Regression III
Lecture 12: TSCS Issues

Dave Armstrong

University of Wisconsin – Milwaukee
Department of Political Science

e: armstrod@uwm.edu
w: www.quantoid.net/ICPSR.php

Issues in TSCS Models

There are a number of issues in TSCS models that we haven't spent a lot of time talking about thus far.

- Dynamics
- Heterogeneity

TSCS data are a particular type of multilevel data where years are the level 1 units and group (e.g., countries, states, people, etc...) are the level 2 units

Dynamics

- This isn't a course on (nor do we have time to cover) time-series analysis as a subject of investigation in its own right.
- There are a couple of reasons why this *may not be* as problematic as we might think.
    - Lots of times, with annual aggregate data, we are unable to take full advantage of all that time-series has to offer us. This is not to say that problems don't exist, rather it is to say that with relatively small T, we will be less sure about how to fix these things
    - Some problems that are "easy" to deal with in single-time-series are not nearly as easily dispensed with in TSCS applications (e.g., stationarity)
- We will focus mainly on the lagged outcome variable (LDV) model.
    - As Beck and Katz (2004) suggest, there is generally little reason to prefer an AR(1) model to the LDV model

Serial Correlation

- One problem for inference with time-series data is serial correlation in the errors.
- We can deal with this if we properly model the serial correlation either through an autoregressive parameter (AR) or include a lagged outcome variable on the RHS of the model.
- Testing for serial correlation - you can use a Lagrange Multiplier Test by doing the following:
    1. Run OLS
    2. Compute residuals
    3. Regress residuals on all explanatory variables and the lagged residual
    4. If the coefficient on the lagged residual is significant, we can reject the null of independent errors

Fixed-Effects models in R

- Fixed-effects models are easy to run in **R**
- You can get the results you want either by using the "demeaning" code above or by treating your unit variable as a factor, as you've seen already, **R** will dummy all of those values out
- It is also possible to include a lagged outcome variable if you like. You could use the time-series part of **R** to do this, but it's actually a bit easier to just write a function to do it:

TSCS Lag in **R**

```
> tscslag <- function(dat, x, id, time){
+ obs <- apply(dat[, c(id, time)], 1, paste, collapse=".")
+ tm1 <- dat[[time]] - 1
+ lagobs <- apply(cbind(dat[[id]], tm1), 1, paste, collapse=".")
+ lagx <- dat[match(lagobs, obs), x]
+ }
```


Fixed Effects with "Sluggish" variables

- One of the big problems people have with fixed effects models is that any heterogeneity that is a function of non-time-varying variables (or slowly changing variables) will be captured by the unit-specific intercepts (i.e., the fixed effects)
- Until recently, there has been little work done on how we might remedy this problem, the advice was either accept it, or use random effects
- I've collected a fair amount of anecdotal evidence suggesting that this is the main reason that people choose random effects over fixed effects, regardless of how reasonable the assumptions are
- Plümper and Troeger (2007) suggest an estimator that will essentially permit an estimation of the fixed effects and more importantly, the effect that non-time-varying variables have on the fixed effects

Plümper and Troeger Model

- Let's assume we have the following DGP:

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + c_i + \varepsilon_{it}$$

  where the z_{mi}'s are non-time-varying variables.
- It might go without saying, though I won't let it, that if the c_i's were a perfect linear combination of the z_{mi}'s, then we wouldn't need to estimate the c_i's.
  - That is to say, that if we could account for all of the differences between units with observed variables, we wouldn't have to worry about fixed effects (i.e., we would have no *unobserved* heterogeneity, it would all be observed

Fixed Effects Vector Decomposition (FEVD)

The model is estimated in three stages

- Estimate the standard fixed effects model to obtain estimates of the unit effects:

$$\begin{aligned} \ddot{\mathrm{y}}_{it} &= \sum_{k=1}^{K} \beta_k^{FE} \ddot{\mathbf{x}}_{it} + \ddot{\mathrm{e}}_{it} \\ \hat{c}_i &= \bar{y}_i - \sum_{k=1}^{K} \beta_k^{FE} \bar{x}_{ki} - \bar{e}_i \end{aligned}$$

- Regress the estimated unit effects $\hat{c}_i$ on the non-time-varying and slowly changing variables:

$$\hat{c}_i = \sum_{m=1}^{M} \gamma_m z_{mi} + h_i$$

where h_i is that part of the unit effect that cannot be explained by the $n-t-v$ variables

FEVD (2)

- In the third stage, we add back in the RHS from stage 2:

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + \delta h_i + \varepsilon_{it}$$

- We do this for two reasons:
  1. This provides the right number of degrees of freedom for statistical tests
  2. Also, this model allows for fixing problems with autocorrelation (by adding a lagged DV) and/or heteroskedasticity (by using a robust variance estimator in the form of a White HCCM or PCSE's)

Fixed Effects Redux

- Fixed effects allow arbitrary correlation between the unit effects and the $\boldsymbol{x}_{it}$ which makes them a less restrictive form of estimation
- Through recent developments, the effects of non-time-varying variables can be captured through the fixed-effects vector decomposition
- With TSCS data of any kind, pains must be taken to "fix" any problems in the data such as serial correlation and unmodelled heterogeneity

Random Effects

- Random effects do not model c_i directly, rather they estimate the mean and variance of a distribution for c_i.
- Wooldridge (2002) suggests that if the units are exchangeable (i.e., their names are irrelevant), then random effects often make sense. If the units are not exchangeable, they are interesting in their own right, then random effects are probably not the right model
- The general model takes this form:

$$y_{it} = \boldsymbol{x}_{it} + \nu_{it} \text{ where } \nu_{it} = c_i + u_{it}$$

- In random effects models, we're basically acknowledging the fact that the errors are not *iid*, but we're not interested in estimating the unit effects

Structure of Error Variance

- We are assuming that the variance-covariance matrix of the errors $\boldsymbol{\Omega}$ is not diagonal, rather there is some non-independence in the following form:

$$\boldsymbol{\Omega} = E(\boldsymbol{v}_i\boldsymbol{v}_i') = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & & \vdots \\ \vdots & & \ddots & \sigma_c^2 \\ \sigma_c^2 & \dots & \sigma_c^2 & \sigma_c^2 + \sigma_u^2 \end{pmatrix}$$

- Since c_i is in the error term, this requires the zero correlation between $\boldsymbol{x}_{it}$ and c_i
- What we are gaining from this more restrictive assumption is a lot of degrees of freedom.

Estimating the Random Effects Model

- We can use feasible generalized least squares (FGLS) [not to be confused with GLM] to estimate the random effects model
- The FGLS estimator of β is:

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}_i\right)^{-1}\left(\sum_{i=1}^{N} \boldsymbol{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}_i\right)$$

- This is akin to weighted least squares, but the "weight" matrix is not diagonal
- You could also do a less restrictive FGLS estimation where you are attempting to estimate all of the off-diagonal elements of $\boldsymbol{\Omega}$ to be different, but this takes a *lot* of data.

Hausman Test for FE vs RE

- Hausman proposed a test of the difference between the RE and FE estimates.
- Before we get to the statistic, there are 2 caveats:
  1. Strict exogeneity is assumed
  2. The test assumes that under the null - zero covariance between $\boldsymbol{x}_i$ and c_i, so it is not actually testing this assumption.
- The test looks at the difference between the two coefficient vectors as FE is consistent even if the $\boldsymbol{x}_i$ and c_i are correlated, but RE is more efficient when they are not.

$$H = (\boldsymbol{b} - \boldsymbol{B})' \left(V(\boldsymbol{b}) - V(\boldsymbol{B})\right)^{-1} (\boldsymbol{b} - \boldsymbol{B})$$

where $\boldsymbol{b}$ and $V(\boldsymbol{b})$ come from the FE model, $\boldsymbol{B}$ and $V(\boldsymbol{B})$ are from the RE model, and $H \sim \chi^2$ with k (length of $\boldsymbol{b}$) degrees of freedom

Data and Model

- These data come originally from Evelyne Huber, Charles Ragin and John D. Stephens (1997) and were updated by David Brady, Jason Beckfield and John D. Stephens in 2004.
- The model we're looking at comes from the book "Development and Crisis of the Welfare States: Parties and Policies in Global Markets" by the above mentioned authors
- Data and replication materials can be obtained from: `http://www.unc.edu/~jdsteph/joint_data.html`

Preparing Data

- We'll standardize the continuous data just to make our lives a bit easier. It will put all of the coefficients on the same metric

```
> dat <- read.dta("cwssmall.dta")
> convars <- c("sstran","flabfo","vturn","pctaged","cgdp",
+     "tunemp","ofdi","openc")
> dat.std <- dat
> dat.std[, convars] <- scale(dat.std[, convars])
```

- It will also help us out to have an ID number that is in increasing consecutive integers

```
> unid <- unique(dat[["idn"]])
> unid <- unid[order(unid)]
> dat[["id"]] <- match(dat[["idn"]], unid)
```


Fixed Effects Model

- We can run the FE model two different ways in **R** , either by including dummy variables for the ID number or using the demeaned data

```
> mod.fe1 <- lm(sstran ~ leftcab + cncrcab + flabfo +
+     vturn + pctaged + cgdp + tunemp  + ofdi + openc +
+     as.factor(idn), data=dat)
```



```
Call:
lm(formula = sstran ~ leftcab + cncrcab + flabfo + vturn + pctaged +
    cgdp + tunemp + ofdi + openc + as.factor(idn), data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1493 -0.8958 -0.1392  0.7999  7.1604

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.939e+00  3.946e+00   1.252 0.211761
leftcab         -5.642e-01  3.453e-01  -1.634 0.103440
cncrcab          1.077e+01  1.405e+00   7.663 3.06e-13 ***
flabfo          -2.735e-05  7.228e-05  -0.378 0.705457
vturn           -5.164e-02  3.960e-02  -1.304 0.193331
pctaged          6.474e+01  2.170e+01   2.983 0.003105 **
cgdp             6.332e-05  5.060e-05   1.251 0.211852
tunemp           4.172e-04  2.230e-04   1.871 0.062424 .
ofdi            -1.509e-05  1.510e-05  -0.999 0.318680
openc           -4.751e-04  1.846e-02  -0.026 0.979481
as.factor(idn)2  7.526e+00  1.416e+00   5.314 2.21e-07 ***
as.factor(idn)3  6.809e+00  2.128e+00   3.199 0.001538 **
as.factor(idn)4  1.734e+00  1.149e+00   1.509 0.132346
as.factor(idn)6  5.319e+00  1.378e+00   3.860 0.000141 ***
as.factor(idn)7  6.053e+00  1.597e+00   3.790 0.000185 ***
as.factor(idn)8  5.637e+00  1.662e+00   3.393 0.000794 ***
as.factor(idn)10 4.715e+00  1.270e+00   3.713 0.000247 ***
as.factor(idn)12 1.057e+01  1.715e+00   6.165 2.48e-09 ***
as.factor(idn)14 2.228e+00  1.584e+00   1.407 0.160656
as.factor(idn)15 6.916e+00  1.694e+00   4.082 5.84e-05 ***
as.factor(idn)17 1.354e+00  1.884e+00   0.719 0.473053
as.factor(idn)18 -9.363e-01  3.832e+00  -0.244 0.807130
---
Signif. codes:  0 â***â 0.001 â**â 0.01 â*â 0.05 â.â 0.1 â â 1

Residual standard error: 1.769 on 277 degrees of freedom
Multiple R-squared: 0.8627, Adjusted R-squared: 0.8523
F-statistic: 82.88 on 21 and 277 DF,  p-value: < 2.2e-16
```


The Random Effects Model

- Remember, the random effects model can be computed with FGLS (feasible generalized least squares)
- The key to getting this "right" in **R** is picking the right correlation structure.
- To get the model that the slides have been talking about you need a "compound symmetric correlation matrix" which estimates the same off-diagonal correlation of the errors $\hat{\sigma}_c$
- We can specify the model in **R** as follows:

```
> library(nlme)
> mod.re <- gls(sstran ~ leftcab + cncrcab +
+     flabfo + vturn + pctaged + cgdp +
+     tunemp  + ofdi + openc, data=dat.std,
+     correlation=corCompSymm(form=~1|idn))
```

- These are linear model coefficients and can be interpreted as linear model coefficients

```
Generalized least squares fit by REML
  Model: sstran ~ leftcab + cncrcab + flabfo + vturn + pctaged + cgdp +      tunemp + ofdi + openc
  Data: dat.std
       AIC      BIC    logLik
  374.2526 418.2497 -175.1263

Correlation Structure: Compound symmetry
 Formula: ~1 | idn
 Parameter estimate(s):
      Rho
0.7102305

Coefficients:
                 Value  Std.Error   t-value p-value
(Intercept) -0.0439012 0.17167055 -0.255729  0.7983
leftcab     -0.1281323 0.07465114 -1.716415  0.0872
cncrcab      2.4435951 0.29948875  8.159222  0.0000
flabfo      -0.2254754 0.15982370 -1.410776  0.1594
vturn       -0.1109588 0.08899634 -1.246779  0.2135
pctaged      0.3410179 0.08819867  3.866475  0.0001
cgdp         0.0573980 0.05314192  1.080089  0.2810
tunemp       0.1698466 0.09888141  1.717680  0.0869
ofdi        -0.0236916 0.04688603 -0.505303  0.6137
openc        0.1011247 0.09938204  1.017535  0.3097

 Correlation:
        (Intr) leftcb cncrcb flabfo vturn  pctagd cgdp   tunemp ofdi
leftcab -0.177
cncrcab -0.103  0.230
flabfo   0.005 -0.016 -0.040
vturn   -0.014  0.113 -0.057  0.029
pctaged  0.026 -0.145 -0.032  0.103 -0.252
cgdp    -0.013  0.082 -0.011 -0.091  0.387 -0.747
tunemp  -0.028  0.167  0.011 -0.378  0.229 -0.020 -0.146
ofdi    -0.012  0.039  0.086 -0.562  0.024  0.097 -0.354  0.305
openc    0.012 -0.050 -0.061  0.218 -0.055  0.036 -0.254  0.081 -0.056

Standardized residuals:
        Min          Q1         Med          Q3         Max
-2.26407449 -0.76062469 -0.04164017  0.59459603  3.33051108

Residual standard error: 0.7150392
Degrees of freedom: 299 total; 289 residual
```


FEVD

- If the fixed-effects porridge is too cold and the random-effects porridge is too hot, the FEVD might be just right
- Remember, this is a three stage procedure
- The authors of the original work we are replicating proposed that a number of institutional variables, two of which we'll investigate: federalism and legacy of authoritarianism

```
> fix.eff <- c(0, coef(mod.fe1)[11:22])
> stage2.mod <- lm(fix.eff[dat[["id"]]] ~ fed + authleg, data=dat)
> s2.resid <- residuals(stage2.mod)
> mod.fevd <- lm(sstran ~ leftcab + cncrcab + flabfo +
+     vturn + pctaged + cgdp + tunemp  + ofdi +
+     openc + authleg + fed + s2.resid, data=dat.std)
```



```
Call:
lm(formula = sstran ~ leftcab + cncrcab + flabfo + vturn + pctaged +
    cgdp + tunemp + ofdi + openc + authleg + fed + s2.resid,
    data = dat.std)

Residuals:
     Min       1Q   Median       3Q      Max
-1.57834 -0.19873 -0.02532  0.17762  1.53662

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.7326858  0.0840735  -8.715 2.40e-16 ***
leftcab     -0.1251934  0.0657498  -1.904 0.057902 .
cncrcab      2.3008162  0.2087005  11.024  < 2e-16 ***
flabfo      -0.0517521  0.0905474  -0.572 0.568078
vturn       -0.1249171  0.0371816  -3.360 0.000886 ***
pctaged      0.2834127  0.0355347   7.976 3.67e-14 ***
cgdp         0.0846646  0.0305876   2.768 0.006009 **
tunemp       0.1768563  0.0799500   2.212 0.027751 *
ofdi        -0.0561203  0.0378910  -1.481 0.139681
openc        0.0006613  0.0341459   0.019 0.984561
authleg      0.4599504  0.0384943  11.949  < 2e-16 ***
fed         -0.2951111  0.0374200  -7.886 6.61e-14 ***
s2.resid     0.2154779  0.0135158  15.943  < 2e-16 ***
---
Signif. codes:  0 â***â 0.001 â**â 0.01 â*â 0.05 â.â 0.1 â â 1

Residual standard error: 0.3777 on 286 degrees of freedom
Multiple R-squared: 0.8631, Adjusted R-squared: 0.8574
F-statistic: 150.3 on 12 and 286 DF,  p-value: < 2.2e-16
```


Hausman Test for Random vs. Fixed Effects

- We can compute the Hausman test we talked about earlier for these models:

```
> b <- mod.fe1$coef[2:10]
> B <- mod.re$coef[2:10]
> vb <- vcov(mod.fe1)[2:10,2:10]
> vB <- vcov(mod.re)[2:10,2:10]
> chi2 <- t(b-B) %*% solve(vb-vB) %*% (b-B)
> 1-pchisq(chi2, length(b))

             [,1]
[1,] 2.346427e-05
```

What would I do?

- This is a nice example because the instructions are not as straightforward here
- There is no empirical evidence that we're getting substantially different results from the FE model versus the RE model
    - This would argue for the RE model because it is more efficient (estimates fewer parameters)
- However, as Wooldridge suggested, if the units are not "exchangeable", this may well make random effects theoretically unappealing, if not empirically so.
- The FEVD shows that there are some interesting things going on with respect to the non-time-varying variables, so if it is our goal to say things about those substantively important variables, I would probably do FEVD
- The outcome of the Hausman test here is basically that there is little difference in the results from both the FE and RE models