Clustered and Panel Data

Brenton Kenkel — PSCI 8357
March 17, 2016

Today we will go through the most popular method for data on units within larger groups, of which panel data is one example. There are numerous considerations for data like this. We could spend an entire semester on panel methods just for linear models, then another on the nonlinear case. The coverage here is just meant to get your feet wet in the world of panel methods, not to give you all the answers. Before you do any real-world research with panel data, I would urge you to look carefully at Wooldridge (2002) or another full-length treatment of panel methods.

Grouped Data

Assume that for each of $i = 1, \ldots, N$ groups we observe $j = 1, \ldots, J$ units. For example, we might observe individual students grouped by classrooms. Or we might observe the same units repeatedly over time—this is called *panel data*, usually replacing j and J with the "time" indices t and T.[1]

Our notation changes a bit with grouped data.

- Unit-level
    - Y_{ij}: the response for the j'th unit of group i (scalar)
    - $x_{ij} = (x_{ij1}, \ldots, x_{ijp})$: the covariates for the j'th unit of group i ($p \times 1$ vector)
- Group-level
    - $Y_i = (Y_{i1}, \ldots, Y_{iJ})$: the collection of responses for group i ($J \times 1$ vector)
    - $\mathbf{X}_i$: the collection of covariates for group i ($J \times p$ matrix)
- Full data

---

[1]Everything we do here would go through if we allowed the number of observations to vary across groups, but the notation would get uglier. In the panel data context we say that the panels are *balanced* if the number of observations is the same for each group and *unbalanced* otherwise.

- $Y = (Y_1, \ldots, Y_N)$: all responses ($NJ \times 1$ vector)
- $\mathbf{X}$: all covariates ($NJ \times p$ matrix)
- $\mathbf{D}$: matrix of dummy variables for group membership ($NJ \times N$ matrix)
- $\mathbf{Z} = \begin{bmatrix} \mathbf{X} & \mathbf{D} \end{bmatrix}$: matrix of all covariates and group membership dummies ($NJ \times (p + N)$ matrix)

Don't get intimidated by all the new notation—it's just book-keeping. If we assume as usual that the conditional expectation of the response is a linear function of the covariates, we have a familiar-looking linear model,

$$Y_{ij} = x_{ij}^\top \beta + \epsilon_{ij},$$

where ϵ_{ij} is unit-specific noise or random error.

The reason we need special methods for grouped data is that the standard Gauss-Markov assumptions are often implausible in a grouped setting. At a minimum, it is unlikely that ϵ_{ij} has constant variance and is uncorrelated across observations. In most applications, we would expect some correlation within groups that cannot be fully explained as a function of observables. As with heteroskedasticity, this means our estimated standard errors are wrong and need to be corrected.

If within-group error correlations are the only issue, then OLS still gives an unbiased and consistent estimate of β, even if the standard errors are troublesome. It gets worse if there is unobserved heterogeneity between groups that is correlated with the covariates. For example, imagine that we have data on students grouped by classrooms, and we are modeling test scores as a function of household income. If students from higher-income families are more likely to be placed with high-ability teachers and teacher ability is unobservable, then OLS is a poor estimator. It will not even be unbiased or consistent, and we must use a different estimator.

Clustered Standard Errors

Let us imagine an example of the first, less serious problem: errors that are not independent across observations within the same group. Suppose that half of the U.S. states received a federal grant to improve elementary education, and the other half did not. Our unit of observation is school districts,

grouped within states. We want to know whether receiving the grant affected test scores. What we cannot measure is how well the states used the money. If some states used it better than others, then we would expect the error term to be correlated across school districts within the state. In a state where the money was used wisely, we would expect most of the schools to do "better than expected"—to have positive residuals, in the language of regression. Conversely, in a state where the money was squandered, we would expect most of the schools to do worse than we would otherwise predict.

This is one example of the general phenomenon Moulton (1986; 1990) identifies—that there is often substantial correlation in the random errors within groups, especially when we are looking at the effects of variables that only vary at the group level. One way I think about it is that, with grouped data and group-level covariates, the effective number of observations is less than the nominal number of observations. If we use the OLS standard errors, we are pretending to have more data than we really do.

When the errors are correlated within groups (and possibly heteroskedastic within and across groups), we can continue to get our coefficient estimates from OLS. We just need to correct the standard errors. The methodology is similar to that of Huber-White standard errors under heteroskedasticity. Assume that the variance matrix of the error term ϵ is a block-diagonal matrix of the form

$$\Omega = \begin{bmatrix} \Omega_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Omega_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Omega_N \end{bmatrix},$$

where each Ω_i is a $J \times J$ matrix of the form

$$\Omega_i = \begin{bmatrix} \sigma_{i1}^2 & \rho_{i12}\sigma_{i1}\sigma_{i2} & \cdots & \rho_{i1J}\sigma_{i1}\sigma_{iJ} \\ \rho_{i12}\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2 & \cdots & \rho_{i1J}\sigma_{i1}\sigma_{iJ} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{i1J}\sigma_{i1}\sigma_{iJ} & \rho_{i2J}\sigma_{i2}\sigma_{iJ} & \cdots & \sigma_{iJ}^2 \end{bmatrix}.$$

If we knew Ω, we could use generalized least squares (a generalization of weighted least squares that we will talk about momentarily) to efficiently estimate β and correctly estimate the standard errors. However, even without

knowing Ω, we may obtain an asymptotically correct estimate of the OLS standard errors, just as we did under heteroskedasticity. The *cluster-robust variance matrix estimate* is

$$\hat{\Sigma}_{\mathrm{CR}} = (\mathbf{X}^\top \mathbf{X})^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i^\top \hat{e}_i \hat{e}_i^\top \mathbf{X}_i \right) (\mathbf{X}^\top \mathbf{X})^{-1}.$$

The earliest derivation of this estimator was Liang and Zeger (1986), and that is whom you should cite when you use it. Arellano (1987) derives the same estimator but is a bit easier to follow. As with the heteroskedasticity-consistent standard error estimator, there are various finite-sample corrections to $\hat{\Sigma}_{\mathrm{CR}}$ that you may want to use; we will not go through those here.

In practice, this estimator is useful only when the number of groups, N, is large. The consensus seems to be that $N = 50$ is large enough (Cameron and Miller 2015), so you Americanists may go ahead and rejoice. However, if you observe a small number of units over a long period of time, the cluster-robust standard error estimator will be severely biased. Beck and Katz (1995) provide a similar estimator for such data, except with the summation in the middle of the "sandwich" taken over the J time periods instead of the N groups.

ASIDE: Never Use Clustered Standard Errors in Logit/Probit

For OLS estimation of linear models, clustered standard errors "work" because $\hat{\beta}_{\mathrm{OLS}}$ is a consistent estimator of β even under non-spherical errors. This is not true for logit, probit, and similar models. If there are correlated errors across observations in a logistic regression model, the usual estimator of β (the maximum likelihood estimator) is *inconsistent*. No matter how big our sample size, it won't necessarily yield a result close to the true coefficients. There is no point in getting a more accurate estimate of the standard errors of such an estimator, because the estimator itself is crap. See Freedman (2006) for more on this point.

Random Effects

Just as Huber-White standard errors do not fix the inefficiency of OLS under heteroskedasticity, cluster-robust standard errors do not fix the inefficiency of

OLS under within-group correlation. To get a handle on the efficiency problem, we will make further assumptions about the source of that correlation.

Suppose that each group has a different intercept, with α_i denoting the intercept for group i. (From here on we will assume, contrary to our usual practice, that x_{ij} does not contain an intercept.) In other words, the baseline level of the response varies across groups. We can then decompose the error term ϵ_{ij} into the group-specific effect α_i and an idiosyncratic unit-level error η_{ij}, giving us the model

$$Y_{ij} = x_{ij}^\top \beta + \underbrace{\alpha_i + \eta_{ij}}_{\epsilon_{ij}}.$$

We will assume that η_{ij} is independent and identically distributed across observations, so the within-group correlation comes exclusively from the fact that the group effect α_i enters the error term for every observation in the group. Let $V[\alpha_i] = \sigma_\alpha^2$ and $V[\eta_{ij}] = \sigma_\eta^2$. Under this model, the covariance structure is

$$\mathrm{Cov}[\epsilon_{ij}, \epsilon_{i'j'}] = \begin{cases} \sigma_\eta^2 + \sigma_\alpha^2 & i = i', j = j', \\ \sigma_\alpha^2 & i = i', j \neq j', \\ 0 & i \neq i'. \end{cases}$$

This is equivalent to assuming Ω takes the block-diagonal form given above, with each Ω_i along the diagonal being

$$\Omega_i = \begin{bmatrix} \sigma_\eta^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\eta^2 + \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\eta^2 + \sigma_\alpha^2 \end{bmatrix}.$$

If we knew σ_α^2 and σ_η^2, we could efficiently estimate β via the *generalized least squares* estimator,

$$\hat{\beta}_{\mathrm{GLS}}(Y, \mathbf{X}, \Omega) = (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} Y.$$

As you surely noticed, this looks identical to the weighted least squares estimator, because WLS is the special case of GLS in which Ω is a diagonal matrix. Anyway, the problem is, we don't know the two error variances. But we can estimate them. In the interests of time, I will skip over exactly how we do this—you can find the details in Johnston and DiNardo (1997, chap. 12). If

we replace σ_α^2 and σ_η^2 in Ω with their corresponding estimates and call the resulting matrix $\hat{\Omega}$, we can use the *feasible GLS estimator*,

$$\hat{\beta}_{\text{FGLS}}(Y, \mathbf{X}, \hat{\Omega}) = (\mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\Omega}^{-1} Y.$$

We call this application of feasible GLS the *random effects estimator*, or $\hat{\beta}_{\text{RE}}$. As long as we have correctly specified the form of Ω and the usual exogeneity condition holds ($E[\epsilon_{ij} | x_{ij}] = 0$), then the random effects estimator is consistent and asymptotically efficient: in large enough samples it gets close to the right answer and no other linear estimator gets closer on average. The small-sample properties are generally unknown.

But I have tried to slip something past you here. I referred to the "usual exogeneity condition" like a bit of a throwaway. But this is unduly restrictive when ϵ_{ij} contains a group-specific effect. It implies that the group-specific effects must be uncorrelated with all of the covariates—a condition that is at best questionable in most observational data settings. Recall the earlier example where kids from higher-income families get placed in classrooms with higher-ability teachers. If $\text{Cov}[x_{ij}, \alpha_i] \neq 0$, then the random effects estimator will be inconsistent. If this is so, we need to bite the bullet and estimate the α_i's.

Fixed Effects

The *fixed effects estimator* is much easier to define than the random effects estimator. It is simply the regression of Y on $\mathbf{Z}$, the matrix that contains both the covariates and the group membership dummies:

$$\hat{\beta}_{\text{FE}}(Y, \mathbf{Z}) = \hat{\beta}_{\text{OLS}}(Y, \mathbf{Z})[1, \ldots, p],$$

in other words, only the first p elements of the $p + N$ OLS estimates. Fixed effects takes the part of the error term that is correlated with the covariates—the group-specific intercepts—and brings them into the covariate matrix, purging the source of omitted variable bias that arises under OLS or random effects.

Most textbook treatments of fixed effect estimators go through a whole rigmarole about computation, because inverting the $(N + p) \times (N + p)$ matrix $\mathbf{Z}^\top \mathbf{Z}$

6

used to be impossible for large N. This is no longer true,[2] so you can safely ignore most of the hand-wringing about computational difficulties.

The standard errors of the fixed effects estimator are usually higher than those of the random effects estimator, since estimating N additional parameters uses a lot of degrees of freedom. This leads us to the following pair of observations.

- If the random effects assumption is met (group-specific effects are uncorrelated with covariates), then the random effects and fixed estimators are both consistent, but fixed effects is less efficient.

- If the random effects assumption is not met, then the random effects estimator is inconsistent while the fixed effects estimator is consistent.

The typical test for whether fixed effects are necessary comes from Hausman (1978). Under the null hypothesis that both estimators are consistent (and thus fixed effects are unnecessary and inefficient), the test statistic

$$H = (\hat{\beta}_{\mathrm{RE}} - \hat{\beta}_{\mathrm{FE}})^\top (\hat{\Sigma}_{\mathrm{FE}} - \hat{\Sigma}_{\mathrm{RE}})^{-1} (\hat{\beta}_{\mathrm{RE}} - \hat{\beta}_{\mathrm{FE}})$$

asymptotically has a χ^2 distribution with p degrees of freedom.

The other main drawback of fixed effects estimation is that you cannot estimate the effects of variables that do not vary within groups. (Why not?) See Greene (2003, sec. 13.5) for estimation strategies with panel data and time-invariant covariates.

One final note: Arellano (1987) shows that the cluster-robust variance matrix estimator can be used with fixed effects. Cameron and Miller (2015) recommend doing so, at least when the number of groups is large.

Summary

- Cluster your standard errors when estimating the effect of a group-level variable with individual-level data, or any other time you suspect correlated errors within groups. This advice applies *only for linear models*.

---

[2]I simulated a dataset with a single covariate, $N = 1,000$, and $J = 1,000$, meaning the fixed effects regression entailed estimating 1,001 parameters from 1,000,000 observations. It took less than two minutes to run on my not-especially-powerful laptop. That said, should you ever have N another order of magnitude above that, the computational tricks given in the textbook treatments will once again be of use.

- Use random effects when it is plausible the group-level effects are uncorrelated with the covariates (i.e., rarely).
- Use fixed effects when the group-level effects may be correlated with the covariates (i.e., usually).

Appendix: Implementation

The methods introduced here can be implemented via the **plm** package.

```r
library("plm")
```

We will use the `Produc` dataset from **plm**, a riveting collection of economic statistics about the U.S. states from 1970 to 1986.

```r
data(Produc, package = "plm")
head(Produc)
```

```
##      state year region  pcap  hwy water util    pc   gsp   emp unemp
## 1 ALABAMA 1970      6 15033 7326  1656 6051 35794 28418 1010   4.7
## 2 ALABAMA 1971      6 15502 7526  1721 6255 37300 29375 1022   5.2
## 3 ALABAMA 1972      6 15972 7765  1765 6442 38670 31303 1072   4.7
## 4 ALABAMA 1973      6 16406 7908  1742 6756 40084 33430 1136   3.9
## 5 ALABAMA 1974      6 16763 8026  1735 7002 42057 33749 1170   5.5
## 6 ALABAMA 1975      6 17316 8158  1752 7406 43972 33604 1155   7.7
```

The functions in **plm** assume that your data are organized like `Produc`, with the grouping variable in the first column and the identification variable (time, in the case of panel data) in the second column. See the **plm** package vignette on how to get datasets not organized this way into line.

We will treat unemployment (`unemp`) as our response and public capital stock (`pcap`) and private capital stock (`pc`) as our covariates. As a benchmark, let's use OLS.

```r
fit_ols <- lm(unemp ~ pcap + pc,
              data = Produc)
summary(fit_ols)
```

```
##
## Call:
```

```
## lm(formula = unemp ~ pcap + pc, data = Produc)
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -3.559 -1.622 -0.337  1.213 11.595
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.20e+00   1.08e-01   57.57   <2e-16
## pcap        1.01e-05   5.52e-06    1.82    0.069
## pc          2.56e-06   2.56e-06    1.00    0.319
##
## Residual standard error: 2.2 on 813 degrees of freedom
## Multiple R-squared:  0.0352, Adjusted R-squared:  0.0328
## F-statistic: 14.8 on 2 and 813 DF,  p-value: 4.79e-07
```

The "pooling" estimator implemented by `plm()` ought to give us the same results.

```
fit_pooling <- plm(unemp ~ pcap + pc,
                   data = Produc,
                   model = "pooling")
summary(fit_pooling)
```

```
## Oneway (individual) effect Pooling Model
##
## Call:
## plm(formula = unemp ~ pcap + pc, data = Produc, model = "pooling")
##
## Balanced Panel: n=48, T=17, N=816
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
##  -3.560  -1.620  -0.337   1.210  11.600
##
## Coefficients :
##             Estimate Std. Error t-value Pr(>|t|)
## (Intercept) 6.20e+00   1.08e-01   57.57   <2e-16
## pcap        1.01e-05   5.52e-06    1.82    0.069
## pc          2.56e-06   2.56e-06    1.00    0.319
##
```

```
## Total Sum of Squares:    4060
## Residual Sum of Squares: 3920
## R-Squared:      0.0352
## Adj. R-Squared: 0.035
## F-statistic: 14.8141 on 2 and 813 DF, p-value: 4.79e-07
```

We can obtain the cluster-robust variance matrix estimate via vcovHC(). Make
sure to specify type = "arellano" so as to get the usual estimator. It is not
entirely clear to me which of the various finite-sample adjustments corresponds
to the defaults in Stata.

```
crvm_pooling <- vcovHC(fit_pooling,
                       method = "arellano",
                       type = "HC1")
summary(fit_pooling, .vcov = crvm_pooling)
```

```
## Oneway (individual) effect Pooling Model
##
## Call:
## plm(formula = unemp ~ pcap + pc, data = Produc, model = "pooling")
##
## Balanced Panel: n=48, T=17, N=816
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
##  -3.560  -1.620  -0.337   1.210  11.600
##
## Coefficients :
##             Estimate Std. Error t-value Pr(>|t|)
## (Intercept) 6.20e+00   2.45e-01   25.34   <2e-16
## pcap        1.01e-05   1.21e-05    0.83     0.41
## pc          2.56e-06   7.30e-06    0.35     0.73
##
## Total Sum of Squares:    4060
## Residual Sum of Squares: 3920
## R-Squared:      0.0352
## Adj. R-Squared: 0.035
## F-statistic: 14.8141 on 2 and 813 DF, p-value: 4.79e-07
```

Notice that our t-statistics are nearly cut in half, even though our variables
have within-group variation (unlike Moulton (1990)'s example).

We can also use `plm()` to estimate a random-effects model.

```r
fit_random <- plm(unemp ~ pcap + pc,
                  data = Produc,
                  model = "random")
summary(fit_random)
```

```
## Oneway (individual) effect Random Effect Model
##     (Swamy-Arora's transformation)
##
## Call:
## plm(formula = unemp ~ pcap + pc, data = Produc, model = "random")
##
## Balanced Panel: n=48, T=17, N=816
##
## Effects:
##                  var std.dev share
## idiosyncratic 3.09    1.76  0.69
## individual    1.36    1.17  0.31
## theta:  0.656
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
##  -3.640  -1.360  -0.286   1.010   9.730
##
## Coefficients :
##             Estimate Std. Error t-value Pr(>|t|)
## (Intercept) 5.67e+00   2.53e-01   22.42  < 2e-16
## pcap        1.55e-09   1.12e-05    0.00  0.99989
## pc          1.59e-05   4.55e-06    3.51  0.00048
##
## Total Sum of Squares:    2920
## Residual Sum of Squares: 2800
## R-Squared:      0.0421
## Adj. R-Squared: 0.042
## F-statistic: 17.88 on 2 and 813 DF, p-value: 2.52e-08
```

And, finally, a fixed-effects model, which `plm()` calls the `"within"` model.

```r
fit_fixed <- plm(unemp ~ pcap + pc,
                 data = Produc,
```

```
                model = "within")
summary(fit_fixed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = unemp ~ pcap + pc, data = Produc, model = "within")
##
## Balanced Panel: n=48, T=17, N=816
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
##  -3.810  -1.170  -0.248   0.947   8.390
##
## Coefficients :
##      Estimate Std. Error t-value Pr(>|t|)
## pcap 2.27e-04   3.18e-05    7.14  2.2e-12
## pc   4.18e-06   6.57e-06    0.64     0.52
##
## Total Sum of Squares:    2770
## Residual Sum of Squares: 2370
## R-Squared:       0.144
## Adj. R-Squared: 0.135
## F-statistic: 64.2646 on 2 and 766 DF, p-value: <2e-16
```

We can extract the fixed-effect estimates themselves via fixef().

```
fixef(fit_fixed)
```

```
##       ALABAMA         ARIZONA         ARKANSAS        CALIFORNIA
##         3.764           3.201            5.021           -24.352
##      COLORADO      CONNECTICUT         DELAWARE           FLORIDA
##         1.907           2.223            5.724            -2.950
##       GEORGIA           IDAHO         ILLINOIS           INDIANA
##         0.091           5.926           -7.483             1.499
##          IOWA          KANSAS         KENTUCKY         LOUISIANA
##         1.176           1.258            2.487             2.395
##         MAINE        MARYLAND    MASSACHUSETTS          MICHIGAN
##         6.133          -0.469           -0.081            -2.073
##     MINNESOTA     MISSISSIPPI         MISSOURI           MONTANA
##        -0.913           4.790            0.576             5.399
```

```
##       NEBRASKA          NEVADA NEW_HAMPSHIRE      NEW_JERSEY
##          1.034           6.122         3.867          -0.476
##     NEW_MEXICO        NEW_YORK NORTH_CAROLINA    NORTH_DAKOTA
##          5.876         -22.417         0.589           3.715
##           OHIO        OKLAHOMA        OREGON    PENNSYLVANIA
##         -5.208           2.212         4.914          -6.741
##   RHODE_ISLAND  SOUTH_CAROLINA   SOUTH_DAKOTA        TENNESSE
##          6.046           3.895         2.943           1.080
##          TEXAS            UTAH       VERMONT        VIRGINIA
##        -11.161           4.234         5.624          -0.961
##     WASHINGTON   WEST_VIRGINIA      WISCONSIN         WYOMING
##          1.081           6.941         0.192           3.799
```

If we wanted to include time dummies as well, we could specify `effect = "twoways"` in the fitting function.

```
fit_fixed_2 <- plm(unemp ~ pcap + pc,
                   data = Produc,
                   effect = "twoways",
                   model = "within")
summary(fit_fixed_2)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = unemp ~ pcap + pc, data = Produc, effect = "twoways",
##     model = "within")
##
## Balanced Panel: n=48, T=17, N=816
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -3.3300 -0.8030 -0.0449  0.7500  6.0200
##
## Coefficients :
##        Estimate Std. Error t-value Pr(>|t|)
## pcap  9.10e-05   2.62e-05    3.47  0.00054
## pc   -1.27e-05   5.04e-06   -2.51  0.01216
##
## Total Sum of Squares:    1260
## Residual Sum of Squares: 1240
```

```
## R-Squared:       0.0164
## Adj. R-Squared: 0.015
## F-statistic: 6.23517 on 2 and 750 DF, p-value: 0.00206
```

```
fixef(fit_fixed_2, effect = "time")
```

```
## 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983
## 3.55 4.28 3.77 3.19 3.83 6.52 5.46 4.94 3.94 3.82 5.18 5.67 7.70 7.68
## 1984 1985 1986
## 5.64 5.49 5.35
```

phtest() implements the Hausman test. Remember that the null hypothesis is that both estimators are consistent; if we reject it, then the random effects estimator is inconsistent and we must use fixed effects.

```
phtest(fit_random, fit_fixed)
```

```
##
##  Hausman Test
##
## data:  unemp ~ pcap + pc
## chisq = 100, df = 2, p-value <2e-16
## alternative hypothesis: one model is inconsistent
```

References

Arellano, M. 1987. "Practitioners Corner: Computing Robust Standard Errors for Within-groups Estimators." *Oxford Bulletin of Economics and Statistics* 49 (4): 431–34.

Beck, Nathaniel, and Jonathan N Katz. 1995. "What to do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89 (03): 634–47.

Cameron, A Colin, and Douglas L Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–72.

Freedman, David A. 2006. "On The So-Called Huber Sandwich Estimator and

Robust Standard Errors." *The American Statistician* 60 (4): 299–302.

Greene, William H. 2003. *Econometric Analysis*. 5th ed. Prentice Hall.

Hausman, J A. 1978. "Specification Tests in Econometrics." *Econometrica* 46 (6): 1251.

Johnston, John, and John DiNardo. 1997. *Econometric Methods*. 4th ed. McGraw-Hill.

Liang, Kung-Yee, and Scott L Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73 (1): 13.

Moulton, Brent R. 1986. "Random group effects and the precision of regression estimates." *Journal of Econometrics* 32 (3): 385–97.

———. 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.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.