Lecture 7: Multilevel and TSCS Models

Dave Armstrong

Western University
Department of Political Science

e: dave.armstrong@uwo.ca

Overview

Broadly speaking, we're going to worry about a few different sets of things today:

1. What are the consequences for our models of having *dependent* data?
2. When we have dependent data, we can make hypotheses about within-unit effects and between-unit effects, what are the consequences of this choice?
3. When we use unit effects to deal with dependent data, are we better off using fixed or random unit effects?
4. How do things get complicated when time is an important variable in our analysis?

Dependent Data

Multilevel data are different than the data we have considered thus far as they have observed data at different levels of aggregation. Some examples:

- Children in classrooms, in schools, in districts, in states.
- Voters in towns, in states
- Democracy measured at different years within each country.

The prominent feature here is that observations can be thought of as "nested" in groups.

- These groups can have attributes of their own that we might care about.

The Independence Assumption

- Recall that OLS assumes that each observation is independent of the others
  - More specifically, the error term, or equivalently the Y-values, are independent of each other
- Although the assumption of independence is rarely perfect, in practice a random sample from a large population provides a close approximation
  - Time-series data, panel data and clustered data often do not satisfy this condition
    - In these cases, dependencies among the errors can be quite strong
- If independence is violated, OLS is no longer the optimal estimation method as standard errors are biased downwards

What happens if we ignore the multilevel nature of the data?

- Calculation of standard errors involves consideration of the sample size in the denominator of the formula:

$$
\begin{aligned}
SE(\bar{x}) &= \frac{S_x}{\sqrt{n}} \\
SE(B) &= \frac{S_E}{\sqrt{\sum (x_i - \bar{x})^2}} \\
&\equiv \frac{S_y}{S_x}\sqrt{\frac{1 - r_{xy}^2}{n - 2}}
\end{aligned}
$$

- Answer to Q1: When the observations are not independent, the effective sample size is smaller than what we observe and thus an adjustment must be made to the standard errors or they will be biased downwards.

Within and Between Effects

The choice of estimator for grouped data comes down to a choice about what sorts of things you want to know. If you want to know something about between unit variance, then you need a between estimator, e.g.:

$$\bar{Y}_{\cdot j} = \delta_0 + \delta_1 \bar{X}_{\cdot j} + \nu_j$$

However, if you want to know something about the relationship of Y and X *within* groups (i.e., among individuals in each group), then you need a within estimator, e.g.:

$$(Y_{ij} - \bar{Y}_{\cdot j}) = \delta_0 + \delta_1 (X_{ij} - \bar{X}_{\cdot j}) + (u_{ij} - \bar{u}_{\cdot j})$$

Some models are compromises that allow us to learn something about both types of relationships.

Snijders and Bosker's Example (1)

- Assume the following two-level artificial data, with 5 groups containing 2 observations each:

j	i	X_{ij}	$\bar{X}_{\cdot j}$	Y_{ij}	$\bar{Y}_{\cdot j}$
1	1	1	2	5	6
1	2	3	2	7	6
2	1	2	3	4	5
2	2	4	3	6	5
3	1	3	4	3	4
3	2	5	4	5	4
4	1	4	5	2	3
4	2	6	5	4	3
5	1	5	6	1	2
5	2	7	6	3	2

- We now fit several models: (1) a total regression, (2) a regression between group means, (3) within group regressions, (4) multilevel model

Snijders and Bosker's Example (2)

1. Total regression $(Y_{ij} \sim X_{ij})$

$$\hat{Y}_{ij} = 5.33 - 0.33X_{ij}$$

2. Between Group Means $(Y_{\cdot j} \sim X_{\cdot j})$

$$\hat{Y}_{\cdot j} = 8 - 1\bar{X}_{\cdot j}$$

3. Within Groups $(Y_{ij} - \bar{Y}_{\cdot j} \sim X_{ij} - \bar{X}_{\cdot j})$

$$\hat{Y}_{ij} = \bar{Y}_{\cdot j} + 1(X_{ij} - \bar{X}_{\cdot j})$$

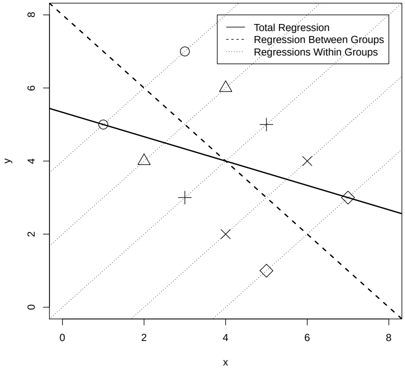
4. Multilevel Regression:

$$\hat{Y}_{ij} = 8.00 - 1.00\bar{X}_{\cdot j} + 1.00(X_{ij} - \bar{X}_{\cdot j})$$

- Equation 3 stems from the individual regressions of Y_i on X_i within each group. It simplifies to this equation only when the slopes for each group are the same
- The multilevel regression (4) writes Y_{ij} as a function of the within group and between group relations between Y and X

Snijders and Bosker's Example (3)

- Notice that both the total regression and the between group regression have done a poor job of capturing the trend in the data
- In fact the regression between groups is in the opposite direction to the individual within group regressions



Multilevel Regression Model: Varying Intercepts

Imagine we have the following model:

$$Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + R_{ij}$$

where

$$\beta_{0j} = \gamma_{00} + U_{0j}$$

The new equation is:

$$Y_{ij} = \gamma_{00} + \beta_1 x_{ij} + U_{0j} + R_{ij}$$

The U_{0j} are deviations in the group intercepts from the overall mean γ_{00}.

U_{0j} - fixed or random?

If we think of U_{0j} as fixed, we estimate a parameter for each of U_{0j} (save 1 for identification).

- These group level parameters are by definition going to be perfectly collinear with any explanatory variable that only varies by group.

If we think of U_{0j} as random, then we think of them as "group residuals", given $\boldsymbol{X}$.

- This is appropriate if groups are "exchangeable" - all drawn from a population of groups.
- There is *one* parameter associated with the groups in this model and that is the variance of the group effects.

Fixed or Random?

1. If the groups are of interest and researchers want to make inferences about the differences between groups, the fixed effects should be used.
2. If the groups are considered to be sampled from a population of groups and the research wants to make inferences pertaining to the population of groups, random effects should be used.
3. If the researcher wants to test propositions about the effects of group level variables - random coefficients should be used (we'll talk more about this later with Pluemper and Troger's piece).
4. If group sizes are small - random effects can help leverage strength across the groups if the assumption of exchangeability holds.
5. Random effects models work best when the assumption of approximate normality of U_{0j} and R_{ij} holds.

Bell and Jones

Bell and Jones make the following suggestion:

- Use both the within and between transformations to identify the different effects.

Mixed Models Example: British Context Data

- The data are a subset from the 1997 British Election Study
  - Collected using a multistage sample where first parliamentary constituencies were randomly selected and then voters were randomly selected from within the selected constituencies
  - There are 2141 observations nested within 136 constituencies
- The variables of interest are:
  - LRSCALE: left-right values scale
  - PROFMAN: % professionals in constituency
  - AGE
  - SEX
  - Degree (respondent obtained a univ degree)
  - INCOME (household income)
  - PANO: identifies the constituency

Hypotheses

1. Household income affects attitudes
   - More specifically, the richer one is, the more right-wing on average their attitudes will be
2. This relationship will differ depending on the area in which one lives - if they live in a rich constituency, they will be more likely to hold right-wing attitudes regardless of their own wealth
   - In other words, we expect their to be random effects for income

- We shall examine both whether the intercept varies across constituencies (indicating on average that overall attitudes differ) and whether the slope for income varies (indicating that income has different effects according to constituency)

Multilevel Regression Models: Within and Between Unit Effects

Hierarchical or Multilevel Models offer us a way to learn something about both within- and between-unit effects. We can formulate the model as follows:

$$Y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + U_{0j} + R_{ij}$$

Here:

- γ_{00} represents the grand mean of Y_{ij}
- γ_{10} is the coefficient relating x_{ij} to Y_{ij}
- U_{0j} is the unit-specific deviation from the grand-mean (a unit-specific intercept shift)
- R_{ij} is the observation-specific residual from the regression line $\gamma_{00} + \gamma_{10} x_{ij} + U_{0j}$

Modeling both Between and Within Effects

We can allow these to be different by including the group specific variables - in this case, the group mean of income:

```
library(foreign)
dat <- read.dta("http://quantoid.net/files/9591/context.dta")
betweendat <- make_between_data(LRSCALE ~ AGE + SEX + INCOME + PROFMAN, dat, id="PANO")

library(lme4)
mod <- lmer(LRSCALE ~ INCOME_b + INCOME_w + AGE_b + AGE_w +
    SEXmen_b + SEXmen_w + PROFMAN_b + (1|PANO), data=betweendat)
s <- summary(mod, corr=FALSE)
round(s$coefficients, 3)
```

```
##             Estimate Std. Error t value
## (Intercept)    9.410      1.021   9.219
## INCOME_b       0.341      0.058   5.848
## INCOME_w       0.157      0.020   7.831
## AGE_b          0.041      0.019   2.179
## AGE_w          0.026      0.005   5.236
## SEXmen_b       1.246      0.679   1.836
## SEXmen_w       0.148      0.161   0.921
## PROFMAN_b      0.032      0.015   2.117
```

Interpretation

- In the example above, the coefficient on `INCOME_w` term suggests that within groups, as individuals have higher levels of income, they have more right-wing attitudes.
- The Between-unit coefficient (`INCOME_b`) suggests that individuals who are in more affluent constituencies (as measured by the mean of income) have intercepts that are on average more right-wing. So, on the whole, richer places are more right-wing in nature.

Obtaining Estimates of U_{0j}

I will give the formula for the empty model (without variables). It gets increasingly more complicated with other data in the model, but **R** will do that for us.

$$
\begin{aligned}
\hat{U}_{0j} &= \lambda_j \hat{\beta}_{0j} + (1 - \lambda_j)\hat{\gamma}_{00} \\
\hat{\gamma_{00}} &= \sum_{j=1}^{N} \frac{n_j}{M} \bar{Y}_{\cdot j} \\
\beta_{0j} &= \bar{Y}_{\cdot j} \\
\lambda_j &= \frac{\tau_0^2}{\tau_0^2 + \frac{\sigma^2}{n_j}}
\end{aligned}
$$

The consequence of this estimation procedure is to pull all observations toward the estimated population mean of γ_{00}.

This is called an *Empirical Bayes* estimate of the unit effect. Now we're all Bayesians. Yay!

Estimates of U_{0j} from R

```r
head(ranef(mod)$PANO)
```

```
##          (Intercept)
##     2.00  0.03742999
##     5.00 -0.22809606
##    22.00 -0.47200641
##    24.00 -0.09535116
##    28.00 -0.01363689
##    33.00 -0.47573655
```

p-values

Note that none of the model summaries come with p-values (this is considered a feature rather than a flaw).[1] There are a few ways you can accomplish this.

```
library(LMERConvenienceFunctions)
pamer.fnc(mod)


##            Df   Sum Sq  Mean Sq F value upper.den.df upper.p.val
## INCOME_b    1 872.7245 872.7245 69.6000         2133      0.0000
## INCOME_w    1 540.3118 540.3118 43.0900         2133      0.0000
## AGE_b       1 101.4241 101.4241  8.0886         2133      0.0045
## AGE_w       1 349.6833 349.6833 27.8873         2133      0.0000
## SEXmen_b    1  40.8009  40.8009  3.2539         2133      0.0714
## SEXmen_w    1  10.6267  10.6267  0.8475         2133      0.3574
## PROFMAN_b   1  56.2162  56.2162  4.4833         2133      0.0343
##            lower.den.df lower.p.val expl.dev.(%)
## INCOME_b           1997      0.0000       2.8885
## INCOME_w           1997      0.0000       1.7883
## AGE_b              1997      0.0045       0.3357
## AGE_w              1997      0.0000       1.1574
## SEXmen_b           1997      0.0714       0.1350
## SEXmen_w           1997      0.3574       0.0352
## PROFMAN_b          1997      0.0344       0.1861
```

[1] https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

P-values II

There is a nice discussion online of a few different methods.[2]

```
library(lmerTest)
mod <- update(mod)
anova(mod)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##             Sum Sq Mean Sq NumDF   DenDF F value    Pr(>F)    
## INCOME_b    428.82  428.82     1  142.84 34.1982 3.247e-08 ***
## INCOME_w    768.89  768.89     1 1998.21 61.3188 7.814e-15 ***
## AGE_b        59.52   59.52     1  134.40  4.7468   0.03110 *  
## AGE_w       343.81  343.81     1 1998.21 27.4188 1.810e-07 ***
## SEXmen_b     42.28   42.28     1  137.27  3.3714   0.06850 .  
## SEXmen_w     10.63   10.63     1 1998.21  0.8475   0.35738    
## PROFMAN_b    56.22   56.22     1  129.03  4.4833   0.03615 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[2] http://mindingthebrain.blogspot.com/2014/02/three-ways-to-get-parameter-specific-p.html

Testing Equality of Between and Within Effects

```r
library(car)
linearHypothesis(mod, "INCOME_b = INCOME_w")
```

```
## Linear hypothesis test
##
## Hypothesis:
## INCOME_b - INCOME_w = 0
##
## Model 1: restricted model
## Model 2: LRSCALE ~ INCOME_b + INCOME_w + AGE_b + AGE_w + SEXmen_b + SEXmen_w +
##     PROFMAN_b + (1 | PANO)
##
##   Df  Chisq Pr(>Chisq)
## 1
## 2  1 8.9427   0.002786 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Time-Series Cross-Sectional Data

TSCS Data have both cross-sectional (data on different units) and time-series (data over time) properties. As such:

- For each unit (e.g., country) there are many different time-periods.
- For each different time-period, there are many different units

In the presence of such data structure, conventional cross-sectional results are *wrong*:

- At least inefficient - standard errors will be wrong.
- Could also be biased depending on the particular nature of the data.

Time-series Issues

- Non-independence: the assumption that all observations are independent is almost certainly violated.
  - Observations close in time for the same unit are often more alike each other than observations far away in time.
  - Consequence I: Serial correlation (errors are correlated with each other).
  - Consequence II: Model mis-specification: if y_t is a function of y_{t-1} and we don't include it, then we have mis-specified the model.
- Trending - series that tend to increase or decrease by the same amount over time.
  - Two series that are trending will tend to show a strong relationship, but the relationship could be spurious as both variables are a function of time.
  - Also violates the assumption of stationarity - finite, constant variance.
- Structural breaks - rapid and immediate changes in mean or variance

Autoregression

While there are lots of things we *could* care about with respect to time-series, for TSCS data, it's almost always about autoregression

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$$

- The α_0 allows the long-run equilibrium to be non-zero.

Assuming $|\alpha_1| < 1$, then

- The series will eventually, though not immediately, return to its long-term equilibrium after an intervention (without further interventions).

Autocorrelation

The correlation between temporally adjacent values of y:

$$Corr\left(y_t, y_{t-1}\right) \equiv \rho \equiv \frac{E\left(\left(y_t - \mu_t\right)\left(y_{t-1} - \mu_{t-1}\right)\right)}{E\left(y_t - \mu_t\right)^2}$$

We can estimate $\hat{\rho}$ by plugging $\bar{y}$ in for μ.

Stationarity

Covariance stationarity is defined by:

- $E(y_t)$ is constant
- Var(y_t) is constant
- Cov$(y_{t=k}, y_{t=k+h})$ depends only on h and not k.

Exogeneity

Strict exogeneity:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t,$$
$$E(\varepsilon_t | \boldsymbol{X}) = 0 \implies E(\varepsilon_t | x_{t+h}) = 0 \quad \forall h$$

If exogeneity is violated, results will be biased. A weaker version (contemporaneous exogeneity) will suffice if T is large.

$$E(\varepsilon_t | x_t) = 0$$

In the ADL model, we assume:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$$

- The α_0 assumes that the series has a non-zero mean.
- In the *static process* the long-run equilibrium is $\beta_0 = \frac{\alpha_0}{1-\alpha_1}$

We will tend to focus on the lagged dependent variable (LDV) model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \varepsilon_t$$

- Short-term (immediate) effect of x_t is β_1.
- Long-term effect of x_t is $\frac{\beta_1}{1-\alpha_1}$

Single vs Multiple Series

Mark's book is really about single-series TS models.

- These things get much more complicated in TSCS data becuse there are, in effect, many different time-series that we're trying to model simultaneously.
- Further, the parameters are often constrained to be the same across the different series, which may or may not be a good idea.

Unit Effects

As mentioned above, the nature of the data produces T observations for each of the N observations.

- Modeling unit effects refers to acknowledging that there may be differences in the average level of the dependent variable across groups.

$$y_i = \alpha_{j[i]} +_i +\varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma_y^2)$$

- After controlling for x, there may remain unexplained variation relating to groups in y. That is what $\alpha_{j[i]}$ is meant to capture.
- Failing to account for $\alpha_{j[i]}$ will usually result in biased estimates of β.

Two Approaches

- Fixed Effects:

$$y_i = \sum_{j=1}^{N} \alpha_j z_{j[i]} + \beta x_i + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma_y^2)$$

- Random Effects

$$y_i = \alpha_{j[i]} + \beta x_i + \varepsilon_i; \quad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2); \quad \varepsilon_i \sim N(0, \sigma_y^2)$$

Within vs Between Effects

Fixed effects is a *within* estimator.

- The interpretation of β_{FE} is that within each observation, a one-unit increase in x corresponds with a β_{FE} unit increase in y over time.
- Since FE is a within estimator, all variables included in the model must vary within unit.

Between effects tell us about the average difference in y by groups, disregarding the variation within units.

- Random effects are *partially pooled* models that can have aspects of both within and between effects.
- RE requires that there is no correlation between x and the unit effect α_j.

Advice: Clark and Linzer

- In the standard case (non-sluggish variables) - both RE and FE do equally well as regards bias, use whichever one you want.
    - Your hypothesis should be the guide - if you have a *within* hypothesis, you should use a *within* estimator.
- With "sluggish" independent variables,
    - RE is better with few observations overall and few observations per unit.
    - FE is better with larger N when the correlation between unit effects and independent variable are moderate to high.
    - FE is better in all cases when the number of observations per unit (times in our case) is bigger than 20.

The plm Package

The `plm` package in R has a number of functions that will help us out here. We'll load in some data to use below:

```
library(foreign)
library(plm)
dat <- read.dta("https://quantoid.net/files/9591/repress_data.dta")
```

Fixed Effects in the plm package

```r
mod1.fe <- plm(repress ~ voice*veto + log(pop) + log(rgdpch) + iwar + cwar, data=dat,
    model = "within", index= c("ccode", "year"), effect="individual")
summary(mod1.fe)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = repress ~ voice * veto + log(pop) + log(rgdpch) +
##     iwar + cwar, data = dat, effect = "individual", model = "within",
##     index = c("ccode", "year"))
##
## Unbalanced Panel: n = 159, T = 6-28, N = 3876
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -4.8942061 -0.6903648  0.0022614  0.7214781  4.4375575
##
## Coefficients:
##               Estimate Std. Error t-value  Pr(>|t|)
## voice         0.063040   0.014638  4.3065 1.702e-05 ***
## veto          0.233100   0.067497  3.4535 0.0005595 ***
## log(pop)     -1.430969   0.146190 -9.7884 < 2.2e-16 ***
## log(rgdpch)  -0.125093   0.089594 -1.3962 0.1627308
## iwar         -0.205658   0.148735 -1.3827 0.1668359
## cwar         -0.633611   0.106042 -5.9751 2.516e-09 ***
## voice:veto    0.013440   0.012657  1.0619 0.2883504
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    5062.1
## Residual Sum of Squares: 4797.3
## R-Squared:      0.052307
## Adj. R-Squared: 0.010159
## F-statistic: 29.2526 on 7 and 3710 DF, p-value: < 2.22e-16
```

Random Effects in the plm package

```r
mod1.re <- plm(repress ~ voice*veto + log(pop) + log(rgdpch) + iwar + cwar, data=dat,
    model = "random", index= c("ccode", "year"), effect="individual")
summary(mod1.re)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = repress ~ voice * veto + log(pop) + log(rgdpch) +
##     iwar + cwar, data = dat, effect = "individual", model = "random",
##     index = c("ccode", "year"))
##
## Unbalanced Panel: n = 159, T = 6-28, N = 3876
##
## Effects:
##                  var std.dev share
## idiosyncratic 1.2931  1.1371 0.579
## individual    0.9409  0.9700 0.421
## theta:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.5683  0.7837  0.7837  0.7718  0.7837  0.7837
##
## Residuals:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -4.5803 -0.7387  0.0145  0.0037  0.7474  4.3187
##
## Coefficients:
##              Estimate Std. Error  t-value  Pr(>|t|)
## (Intercept)  3.822831   0.636985   6.0014 2.135e-09 ***
## voice        0.069137   0.014104   4.9020 9.875e-07 ***
## veto         0.203971   0.066453   3.0694 0.0021598 **
## log(pop)    -0.607796   0.048474 -12.5385 < 2.2e-16 ***
## log(rgdpch)  0.191204   0.053444   3.5777 0.0003509 ***
## iwar        -0.150622   0.149752  -1.0058 0.3145696
## cwar        -0.774438   0.106113  -7.2983 3.520e-13 ***
## voice:veto   0.057605   0.011702   4.9226 8.898e-07 ***
```

Test for Serial Correlation

Fixed Effects:

```
pbgtest(mod1.fe, order=1)
```

```
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel
##  models
##
## data:  repress ~ voice * veto + log(pop) + log(rgdpch) + iwar + cwar
## chisq = 731.08, df = 1, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

```
pwartest(mod1.fe)
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:  mod1.fe
## F = 340.26, df1 = 1, df2 = 3715, p-value < 2.2e-16
## alternative hypothesis: serial correlation
```

Random Effects

```
pbgtest(mod1.re, order=1)
```

```
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel
##  models
##
## data:  repress ~ voice * veto + log(pop) + log(rgdpch) + iwar + cwar
## chisq = 852.02, df = 1, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

Including a Lag of y, (FE)

```
mod2.fe <- plm(repress ~ lag(repress, 1) +  voice*veto + log(pop) + log(rgdpch) + iwar + cwar,
    data=dat, model = "within", index= c("ccode", "year"), effect="individual")
summary(mod2.fe)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = repress ~ lag(repress, 1) + voice * veto + log(pop) +
##     log(rgdpch) + iwar + cwar, data = dat, effect = "individual",
##     model = "within", index = c("ccode", "year"))
##
## Unbalanced Panel: n = 159, T = 5-27, N = 3712
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -4.109986 -0.584769  0.012532  0.589705  3.923199
##
## Coefficients:
##                   Estimate Std. Error t-value  Pr(>|t|)
## lag(repress, 1)  0.4563965  0.0145378 31.3937 < 2.2e-16 ***
## voice            0.0325649  0.0132895  2.4504 0.0143167 *
## veto             0.1305343  0.0608242  2.1461 0.0319333 *
## log(pop)        -0.7825166  0.1371398 -5.7060 1.252e-08 ***
## log(rgdpch)     -0.1098530  0.0820230 -1.3393 0.1805605
## iwar            -0.0075908  0.1332022 -0.0570 0.9545587
## cwar            -0.3415915  0.0936629 -3.6470 0.0002691 ***
## voice:veto       0.0027028  0.0113383  0.2384 0.8116001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    4689.6
## Residual Sum of Squares: 3484.4
## R-Squared:      0.25698
## Adj. R-Squared: 0.22219
## F-statistic: 153.262 on 8 and 3545 DF, p-value: < 2.22e-16
```

Including a Lag of y, (RE)

```r
mod2.re <- plm(repress ~ lag(repress, 1) +  voice*veto + log(pop) + log(rgdpch) + iwar + cwar,
    data=dat, model = "random", index= c("ccode", "year"), effect="individual")
summary(mod2.re)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = repress ~ lag(repress, 1) + voice * veto + log(pop) +
##     log(rgdpch) + iwar + cwar, data = dat, effect = "individual",
##     model = "random", index = c("ccode", "year"))
##
## Unbalanced Panel: n = 159, T = 5-27, N = 3712
##
## Effects:
##                  var std.dev share
## idiosyncratic 0.9829  0.9914     1
## individual    0.0000  0.0000     0
## theta:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       0       0       0       0       0
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -3.846762 -0.673503  0.042165  0.593528  4.428915
##
## Coefficients:
##                   Estimate Std. Error  t-value  Pr(>|t|)
## (Intercept)      0.4037890  0.1830514   2.2059 0.0274538 *
## lag(repress, 1)  0.7021257  0.0113094  62.0835 < 2.2e-16 ***
## voice            0.0368393  0.0097560   3.7761 0.0001618 ***
## veto             0.0696634  0.0484144   1.4389 0.1502640
## log(pop)        -0.1452212  0.0127959 -11.3491 < 2.2e-16 ***
## log(rgdpch)      0.0956760  0.0173440   5.5164 3.698e-08 ***
## iwar            -0.1057841  0.1327086  -0.7971 0.4254348
## cwar            -0.2610067  0.0908406  -2.8732 0.0040860 **
```

Test for Serial Correlation

Fixed Effects:

```
pbgtest(mod2.fe, order=1)
```

```
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel
##  models
##
## data:  repress ~ lag(repress, 1) + voice * veto + log(pop) + log(rgdpch) +     iwar + cwar
## chisq = 65.318, df = 1, p-value = 6.375e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

```
pwartest(mod2.fe)
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:  mod2.fe
## F = 2.527, df1 = 1, df2 = 3551, p-value = 0.112
## alternative hypothesis: serial correlation
```

Random Effects

```
pbgtest(mod2.re, order=1)
```

```
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel
##  models
##
## data:  repress ~ lag(repress, 1) + voice * veto + log(pop) + log(rgdpch) +     iwar + cwar
## chisq = 216.03, df = 1, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

Year Fixed Effects

Often times, work will also include year fixed effects, too, in an effort to take out the effect of common time-trends. You can do this in `plm` as follows:

```
mod2.fe <- plm(repress ~ lag(repress, 1) +  voice*veto + log(pop) + log(rgdpch) + iwar + cwar,
    data=dat, model = "within", index= c("ccode", "year"), effect="twoway")
summary(mod2.fe)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = repress ~ lag(repress, 1) + voice * veto + log(pop) +
##     log(rgdpch) + iwar + cwar, data = dat, effect = "twoway",
##     model = "within", index = c("ccode", "year"))
##
## Unbalanced Panel: n = 159, T = 5-27, N = 3712
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -4.1482557 -0.5569854  0.0081578  0.5817876  3.9282476
##
## Coefficients:
##                  Estimate Std. Error t-value  Pr(>|t|)
## lag(repress, 1)  0.4549601  0.0146216 31.1157 < 2.2e-16 ***
## voice            0.0405794  0.0133393  3.0421 0.0023667 **
## veto             0.1510985  0.0613339  2.4635 0.0138048 *
## log(pop)        -0.2189679  0.2139687 -1.0234 0.3062060
## log(rgdpch)      0.0431622  0.1011534  0.4267 0.6696235
## iwar            -0.0490499  0.1340486 -0.3659 0.7144534
## cwar            -0.3338612  0.0945134 -3.5324 0.0004171 ***
## voice:veto       0.0067431  0.0112837  0.5976 0.5501468
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    4405.5
```

Fixing Problems: Serial Correlation and Contemporaneous Correlation

Below will give you the Driscoll-Kraay robust standard errors.

```
library(lmtest)
coeftest(mod2.fe, vcov.=vcovSCC)
```

```
##
## t test of coefficients:
##
##                   Estimate Std. Error t value  Pr(>|t|)
## lag(repress, 1)  0.4549601  0.0403850 11.2656 < 2.2e-16 ***
## voice            0.0405794  0.0106001  3.8282 0.0001313 ***
## veto             0.1510985  0.0504078  2.9975 0.0027409 **
## log(pop)        -0.2189679  0.1915983 -1.1428 0.2531791
## log(rgdpch)      0.0431622  0.1111788  0.3882 0.6978744
## iwar            -0.0490499  0.0938329 -0.5227 0.6011904
## cwar            -0.3338612  0.0741526 -4.5024 6.938e-06 ***
## voice:veto       0.0067431  0.0186373  0.3618 0.7175172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

These have only been shown to work with fixed effects.

Fixing Problems: Serial Correlation and Heteroskedasticity

```r
coeftest(mod2.fe, vcov.=pvcovHC, type="HC3", method="arellano")
```

```
##
## t test of coefficients:
##
##                   Estimate Std. Error t value  Pr(>|t|)
## lag(repress, 1)  0.4549601  0.0256105 17.7646 < 2.2e-16 ***
## voice            0.0405794  0.0173420  2.3399  0.019342 *
## veto             0.1510985  0.0661323  2.2848  0.022384 *
## log(pop)        -0.2189679  0.2800880 -0.7818  0.434395
## log(rgdpch)      0.0431622  0.1257239  0.3433  0.731386
## iwar            -0.0490499  0.1422551 -0.3448  0.730264
## cwar            -0.3338612  0.1077305 -3.0990  0.001957 **
## voice:veto       0.0067431  0.0150327  0.4486  0.653775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```