# 17.874 Lecture Notes
# Part 6: Panel Models

# 6. Panel Models

Panel designs are used to capture two sorts of research design ideas.

First, the literature on quasi-experimentation defines a quasi- or natural experiment as an intervention that occurs in one time and changes the mean or path of the dependent variable. Interventions may be all manner of changes in an independent variable of interests. It is perhaps most natural to think of them as controlled application of some factor, such as occurs with field experiments. But, interventions may also be important events that the researcher conjectures affect behavior exogenously. The nature of the timing of the intervention gives some leverage over the question of causality. Because the intervention precedes the change in behavior, the intervention plausibly causes the outcome.

Second, an important approach to eliminating omitted variables is to collect data for units such that the plausible omitted variables are constant or nearly constant across units. This is sometimes referred to as *heterogeneity*. For example, demographics and civic culture are thought to change very slowly, especially compared to short term fluctuations in the economy or changes in the political climate.

Consider the problem of estimating the effect of imprisonment on crime. The simple correlation between the per capita prison population and property crimes per capita is positive. Controlling for demographic and short-run economic factors the relationship between imprisonment and property crime remains positive. What sorts of interventions might you imagine? How might you set up a study to control for the other potential unmeasured factors, such as culture?

Another example comes from election statistics. In US elections there are, typically, two unmeasured factors the normal party division of the vote and the differential qualities of the competiting candidates. We do not directly observe the normal party division of the vote in elections. Rather, we typically use proxy variables, most commonly the vote in the Presidential or Governor election in the districts of interest. One might consider the normal vote

as unmeasured heterogeneity in the electorate across geographic areas. Candidate quality is also thought of as a sort of unmeasured or unmeasureable factor. Researchers use proxy variables such as prior office holding experience and indications of whether a candidate is in a scandal.

In "Decomposing Sources of the Incumbency Advantage in U.S. House Elections," Steve Levitt and Catherine Levine use elections in which the same incumbent and challenger face each other to hold constant the differential quality of the candidates and the normal vote. In "Old Voters, New Voters, and the Office Holder Advantage", Jim Snyder, Charles Stewart, and I study county level election returns within congressional districts following redistricting to measure the differential between areas where an incumbent is known and areas where he or she is new, controlling for the characteristics of the incumbent and the challenger. In "The Incumbency Advantage In US Elections", Jim Snyder and I pool all statewide elections in a given year (or decade) to capture the normal vote, without using proxies.

Both of these design ideas – quasi-experimentation and heterogeneity – involve using multiple measurements of the same units to untangle the effects of the variables of interest and other factors.

### 6.1. Notation

We will need to keep straight two different sorts of indexes. Let $i$ denote individuals or units (such as countries). Let $t$ index the repeated observation of those units, typically time. Let $\mathbf{x}'_{\mathrm{it}}$ be a vector of K regressors for observation $i$ at time $t$. Let $\mathbf{z}'_{\mathrm{i}}$ be a set of factors that represent the heterogeneity across units. We may express $y_{it}$ in terms of the variables $\mathbf{x}_{\mathrm{it}}$, and the heterogeneous effects $\alpha_i$s and $\lambda_t$s:

$$y_{it} = \mathbf{x}_{\mathrm{i}}\mathbf{t}'\beta + \mathbf{z}'_{\mathrm{i}}\alpha + \lambda_{\mathrm{t}} + \epsilon_{\mathrm{it}}.$$

Because $z_i$ is just an indicator of the unit, it is more commonly written $\alpha_i$. In other words, each unit has its own intercept. There is no constant in this model. One might also include

a constant and measure the heterogeneity as deviations around the grand mean:

$$y_{it} = \mathbf{x_i t}'\beta + \alpha + \mathbf{d_i} + \lambda_\mathrm{t} + \epsilon_\mathrm{it}.$$

Two sorts of effects might be thought to occur: fixed-effects and Random-effects. Fixed-effects: $d_i$ is correlated with $x_{it}$. Random-effects: $d_i$ is uncorrelated with $x_{it}$.

Ecological Regression and Aggregation as examples.

$$P(Y=1) = P(Y=1|X=1)P(X=1) + P(Y=1|X=0)P(X=0)$$

$$y_{it} = \alpha_{it}X_{it} + \beta_{it}(1 - X_{it})$$

$$y_{i.} = \alpha_i x_{i.} + \beta_i(1 - x_{i.})$$

Regress $y_i$ on $x_i$, assuming "random coefficient" that is unrelated to $x_i$. This will yield unbiased estimates of $\alpha$ and $\beta$. (Goodman 1957).

6.2. Fixed-Effects Regression

Differencing

Define the difference operator as the differential between successive values of $t$:

$$\Delta y_{it} = y_{i,t} - y_{i,t-1}$$

Differencing removes the unknown parameter $\alpha_i$:

$$\Delta y_{it} = \Delta \mathbf{x_{it}}' + \mathbf{\Delta}\epsilon_\mathrm{it}$$

Regressing $\Delta y_{it}$ on $\Delta \mathbf{x_{it}}'$ yields $b_k = \sum_i \sum_t \Delta x_{kit} \Delta y_{it} \sum_i \sum_t \Delta x_{kit}^2$. $E[b_k] = \beta_k$, so long as $\Delta x_{kit}$ is uncorrelated with $\Delta \epsilon_{it}$.

Note: differencing subtracts out the means of the variables.

The benefit of this approach is that it mimics the quasi-experiment concept and the approximate interpretation of a regression as the effect of a change in $X$ on a change in

$Y$. One drawback is that it does not allow us to estimate or observe the $\alpha_i$s, which are of practical interest and are statistically important for choice of Fixed or Random Effects. If the effects are unrelated to $\epsilon_i$, we can find a more efficient estimator. This estimator, however, is guaranteed to be unbiased.

Example: Crime Data.

Least Squares Dummy Variables

An alternative way to set up the problem uses dummy variables. Stack the data as follows.

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ . \\ . \\ . \\ y_{1T} \\ y_{21} \\ . \\ . \\ . \\ y_{2T} \\ . \\ . \\ . \\ y_{nT} \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} x_{1,11}, x_{2,11}, ..., x_{k,11} \\ x_{1,12}, x_{2,12}, ..., x_{k,12} \\ . \\ . \\ . \\ x_{1,1T}, x_{2,1T}, ..., x_{k,1T} \\ x_{1,21}, x_{2,21}, ..., x_{k,21} \\ . \\ . \\ . \\ x_{1,2T}, x_{2,2T}, ..., x_{k,2T} \\ . \\ . \\ . \\ x_{1,nT}, x_{2,nT}, ..., x_{k,nT} \end{pmatrix}$$

Define a set of dummy variables that identify each group as:

$$\mathbf{D} = (\, D_1, D_2, ...D_n \,) = \begin{pmatrix} 1,0,0,0,...,0 \\ 1,0,0,0,...,0 \\ . \\ . \\ 1,0,0,0,...,0 \\ 0,1,0,0,...,0 \\ . \\ . \\ 0,1,0,0,...,0 \\ 0,0,1,0,...,0 \\ . \\ . \\ 0,0,1,0,...,0 \\ . \\ . \\ 0,0,0,0,...,1 \\ . \\ . \\ 0,0,0,0,...,1 \end{pmatrix}$$

We may set up the problem as regression model with $n$ dummy variables estimated from $nT$ observations:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{D}\alpha + \epsilon$$

Least squares regression estimates the coefficient vector $(\beta', \alpha')$. This estimation approach is known as Least Squares Dummy Variables.

It is instructive to consider the estimator of just the vector $\beta$, partialling out the fixed effects. Define $M_0 = I - \frac{1}{T}\mathbf{i}\mathbf{i}'$. Premultiplying $M_0$ times any variable deviates all variables from their means. That is $M_0\mathbf{y} = (\mathbf{y}_1 - \bar{\mathbf{y}}, \mathbf{y}_2 - \bar{\mathbf{y}}...)'$. For each variable, Least Squares Dummy Variables deviates the observations within each unit from the unit mean.

$$\mathbf{M_D} = \begin{pmatrix} M_0,0,0,0,0,...,0 \\ 0,M_0,0,0,0,...,0 \\ 0,0,M_0,0,0,...,0 \\ 0,0,0,M_0,0,...,0 \\ . \\ . \\ . \\ 0,0,0,0,0,...,M_0 \end{pmatrix}$$

Create the transformed data $\mathbf{y}_* = \mathbf{M_D y}$ and $\mathbf{X}_* = \mathbf{M_D X}$. The LSDV estimator of $\beta$ is

$$b_{LSDV} = (\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*'\mathbf{y}_* = (\mathbf{X}'\mathbf{M}_\mathbf{D}'\mathbf{M_D X})^{-1}\mathbf{X}'\mathbf{M}_\mathbf{D}'\mathbf{M_D y} = (\mathbf{X}'\mathbf{M_D X})^{-1}\mathbf{X}'\mathbf{M_D y}.$$

The last equality holds because $M_0$ is symmetric and idempotent. When multiplied times its transpose it returns itself.

$$\begin{aligned} E[b_{LSDV}] &= E[(\mathbf{X}'\mathbf{M_D X})^{-1}\mathbf{X}'\mathbf{M_D y}] = \mathbf{E}[(\mathbf{X}'\mathbf{M_D X})^{-1}\mathbf{X}'\mathbf{M_D}(\mathbf{X}\beta + \mathbf{D}\alpha + \epsilon)] \quad (1) \\ &= E[\beta + \mathbf{0} + (\mathbf{X}'\mathbf{M_D X})^{-1}\mathbf{X}'\mathbf{M_D}\epsilon] = \beta. \quad (2) \end{aligned}$$

The regression error variance is:

$$\sigma_\epsilon^2 = \frac{\sum_{i=1}^n e_i^2}{nT - n - K}$$

We may use these results to derive the variance of $\mathbf{b}$, which we will compare with other estimators:

$$\begin{aligned} V(b_{LSDV}) &= E[((\mathbf{X}'\mathbf{M_D X})^{-1}\mathbf{X}'\mathbf{M_D}\epsilon)((\mathbf{X}'\mathbf{M_D X})^{-1}\mathbf{X}'\mathbf{M_D}\epsilon))'] \quad (3) \\ &= E[(\mathbf{X}'\mathbf{M_D X})^{-1}\mathbf{X}'\mathbf{M_D}\epsilon)(\epsilon'\mathbf{M_D X}(\mathbf{X}'\mathbf{M_D X})^{-1})] \quad (4) \\ &= \sigma_\epsilon^2(\mathbf{X}'\mathbf{M_D X})^{-1} = \sigma_\epsilon^2 \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{t}=1}^{\mathrm{T}} (\mathbf{x_{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})(\mathbf{x_{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})' \quad (5) \end{aligned}$$

6.3. Random-Effects and GLS

When $\alpha_i$ is uncorrelated with $X$ or when the between and within regressions are approximately the same, we can make more efficient use of the data with a random effects model. Rather that estimate a simple mean for each, assume that each unit is drawn from a distribution, independently of $X$. We have then two components in the error term:

$$y_{it} = \mathbf{x_{it}}'\beta + \mathbf{u_{it}},$$

$$u_{it} = \alpha_i + \epsilon_{it},$$

where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and $\alpha_i \sim N(0, \sigma_\alpha^2)$

The error structure of this model becomes

$$\mathbf{\Sigma} = \mathbf{E}[(\alpha+\epsilon)(\alpha+\epsilon)'] = \begin{pmatrix} \sigma_\alpha^2+\sigma_\epsilon^2, & \sigma_\alpha^2, & \sigma_\alpha^2, & ..., & \sigma_\alpha^2 \\ \sigma_\alpha^2, & \sigma_\alpha^2+\sigma_\epsilon^2, & \sigma_\alpha^2, & ..., & \sigma_\alpha^2 \\ \sigma_\alpha^2, & \sigma_\alpha^2, & \sigma_\alpha^2+\sigma_\epsilon^2, & ..., & \sigma_\alpha^2 \\ & & . & & \\ & & . & & \\ & & . & & \\ \sigma_\alpha^2, & \sigma_\alpha^2, & \sigma_\alpha^2, & ..., & \sigma_\alpha^2+\sigma_\epsilon^2 \end{pmatrix} = \sigma_\alpha^2 \mathbf{i}\mathbf{i}' + \sigma_\epsilon^2 \mathbf{I}.$$

A GLS estimator is to transform the variables with the following:

$$\mathbf{\Sigma}^{1/2} = \frac{\mathbf{1}}{\sigma_\epsilon}[\mathbf{I} - \frac{\theta}{\mathbf{T}}\mathbf{i}\mathbf{i}'],$$

where

$$\theta = 1 - \frac{\sigma_\epsilon}{\sqrt{\sigma_\epsilon^2 + T\sigma_\alpha^2}}$$

Direct estimation of $\sigma_\epsilon^2$ and $\sigma_\alpha^2$ from the overall regression is not obvious because each error is a draw from both distributions. The total regression yields an error variance that, in expectation, equals $\sigma_u^2 = \sigma_\alpha^2 + \sigma_\epsilon^2$. We can estimate the relevant variance components as follows:

1. Estimate $\beta$ using LSDV. This provides an unbiased estimate of $\beta$

2. Estimate $\sigma_\epsilon^2$ using the average within error variance from the LSDV.

$$\bar{s}_e^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T (e_{it} - e_{i.})^2}{nT - n - K - 1}$$

3. Estimate $\sigma_u^2$ from the total regression. An unbiased estimate of $\sigma_\alpha^2 = \hat{\sigma}_u^2 - \bar{s}_e^2$.

We use these estimates to construct: $\hat{\mathbf{\Sigma}}^{1/2} = \frac{1}{\hat{\sigma}_\epsilon}[I - \frac{\hat{\theta}}{T}\mathbf{i}\mathbf{i}']$. Premultiplying the data by $\hat{\mathbf{\Sigma}}^{1/2}$ provides a consistent, feasible GLS estimate. This is the random effects model estimated in your computer program.

When is a random effects specification more appropriate than least squares or fixed-effects? Several tests are possible. The Lagrange multiplier test considers whether the

variance of the $\alpha$'s is zero. Likelihood ratio tests can also be used to test whether this variance is zero. A non-zero variance is sufficient for a random effects model, and establishes that random effects should be used in stead of least squares. When you have panel data that is nearly always the case.

A more important hypothesis is whether we should use fixed or random effects. Hausman developed a straightforward test based on the Wald test. The assumed hypothesis in a random effects model holds that the regressors are independent of the $\alpha$'s. Under this assumption both LSDV and GLS are unbiased; however, GLS is more efficient. If the assumption is false LSDV is unbiased. Hence, we may compare the bias due to the GLS estimator $\hat{\beta}$ against the efficiency loss with LSDV estimator $\mathbf{b}$.

The Wald statistic is the squared difference between the inefficient but unbiased estimator and the efficient but potentially biased estimator, divided by the variance of the difference. The difference bewteen the parameter vectors $[\mathbf{b} - \hat{\beta}]$. The Variance of the difference is $V[\mathbf{b} - \hat{\beta}]$. Hausman showed that the Variance of the difference equals the difference of the variance!.

$$V[\mathbf{b} - \hat{\beta}] = \mathbf{V}[\mathbf{b}] + \mathbf{V}[\beta] - \mathbf{Cov}[\mathbf{b}, \hat{\beta}] - \mathbf{Cov}[\hat{\beta}, \mathbf{b}]$$

Hausman shows that $Cov[\mathbf{b} - \hat{\beta}, \hat{\beta}] = \mathbf{0}$. Hence,

$$Cov[\mathbf{b} - \hat{\beta}, \hat{\beta}] = \mathbf{Cov}[\mathbf{b}, \hat{\beta}] - \mathbf{V}[\hat{\beta}] = \mathbf{0}$$

So,

$$Cov[\mathbf{b}, \hat{\beta}] = \mathbf{V}[\hat{\beta}]$$

This implies that

$$V[\mathbf{b} - \hat{\beta}] = \mathbf{V}[\mathbf{b}] - \mathbf{V}[\beta]$$

and that the Wald test is

$$W = (\mathbf{b} - \hat{\beta})'[\mathbf{V}[\mathbf{b}] - \mathbf{V}[\beta]]^{-1}(\mathbf{b} - \hat{\beta})$$

One problem with this test is that the Variance-Covariance matrix is not guaranteed to

be positive-definite. While the result holds in expectation and as $n$ becomes very large, it can yield negative variances in finite samples. This is the case with Levitt's crime data.

6.4. Comparison of OLS, Fixed-Effects, and Random-Effects

Another way to understand the issues involved here is in terms of the analysis of variance and covariance. There are different variances and covariances to model.

Ultimately, we wish to explain the total variance of a variable.

$$S_{xx}^T = \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{\mathrm{i\,t}} - \bar{\mathbf{x}}_{\mathrm{..}})(\mathbf{x}_{\mathrm{i\,t}} - \bar{\mathbf{x}}_{\mathrm{..}})'$$

Within each group the variances and covariances are:

$$S_{xx}^W = \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{\mathrm{i\,t}} - \bar{\mathbf{x}}_{\mathrm{i.}})(\mathbf{x}_{\mathrm{i\,t}} - \bar{\mathbf{x}}_{\mathrm{i.}})'$$

$$S_{xy}^W = \sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \bar{y}_{i.})(\mathbf{x}_{\mathrm{i\,t}} - \bar{\mathbf{x}}_{\mathrm{i.}})'$$

These are the averages of the variances within each group.

Between groups the variances are

$$S_{xx}^B = \sum_{i=1}^{n} T(\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})(\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})'$$

$$S_{xy}^B = \sum_{i=1}^{n} T(\bar{y}_{i.} - \bar{y}_{..})(\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})'$$

The Total Variance and Covariance can be expressed as the sum of the Within and Between components.

$$S_{xx}^T = \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{..}})(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{..}})' = \sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{t=1}}^{\mathrm{T}}(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}} + \bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}} + \bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})'$$

$$= \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})' + (\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})' + (\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})(\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})' + (\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})'$$

$$= \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})' + \sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{t=1}}^{\mathrm{T}}(\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})' + \sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{t=1}}^{\mathrm{T}}(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})(\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})' + \sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{T}(\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})'$$

$$= \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})(\mathbf{x}_{\mathrm{it}} - \bar{\mathbf{x}}_{\mathrm{i.}})' + \sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{T}(\bar{\mathbf{x}}_{\mathrm{i.}} - \bar{\mathbf{x}}_{\mathrm{..}})' = \mathbf{S}_{\mathrm{xx}}^{\mathrm{W}} + \mathbf{S}_{\mathrm{xx}}^{\mathrm{B}}$$

Using these expressions we can relate the three conditional means, as follows. The slope in the pooled data (Total) will be a weighted average of the average slope within groups and the slope of the means between groups.

$$b^T = F^W b^W + F^B b^B$$

where $F^W = [S^W_{xx} + S^B_{xx}]^{-1} S^W_{xx} = I - F^B$.

We may nest random effects within this structure. Let $G^W = [S^W_{xx} + \frac{\sigma^2_\epsilon}{\sigma^2_\epsilon + T\sigma^2_\alpha} S^B_{xx}]^{-1} S^W_{xx}$. If the ratio $\frac{\sigma^2_\epsilon}{\sigma^2_\epsilon + T\sigma^2_\alpha} = 1$, then OLS is efficient. However, to the extent that this raio is less than 1, OLS will be inefficient, because it will give too much weight to the between regression and variance. The GLS estimator shifts the weight of the estimators based on the percent of the variance in $u$ accounted for by the within group error and the between group error.

Of note, this is under the assumption that $\alpha$'s are uncorrelated with $X$. We may use this result to consider how different the total regression is from the within (fixed effects) model.

$$b^T - b^W = (F^W - I)b^W + F^B b^B = (I - F^W)(b^B - b^W)$$

Taking expectations of the left side we can express this as $E[b^T] - \beta$, or the bias in $b^T$. Analyzing the right side, we see that the bias in the OLS regression equals the bias in the between regression times the fraction of the variance in $X$ is that is accounted for by the between regression. If there is little bias in the between regression, then the total regression will not be seriously biased. We might tolerate some small degree of bias in order to make more efficient use of the data.

### 6.5. Non-spherical errors in panels

Heteroskedasticity and autocorrelation in panel models can be dealt with using the same methods developed under OLS. Robust standard errors may be used to correct for heteroskedaticity. We typically will use the within group residuals in estimating the variance parameter. The Prais-Winsten transformation may be used to adjust for autocorrelation. Again, the within group residuals are used to estimate $\rho$. When both are problems, we first use Prais-Winsten, and then, using the autocorrelation corrected standard errors, construct the GLS estimator.

These problems are second-order issues, as they affect the efficiency of the estimates. In relatively short panels, however, these differences can have noticeable affects on estimates, producing small sample biases. In fact, in short samples, autocorrelation can cause LSDV and differences on differences to diverge somewhat.