e_{i(t-1)} + z_{it}$$

where z_{it} is **iid**. We could use **xtregar y x, fe**, **xtpcse y x, c(p)**, or **xtreg y x, fe cluster()**. How do these compare in finite samples? We can use MC simulation to evaluate the two approaches.

Additionally, Wooldridge (2002, pp.282-283) derives a simple test for autocorrelation in panel-data models, and the user-written program `xtserial` (Drukker 2003) performs this test in Stata. We can compare the performance of `xtserial` and `cltest` using MC simulation.

SE Estimates with Autocorrelation

Suppose $t \in \{1,2,3,4,5,6,7\}$ and $x = t - 4$ with $y = x + e$ and $M = 100$ (i.e. we are estimating a trend line $\beta = 1$ and there are 100 clusters). Suppose u_i is mean zero and uniform on $(-.5, .5)$. Here is a comparison of the reported and true SD of the OLS estimates (see also Diggle et al. 2002 Figure 1.7):

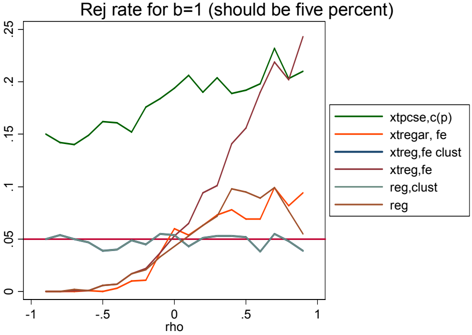


Rejection Rates, AR(1) Errors

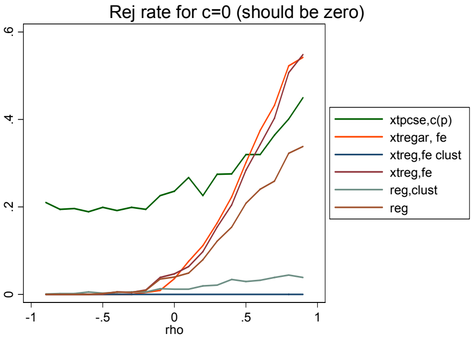
Mean rejection rates of $\beta = 1$ with nominal size 0.05

rho	reg	xtreg, fe	xtpcse	xtregar	reg, clust	xtreg, fe clust
-.9	0	0	.15	0	.05	.05
-.5	.006	.006	.162	0	.039	.039
-.1	.033	.037	.184	.037	.055	.055
0	.043	.053	.194	.06	.054	.054
.1	.053	.065	.206	.054	.043	.043
.5	.095	.156	.192	.069	.052	.052
.9	.055	.243	.21	.094	.039	.039

Rej Rates, AR(1) Errors

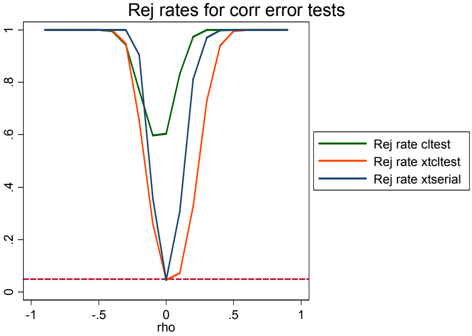


Rej Rates, AR(1) Errors



Tests for Clustering, AR(1) Errors

The test for clustering after **reg** is **cltest**, and the test for clustering after **xtreg, fe** is **xtcltest** (to be available from SSC shortly). It performs nearly as well as **xtserial** (which by construction is the correct test for this particular variety of clustering):



More Examples and Simulations?

We plan to turn this talk into a Stata Journal submission. Any suggestions on additional topics that you feel should be included are welcomed—contact Austin at austinnichols@gmail.com if you like.

References

Arellano, Manuel. 1987. "Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics,* 49: 431-34.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics,* 119(1): 249-275.

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. 2007. "Enhanced routines for instrumental variables/GMM estimation and testing." Unpublished working paper, forthcoming.

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. 2003. "Instrumental variables and GMM: Estimation and testing." *Stata Journal*, StataCorp LP, vol. 3(1), 1-31. Also Boston College Department of Economics Working Paper No 545

Binder, D. A. 1983. "On the variances of asymptotically normal estimators from complex surveys." *International Statistical Review,* 51: 279-292.

Binder, D. A. and Z. Patak. 1994. "Use of estimating functions for estimation from complex surveys." *Journal of the American Statistical Association,* 89(427): 10351043.

Cameron, Colin A., Jonah Gelbach, and Douglas L. Miller. 2006a. "Robust Inference with Multi-way Clustering." NBER Technical Working Paper No. 327

Cameron, Colin A., Jonah Gelbach, and Douglas L. Miller. 2006b. "Bootstrap-Based Improvements for. Inference with Clustered Errors."

Diggle, Peter J., Patrick Heagerty, Kung-Yee Liang, and Scott L. Zeger. 2002. *Analysis of Longitudinal Data, Second Edition.* Oxford University Press.

Drukker, David M. 2003. "Testing for serial correlation in linear panel-data models." *Stata Journal,* (3)2: 1-10.

Eicker, F. 1967. "Limit theorems for regressions with unequal and dependent errors." *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability,* Berkeley: University of California Press, 1: 59-82.

Froot, K. A. 1989. "Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data." *Journal of Financial and Quantitative Analysis,* 24: 333355.

Fuller, W. A. 1975. "REgression analysis for sample survey." *Sankhyā, Series C* 37: 117-132.

Gail, M. H., W. Y. Tan, and S. Piantodosi. 1988. "Tests for no treatment effect in randomized clinical trials." *Biometrika* 75: 57-64.

Hedges, Larry V. 2006. "Effect Sizes in Cluster-Randomized Designs." IPR WP-06-13.

Huber, P. J. 1967. "The behavior of maximum likelihood estimates under non-standard conditions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* Berkeley: University of California Press, 1, 221-233.

Kent, J. T. 1982. "Robust properties of likelihood ratio tests." *Biometrika,* 67: 19-27.

Kézdi, Gábor. 2004. "Robust Standard Error Estimation in Fixed-Effects Panel Models." *Hungarian Statistical Review* Special(9): 96-116.

Kish, Leslie. 1965. *Survey Sampling.* New York, NY: John Wiley and Sons.

Kish, Leslie and M. R. Frankel. 1974. "Inference from complex samples." *Journal of the Royal Statistical Society, Series B* 36: 1-37.

Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika,* 73: 13-22.

Lin, D. Y. and L. J. Wei. 1989. "The robust inference for the Cox proportional hazards model." *Journal of the American Statistical Association,* 84: 1074-1078.

Moulton, Brent R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *The Review of Economics and Statistics,* 72(2): 334-338.

Newey, W. K. and K. D. West. 1987. "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica,* 55: 703708.

Rabe-Hesketh, Sophia and Anders Skrondal. 2005. *Multilevel and Longitudinal Modeling Using Stata.* College Station, TX: Stata Press.

Rogers, William H. 1993. "sg17: Regression standard errors in clustered samples." *Stata Technical Bulletin* 13: 19-23.

Royall, R. M. 1986. "Model robust confidence intervals using maximum likelihood estimators." *International Statistical Review,* 54: 221-226.

Schaffer, Mark E. 2007. "cltest: Stata module to test for clustering of residuals."

Sribney, William. 1998. "Comparison of standard errors for robust, cluster, and standard estimators." Stata website.

Stock, James H. and Mark W. Watson. 2006. "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression." NBER Technical Working Paper 323.

White, Halbert. 1980. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica,* 48: 817-830.

White, Halbert. 1982. "Maximum likelihood estimation of misspecified models." *Econometrica,* 50: 1-25.

Williams, R. L. 2000. "A note on robust variance estimation for cluster-correlated data." *Biometrics,* 56: 645646.

Wooldridge, Jeffrey M. 2003. "Cluster-Sample Methods in Applied Econometrics." *The American Economic Review,* 93(2): 133-138.

Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press. Available from Stata.com's bookstore.