Web Appendix to "Empirical Strategies for Various Manifestations of Multilevel Data"

Robert J. Franzese, Jr.
Associate Professor of Political Science
The University of Michigan, Ann Arbor
franzese@umich.edu
http://www-personal.umich.edu/~franzese

Within, Between, and Shrinkage Estimators

We should also consider two other approaches for multilevel models, the within and between and the shrinkage estimators. Fixed-effects or least-squares dummy-variables (LSDV), closely related to the within estimator, is just the micro model with the full set of macro-level indicators (e.g., country dummies): $y_{ij}=\beta_{0j}+\beta_1 x_{ij}+\varepsilon_{ij}$. Without a contextually variant estimate of the effect of x_{ij}, i.e., without interactions of these macro-indicators or of z_j with x_{ij}, LSDV estimates will be BLUE *iff* $\gamma_{11}=0$, $u_{1j}=0$, and $V(\boldsymbol{\varepsilon})=\sigma^2\boldsymbol{I}$, i.e., if and only if the effect of x_{ij} does not vary across macro-units j deterministically or stochastically and the sole remaining error component,[1] ε_{ij}, is spherical. If the first condition, $\gamma_{11}=0$, is violated, then the LSDV suffers omitted-variable bias; context conditionality must be modeled. If second, $u_{1j}=0$, is violated, then we have the same random-coefficient situation seen in other pooled-OLS cases, and the same FGLS or consistent-standard-error strategies are advised. Violation of the third condition $V(\boldsymbol{\varepsilon})=\sigma^2\boldsymbol{I}$ is also as before, with the standard non-spherical-errors redresses again emerging (FGLS and/or consistent standard-errors). The conditional-mean effects of z_j, i.e. γ_{01}, are not directly retrievable in LSDV, being subsumed within the fixed effects, $\hat{\beta}_{0j}$, but a second-stage estimate regressing these on z_j can retrieve them (Jusko and Shively or Lewis and Linzer). We already discussed above the full within estimator, which allows both intercepts and coefficients to differ arbitrarily across j, calling it the dummy-interaction model. The between estimator, which simply regresses the

[1] The $u_{0j}=0$ by inclusion of the j dummies.

within j averages of y_{ij} on z_j and within-j averages of x_{ij} (and possibly also within-j averages of $z_j x_{ij}$), on the other hand, is obviously wasteful, potentially tremendously so, as it destroys all information from variations of $\{y_{ij},x_{ij}\}$ around $\{\bar{y}_j, \bar{x}_j\}$. It is "accurately inefficient" thusly *iff* each i within j is literally a replicant. Furthermore, this estimate is subject to potential "ecological fallacy" if one attempts to infer micro-level behavior from these aggregate-level correlations. Thus, unless J is very large and I very small, or each i is a replicant so in truth $I=1$, the between estimator would rarely be advisable on its own.

The last class of estimators, HLM, random-effects/random-coefficients, or *shrinkage* estimators, are weighted-average compromises of within and between estimators that *shrink* the full within (i.e., dummy-interaction) estimates toward between estimates (i.e., the regression of j averages) based on their relative variances. Intuitively, shrinkage estimators "borrow strength" from the other (~j) units to enhance the β_{0j} and β_{1j} estimates. Bowers and Drake (this volume) discuss HLM extensively, so we will not. We do want to note some important simulation results regarding such estimators due to Beck and Katz (2005) though. Their results derive from simulations of pure random-effects/random-coefficients without clustering and $J=20$, but a few important implications apply and strongly merit mention here. First, they find that shrinkage estimators do not, in fact, shrink noticeably from unit-by-unit OLS if I is large, so, in cross-country survey data for example, HLM estimates of $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$ will not differ much from their separate-subsample estimates. Second, by a mean-squared-error (of coefficient-estimates) criterion, with *unmodeled*[2] parameter heterogeneity, $V(u_{1j})$ across j, set to 1.8, fully pooled OLS—i.e., constant, equal coefficients on x_{ij} across all j—dominates or is about equivalent to the various shrinkage estimators if I is quite small (less than about 20). By I of about 40, conversely,

[2] Beck and Katz (2005) compare constant-, i.e. context-*un*conditional-, coefficient pooled-OLS to separate-subsample OLS. Accordingly, all of the parameter hereogeneity that they consider in their simulations is unmodeled heterogeneity with regard to our purposes of comparing pooled linear-interactive OLS to separate-subsample OLS.

unit-by-unit OLS—i.e., arbitrarily differing coefficients across j by separate-sample or dummy-interaction—has become nearly equivalent to the shrinkage estimators. Analogously, if *unmodeled* parameter heterogeneity is small then fully-pooled OLS dominates (less than about 1.5) or is about equivalent (less than nearly 2); if $V(u_{1j})$ large (about 4), then unit-by-unit OLS is about equivalent to the shrinkage estimators. All of this leaves a range, generously, of about I=10-40 and *unmodeled* $V(\beta_{1j})$=2-4 in which HLM might yield noticeable mean-squared-error advantages, with $J=20$ that is. Analytic results and other simulation studies have shown far greater utility for such models in samples with much greater numbers of macro-level units.