Chapter 14 Advanced Panel Data Methods

1. David Neumark and William Wascher published a study in 1992 of the effect of minimum wages on teenage employment using a U.S. state panel. The paper used annual observations for the years 1977-1989 and included all 50 states plus the District of Columbia. The estimated equation is of the following type:

*Eit* = β0 + β1 (*Mit* /*Wit* ) + γ2*D*2*i* + • • • + γ*nD*51*i* + δ2*B*2*t* + • • • + δTB13t + *uit*

where *E* is the employment to population ratio of teenagers, *M* is the nominal minimum wage, and *W* is average wage in the state. In addition, other explanatory variables, such as the prime-age male unemployment rate, and the teenage population share were included. D2-D51 are the state dummies, B2-B13 are the time dummies.

(a) Briefly discuss the advantage of using panel data in this situation rather than pure cross-sections or time series.

*There are likely to be omitted variables in the above regression. On way to deal with some of these is to introduce state and time dummy variables to capture the fixed component of the error term. State dummy variables will capture the influence of omitted variables that are state specific and do not vary over time, while the time dummies will capture any country-wide variables that are common to all states at a point in time. (Furthermore, there are also more observations when using panel data, resulting in more variation.)*

(b) Estimating the model by OLS but including only time dummy variables results in the following output

*Eit* = β0 – 0.33 × (*Mit* /*Wit* ) + 0.35 × (*SHY it* ) – 1.53 × *uram it* ; =.20

 (.08) (.28) (.13)

where *SHY* is the proportion of teenagers in the population, and *uram* is the prime-age male unemployment rate. Coefficients for the time fixed effects are not reported.

Numbers in parentheses are homoskedasticity-only standard errors.

Comment on the above results. Are the coefficients statistically significant? Since these are level regressions, how would you calculate elasticities?

*There is a negative relationship between minimum wages and the employment to population ratio. Increases in the share of teenagers in the population result in higher employment to population ratio, and increases in the prime-age male unemployment rate lower the employment to population ratio. 20 percent of employment to population of teenagers variation is explained by the above regression. The relative minimum wage and the prime-age male unemployment rate are significant using a 1% significance level, while the proportion of teenagers in the population is not.*

*The important point is to note that elasticities vary with levels. One possibility is to report elasticities at the sample means.*

(c) Adding state fixed effects changes the above equation as follows:

*Eit* = β0 + 0.07 × (*Mit* /*Wit* ) – 0.19 × (*SHY it* ) – 0.54 × *uram it* ; = 0.69

(0.10) (0.22) (0.11)

Compare the two results. Why would the inclusion of state dummy variables change the coefficients in this way?

*The parameter of interest here is the coefficient on the relative minimum wage. While it was highly significant in the previous regression, it has now changed signs and is statistically insignificant. The explanatory power of the equation has increased substantially. The size of the other two coefficients has also decreased. The results suggest that omitted variables, which are now captured by state dummy variables, were correlated with the regressors and caused omitted variable bias.*

(d) The significance of each coefficient decreased, yet  increased. How is that possible? What does this result tell you about testing the hypothesis that all of the state fixed effects can be restricted to have the same coefficient? How would you test for such a hypothesis?

*The influence of the state dummy variables is large. These are bound to be statistically significant and the hypothesis to restrict these coefficients is zero is bound to fail. Since these are linear hypothesis that are supposed to hold simultaneously, an F-test is appropriate here.*

1. (From 14.6) Using the “cluster” option in STATA, the fully robust (robust to serial correlation and heteroskedasticity) standard errors for the pooled OLS estimates in Table 14.2 are

se(educ) = .011, se(black) =.051, se(hispan) =.039, se(exper) =.020,

se(exper2) =.001, se(married) =.026, se(union) =.027

1. How do these standard errors generally compare with the nonrobust ones? Why?

*The fully robust standard errors are larger in each case, roughly double for the time-constant regressors educ, black, and hispan. On the time-varying explanatory variables married and union, the fully robust standard errors are roughly 60 percent larger. The differences are smaller for exper and exper2 but hardly trivial. We expect this if we think the composite error term, , contains an unobserved effect, . This induces positive serial correlation and, as we saw in Section 12.1 for time series, the usual OLS standard errors tend to understate the actual sampling variation in the OLS estimates. The same holds true for pooled OLS with panel data.*

1. How do the robust standard errors for the pooled OLS compare with the standard errors for random effects? Does it seem to matter whether the explanatory variable is time-varying or time-constant?

*On the time constant explanatory variables educ, black, and hispan, the RE standard errors and the robust standard errors for pooled OLS are roughly the same. (The coefficient estimates are very similar, too.) The main differences arise in the standard errors (and coefficients) on the time-varying explanatory variables. For example, the RE standard errors on the married and union coefficients are .017 and .018, respectively, compared with the robust standard errors for pooled OLS of .026 and .027. We expect this to be true because, under the under the RE assumptions, RE is more efficient than pooled OLS.*

1. A fellow classmate has written their own code to estimate within, between, and random effects models with balanced panels. As a check on their work, they estimate between, within and random effects models for the one-way effects models with one covariate.
2. The equation that describes the fixed and random effects models is Yit=Xitβ + ui + εit. The results are summarized below. You look at the results and tell the student they have a programming error. What tipped you off?

Parameter Estimates and Standard Errors

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Between | Fixed | Random |
| Xit | 0.01092(0.00214) | 0.03285(0.00251) | 0.03312(0.00267) |

*The random effects estimator can be thought of as a weighted average of the “between” and “within” (fixed effect) estimators. Here it is larger than both, so these results cannot be right.*

1. Your classmates fixes their coding error and generates the following results for a different problem. These are correct. What is the Hausman test statistic for the null hypothesis that ui and Xit are uncorrelated? Can you reject or not reject the null?

Parameter Estimates and Standard Errors

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Between | Fixed | Random |
| Xit | 0.11012(0.04412) | 0.04813(0.02429) | 0.0635(0.02137) |

* = -.0154/.0112 =1.33 which is distributed as a t-statistic. This is clearly not significant at any conventional level. This means that we cannot statistically distinguish the two estimates, leading us to not be able to reject the null.*

1. Download the Stata dataset called **panel\_hw.dta**. This dataset examines voter turnout in 49 US states (Louisiana is omitted because of an unusual election in 1982) plus the District of Columbia over 11 elections (contains data on 50 units over 11 time periods). Submit write-ups for these problems as well as the log files.
	1. Regress turnout as a percent of voting age population on the number of days before the general election by which an individual needs to register, state per capita income, the dummy variable for midterm elections, and the dummy variables for West North Central, the South, and the Border states.

. regress vaprate gsp midterm regdead WNCentral South Border

 Source | SS df MS Number of obs = 550

-------------+------------------------------ F( 6, 543) = 182.08

 Model | 38318.2811 6 6386.38019 Prob > F = 0.0000

 Residual | 19045.9213 543 35.0753616 R-squared = 0.6680

-------------+------------------------------ Adj R-squared = 0.6643

 Total | 57364.2025 549 104.488529 Root MSE = 5.9224

------------------------------------------------------------------------------

 vaprate | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 gsp | -.0382637 .0178458 -2.14 0.032 -.073319 -.0032085

 midterm | -13.37857 .5072368 -26.38 0.000 -14.37495 -12.38218

 regdead | -.2171977 .0333501 -6.51 0.000 -.2827088 -.1516867

 WNCentral | 3.736791 .799348 4.67 0.000 2.166598 5.306984

 South | -7.969966 .7031458 -11.33 0.000 -9.351186 -6.588747

 Border | -7.035122 .8146468 -8.64 0.000 -8.635368 -5.434877

 \_cons | 61.87341 1.02298 60.48 0.000 59.86393 63.88289

------------------------------------------------------------------------------

. test WNCentral South Border

 ( 1) WNCentral = 0

 ( 2) South = 0

 ( 3) Border = 0

 F( 3, 543) = 73.05

 Prob > F = 0.0000

Which coefficients are significant? Are there any regional effects of these regions? Use F-test to determine this.

*All of the independent variables are significant. Yes—regional effects are significant both separately and jointly using the F test.*

* 1. Part (i) assumed that pooling the data was valid. Instead, estimate this with a fixed effects regression. Which variables are omitted from the estimation? Why?

. iis stcode

. tis year

. xtreg vaprate midterm gsp regdead WNCentral South Border, fe

Fixed-effects (within) regression Number of obs = 550

Group variable: stcode Number of groups = 50

R-sq: within = 0.7363 Obs per group: min = 11

 between = 0.0096 avg = 11.0

 overall = 0.4275 max = 11

 F(2,498) = 695.15

corr(u\_i, Xb) = 0.0023 Prob > F = 0.0000

------------------------------------------------------------------------------

 vaprate | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 midterm | -13.37218 .3586782 -37.28 0.000 -14.07689 -12.66747

 gsp | -.0246353 .0171186 -1.44 0.151 -.058269 .0089983

 regdead | (dropped)

 WNCentral | (dropped)

 South | (dropped)

 Border | (dropped)

 \_cons | 54.67635 .5900227 92.67 0.000 53.51711 55.83559

-------------+----------------------------------------------------------------

 sigma\_u | 6.6883066

 sigma\_e | 4.187412

 rho | .71840339 (fraction of variance due to u\_i)

------------------------------------------------------------------------------

F test that all u\_i=0: F(49, 498) = 12.00 Prob > F = 0.0000

. estimates store fe

*Regdead and the regional dummies dropped because these do not vary within states over time. Obviously, the region of the country a state is in does not vary. The results indicate that the number of days before the general election by which an individual needs to register also did not change within states during this time period and therefore also dropped out..*

* 1. Test for whether there is evidence of unobserved heterogeneity—that is, whether a fixed effects model is more appropriate than the pooled model. What is your testing hypothesis? What is your test statistic? Is pooling appropriate in light of the results of your test?

*The testing hypothesis is that the state dummies are equal to zero. The F-test is a test of this. The final line of the table is*

F test that all u\_i=0: F(49, 498) = 12.00 Prob > F = 0.0000

*We can reject the null that the dummies are all equal to zero with a p-value of .0000. As a result, we conclude that pooling is not appropriate—the state fixed effects are significant.*

* 1. Now estimate a random-effects model. Why do those results differ from the fixed effects results? Is there evidence of unobserved heterogeneity? Show your testing hypothesis, and decide on it. Are some variables omitted from the estimation? Why or why not?

. xtreg vaprate midterm gsp regdead WNCentral South Border, re

Random-effects GLS regression Number of obs = 550

Group variable: stcode Number of groups = 50

R-sq: within = 0.7363 Obs per group: min = 11

 between = 0.5741 avg = 11.0

 overall = 0.6677 max = 11

Random effects u\_i ~ Gaussian Wald chi2(6) = 1452.03

corr(u\_i, X) = 0 (assumed) Prob > chi2 = 0.0000

------------------------------------------------------------------------------

 vaprate | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 midterm | -13.37301 .358402 -37.31 0.000 -14.07547 -12.67056

 gsp | -.0264047 .0165908 -1.59 0.111 -.0589221 .0061126

 regdead | -.2183449 .0859471 -2.54 0.011 -.3867981 -.0498918

 WNCentral | 3.772746 2.058301 1.83 0.067 -.2614497 7.806942

 South | -7.904291 1.798542 -4.39 0.000 -11.42937 -4.379214

 Border | -7.07554 2.096786 -3.37 0.001 -11.18516 -2.965916

 \_cons | 61.51503 2.225863 27.64 0.000 57.15242 65.87764

-------------+----------------------------------------------------------------

 sigma\_u | 4.4345407

 sigma\_e | 4.187412

 rho | .5286392 (fraction of variance due to u\_i)

------------------------------------------------------------------------------

. estimates store re

*Note that this differs from the fixed effects estimates for several reasons. First, we get estimates for the time-invariant variables. Second, these results use the variation between states as well as the variation within states across time, although the effect of using this variation is pretty small on both the magnitude of the coefficients and their standard errors.*

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

 vaprate[stcode,t] = Xb + u[stcode] + e[stcode,t]

 Estimated results:

 | Var sd = sqrt(Var)

 ---------+-----------------------------

 vaprate | 104.4885 10.22196

 e | 17.53442 4.187412

 u | 19.66515 4.434541

 Test: Var(u) = 0

 chi2(1) = 673.91

 Prob > chi2 = 0.0000

*This indicates that we can reject the null that var(ui) = 0 with a high degree of confidence (p value = .000). This implies that there is serial correlation in our errors, and that the random effects model is more appropriate than an OLS model.*

* 1. Which model is more appropriate, FE or RE? What is your underlying testing hypothesis? What implication does the null hypothesis have? Discuss the tradeoffs between using pooled OLS, fixed-effects, and random-effects for this model.

. hausman fe re

 ---- Coefficients ----

 | (b) (B) (b-B) sqrt(diag(V\_b-V\_B))

 | fe re Difference S.E.

-------------+----------------------------------------------------------------

 midterm | -13.37218 -13.37301 .000829 .0140729

 gsp | -.0246353 -.0264047 .0017694 .0042182

------------------------------------------------------------------------------

 b = consistent under Ho and Ha; obtained from xtreg

 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

 Test: Ho: difference in coefficients not systematic

 chi2(2) = (b-B)'[(V\_b-V\_B)^(-1)](b-B)

 = 0.18

 Prob>chi2 = 0.9158

*This test shows us whether the coefficients under the random effects model differ statistically from the fixed effects model. The null hypothesis is that the coefficients are not different. In this case, they are not—the p-value associated with a test that they are the same is .9158, so we cannot reject the null hypothesis. The implication is that while different states may have systematically different errors, the errors are not correlated with the X variables. As a result, we would prefer the somewhat more efficient estimates from the random effects model. (Note, however, that there is little difference between the standard errors of the fixed and random effects models.)*

**Replication Exercise**

One important question in public finance is whether the structure of the welfare system affects marriage and family formation. This assignment examines the impact of welfare waivers (IMPDUM) on the propensity to be a never married woman (NEVERMAR). Prior to the 1996 federal Welfare Reform Act, states could apply for waivers to experiment with alternative systems—here related to marriage. For more information, see Marianne Bitler, Jonah Gelbach, Hilary Hoynes, and Madeline Zovodny (2004) “The Impact of Welfare Reform on Marriage and Divorce.”

This assignment is focused on identifying and implementing a model using variation across states and over time in a particular policy.

Download the dataset **wr-nevermar.dta**. The dataset consists of a sample of black women ages 16-54 with a high school education or less from 1989-1996. The data come from the March CPS.

1. Summarize and describe the data. Graph the mean of NEVERMAR by year. Describe the trend in this variable.

 (use the egen command to create the mean of this variable by year)

. sum

 Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

 state | 11367 49.84666 20.05837 11 95

 year | 11367 92.30289 2.268571 89 96

 rmaxpay | 11367 4.774162 2.033734 1.47304 12.5996

 urate | 11367 6.251227 1.358195 2.2 11.4

 empgrat | 11367 1.509997 1.736326 -4.594637 9.862548

-------------+--------------------------------------------------------

 impdum | 11367 .1185009 .3232145 0 1

 everwaiv | 11367 .6145861 .4867144 0 1

 black | 11367 1 0 1 1

 msadum | 11367 .8155186 .3878933 0 1

 age1625 | 11367 .3205771 .466719 0 1

-------------+--------------------------------------------------------

 age2634 | 11367 .2479106 .4318187 0 1

 age3544 | 11367 .2458872 .4306309 0 1

 age4554 | 11367 .1856251 .3888209 0 1

 nevermar | 11367 .4980206 .5000181 0 1

. des

state byte %9.0g State id number

year byte %8.0g Survey year

rmaxpay float %9.0g Max bens for fam of 3, in 1000s of 1997$

urate float %9.0g State unemployment rate

empgrat float %9.0g State empl growth rate

impdum byte %9.0g =1 if waiver implemented before march of this year, & tanf not yet

everwaiv byte %9.0g state ever had a major waiver

black byte %8.0g non-Hispanic black

msadum byte %9.0g =1 if in msa

age1625 byte %8.0g age between 16 and 25

age2634 byte %8.0g age between 26 and 34

age3544 byte %8.0g age between 35 and 44

age4554 byte %8.0g age between 45 and 54

nevermar float %9.0g =1 if never married

egen meannevermar = mean(nevermar), by(year)

twoway (line meannevermar year, sort), ytitle(Percent women never married--mean across states)



1. Now graph the mean of the policy variable IMPDEM over time. How is that variable trending over time? Discuss the trends in (1) and (2) together.

egen meanimpdum = mean(impdum), by(year)

twoway (line meanimpdum year, sort), ytitle(Percent states with waiver)



*Both variables are trending upward. The percent never married increased, and then dropped. After 1993, it picked up again. State waivers began in 1993 as well and expanded rapidly.*

1. Collapse the data to the year level and estimate the model:

NEVERMARt = α +γIMPDUMt + et

Where is the identification coming from in this time series model? Why might this be a biased estimate of γ? If it is biased, what do you think the key omitted variable is? Why don’t you control for it? Does this empirical identification strategy make sense?

Don’t forget to preserve the data before you do this and restore after. Read up on the collapse command—you want to collapse by year.

. preserve

. collapse nevermar impdum, by(year)

. reg nevermar impdum

 Source | SS df MS Number of obs = 8

-------------+------------------------------ F( 1, 6) = 11.50

 Model | .003838615 1 .003838615 Prob > F = 0.0146

 Residual | .002002373 6 .000333729 R-squared = 0.6572

-------------+------------------------------ Adj R-squared = 0.6001

 Total | .005840988 7 .000834427 Root MSE = .01827

------------------------------------------------------------------------------

 nevermar | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 impdum | .1466028 .0432267 3.39 0.015 .0408309 .2523747

 \_cons | .4806534 .0086008 55.88 0.000 .4596079 .5016989

------------------------------------------------------------------------------

. restore

*The identification here is coming from the time series variation—essentially, this is a regression correlating information in the two graphs we just produced. This indicates what we saw--that waivers and percent never married both increased over this time period. However, there are many reasons why this is likely to be biased. A major concern is that states may have chosen to use waivers* **because** *they were experiencing the rise in percent unmarried. The omitted variable would be anything that is leading to higher rates of never married women. This is hard to control for because there are many reasons why this might be on the rise, including some “cultural” explanations that are hard to quantify. This empirical identification strategy would NOT make sense if the trends in never married are themselves causing the adoption of waivers. Note that this observation of concurrent time trends strongly suggests that these will need to be controlled for in the analysis.*

1. Now estimate a cross sectional model for 1995:

NEVERMARi,s,95 = α +γIMPDUMs,95 + εi,s,95

Where is the identification coming from in this cross sectional model? Why might this be a biased estimate of γ? If it is biased, what do you think the key omitted variable is? Why don’t you control for it? Does this empirical identification strategy make sense? What if you had estimated the model using only data from 1991?

. reg nevermar impdum if year==95

 Source | SS df MS Number of obs = 1319

-------------+------------------------------ F( 1, 1317) = 0.22

 Model | .055724792 1 .055724792 Prob > F = 0.6367

 Residual | 328.791887 1317 .249652154 R-squared = 0.0002

-------------+------------------------------ Adj R-squared = -0.0006

 Total | 328.847612 1318 .249505017 Root MSE = .49965

------------------------------------------------------------------------------

 nevermar | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 impdum | .0144801 .0306489 0.47 0.637 -.0456459 .0746061

 \_cons | .5221053 .0162109 32.21 0.000 .4903033 .5539072

------------------------------------------------------------------------------

*Here the identification is coming from cross state differences—comparing states in 1995 that have and have not adopted waivers. Again, this may be biased if states that adopted waivers are systematically different from states that did not. See above for the same sorts of issues about omitted variables. If we had estimated this in 1991, there would have been no estimate on IMPDUM because no states had waivers in that year—there was no cross sectional variation.*

1. Now return to the full model. Estimate the simplest model

NEVERMARi,s,t = α +γIMPDUMs,t + εi,s,t

Interpret the magnitude of the coefficient on IMPDUM. (Remember when you do that, NEVERMAR is 0/1). Is this what you expected to find? Relate your answer to (3).

. reg nevermar impdum

 Source | SS df MS Number of obs = 11367

-------------+------------------------------ F( 1, 11365) = 3.71

 Model | .926411264 1 .926411264 Prob > F = 0.0542

 Residual | 2840.77905 11365 .249958562 R-squared = 0.0003

-------------+------------------------------ Adj R-squared = 0.0002

 Total | 2841.70546 11366 .250018077 Root MSE = .49996

------------------------------------------------------------------------------

 nevermar | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 impdum | .0279323 .0145091 1.93 0.054 -.0005079 .0563726

 \_cons | .4947106 .0049946 99.05 0.000 .4849203 .5045008

*These results that adopting a waiver leads to a 2.7 percentage point reduction in the fraction of women who have never married. No—this is not what we expected to find. Note that this result is between the time series and cross sectional results—the time series variation appears to be partly responsible for these results.*

1. Reestimate the model in (5) accounting for the fact that IMPDUM varies only at the state level, while the data is at the individual level. In STATA, use

reg yvar xvars, cluster(state)

 You should cluster your standard errors for the rest of this problem.

. reg nevermar impdum, cluster(state)

Linear regression Number of obs = 11367

 F( 1, 50) = 1.02

 Prob > F = 0.3175

 R-squared = 0.0003

 Root MSE = .49996

 (Std. Err. adjusted for 51 clusters in state)

------------------------------------------------------------------------------

 | Robust

 nevermar | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 impdum | .0279323 .0276615 1.01 0.317 -.0276275 .0834922

 \_cons | .4947106 .0106169 46.60 0.000 .4733859 .5160352

------------------------------------------------------------------------------

*Note that only the standard errors change, and that they are larger as we might expect.*

1. Add demographic variables and economic variables to the regression (age variables, MSADUM, URATE, EMPGRAT, RMAXPAY). Interpret the coefficient sand relate them to what you might have predicted based on economic theory. Be careful with units when you interpret the variables.

. reg nevermar impdum age2634 age3544 age4554 msadum urate empgrat rmaxpay, cluster(state)

Linear regression Number of obs = 11367

 F( 8, 50) = 1353.26

 Prob > F = 0.0000

 R-squared = 0.3387

 Root MSE = .40676

 (Std. Err. adjusted for 51 clusters in state)

------------------------------------------------------------------------------

 | Robust

 nevermar | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 impdum | .0368661 .0229587 1.61 0.115 -.0092478 .0829799

 age2634 | -.3763929 .0130164 -28.92 0.000 -.402537 -.3502487

 age3544 | -.6116304 .0125927 -48.57 0.000 -.6369235 -.5863373

 age4554 | -.7420067 .0130351 -56.92 0.000 -.7681884 -.715825

 msadum | .0434587 .0167829 2.59 0.013 .0097492 .0771682

 urate | .0034714 .0048595 0.71 0.478 -.0062893 .0132321

 empgrat | -.0087565 .0056462 -1.55 0.127 -.0200971 .0025841

 rmaxpay | .0008234 .0049802 0.17 0.869 -.0091795 .0108264

 \_cons | .8272401 .0401193 20.62 0.000 .746658 .9078221

------------------------------------------------------------------------------

*This indicates that women 26-34 are 37 percentage points less likely to have never married that women 16-35 (the omitted category). Living in a MSA is associated with a 4 percent higher probability of being never married. A one percentage point increase in the unemployment rate is association with .3 percentage point increase in the probability of being never married, although this is not statistically significant. A one thousand dollar increase in the state maximum benefits for a family of three is associated with a .08 percentage point increase in the probability of being never married; again this is not significant.*

1. Add year fixed effects to the model (keep in the demographic and economic variables). That is, estimate

NEVERMARi,s,t = α +γIMPDUMs,t + βX i,s,t + ηt + εi,s,t

Use the xi: reg command to do this so you can see the coefficients on the fixed effects.

How does the identification in THIS model differ from that in (3)? What happens to the coefficient on IMPDUM? Explain why the coefficient changed as it did using reasoning related to omitted variable bias.

. xi: reg nevermar impdum age2634 age3544 age4554 msadum urate empgrat rmaxpay i.year, cluster(state)

i.year \_Iyear\_89-96 (naturally coded; \_Iyear\_89 omitted)

Linear regression Number of obs = 11367

 F( 15, 50) = 1109.32

 Prob > F = 0.0000

 R-squared = 0.3429

 Root MSE = .40559

 (Std. Err. adjusted for 51 clusters in state)

------------------------------------------------------------------------------

 | Robust

 nevermar | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 impdum | -.0063255 .0242885 -0.26 0.796 -.0551103 .0424594

 age2634 | -.3761322 .0130121 -28.91 0.000 -.4022677 -.3499966

 age3544 | -.6135607 .0123234 -49.79 0.000 -.638313 -.5888084

 age4554 | -.7441743 .0126631 -58.77 0.000 -.769609 -.7187397

 msadum | .0367137 .015829 2.32 0.024 .0049203 .0685071

 urate | .0033846 .0050333 0.67 0.504 -.0067251 .0134943

 empgrat | -.0155337 .0060332 -2.57 0.013 -.0276518 -.0034156

 rmaxpay | .0017219 .0048723 0.35 0.725 -.0080644 .0115081

 \_Iyear\_90 | .010087 .0166175 0.61 0.547 -.0232901 .0434642

 \_Iyear\_91 | -.0035181 .0261727 -0.13 0.894 -.0560875 .0490512

 \_Iyear\_92 | .0198713 .020406 0.97 0.335 -.0211153 .060858

 \_Iyear\_93 | .0487032 .0205992 2.36 0.022 .0073285 .0900779

 \_Iyear\_94 | .0730097 .0198747 3.67 0.001 .0330902 .1129291

 \_Iyear\_95 | .1017709 .0177305 5.74 0.000 .0661581 .1373837

 \_Iyear\_96 | .0866582 .0198517 4.37 0.000 .046785 .1265314

 \_cons | .8062789 .0456483 17.66 0.000 .7145917 .8979661

*Now the identification is coming from average cross-state differences in the faction never married-- this is comparing the average difference between states with and without waivers in each year and averaging that difference for all years. Note that the coefficient is much less positive—the sign has switched. This is because of the overall effects of time—the percent never married was rising in ALL states over time, just as the number of waivers was rising. Time (or time proxying for time-varying factors that raised the percent never married) is the omitted variable that is now controlled for.*

1. Add state fixed effects to the model along with the year fixed effects.

NEVERMARi,s,t = α +γIMPDUMs,t + βXi,s,t + ξs + ηt + εi,s,t

Describe how the identification in THIS model differs from that in (3). What happens to the coefficient on IMPDUM? Explain why the coefficient changed as it did using reasoning related to omitted variable bias.

. xi: reg nevermar impdum age2634 age3544 age4554 msadum urate empgrat rmaxpay i.year i.state, cluster(state)

i.year \_Iyear\_89-96 (naturally coded; \_Iyear\_89 omitted)

i.state \_Istate\_11-95 (naturally coded; \_Istate\_11 omitted)

Linear regression Number of obs = 11367

 F( 14, 50) = .

 Prob > F = .

 R-squared = 0.3522

 Root MSE = .4036

 (Std. Err. adjusted for 51 clusters in state)

------------------------------------------------------------------------------

 | Robust

 nevermar | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 impdum | -.0367239 .024847 -1.48 0.146 -.0866306 .0131829

 age2634 | -.3771264 .0130595 -28.88 0.000 -.4033572 -.3508957

 age3544 | -.6153964 .012548 -49.04 0.000 -.6405998 -.5901931

 age4554 | -.7461648 .0131186 -56.88 0.000 -.7725142 -.7198154

 msadum | .0128066 .014799 0.87 0.391 -.016918 .0425312

 urate | -.0048936 .0077298 -0.63 0.530 -.0204194 .0106322

 empgrat | .0038596 .0049854 0.77 0.442 -.0061539 .0138731

 rmaxpay | -.024263 .016962 -1.43 0.159 -.0583323 .0098063

 \_Iyear\_90 | .0202421 .0143847 1.41 0.166 -.0086505 .0491347

 \_Iyear\_91 | .0548824 .0209919 2.61 0.012 .0127189 .0970459

 \_Iyear\_92 | .0528563 .0185285 2.85 0.006 .0156407 .0900719

 \_Iyear\_93 | .0459372 .0202839 2.26 0.028 .0051958 .0866786

 \_Iyear\_94 | .0615356 .0202155 3.04 0.004 .0209315 .1021397

 \_Iyear\_95 | .0823574 .0160891 5.12 0.000 .0500415 .1146733

 \_Iyear\_96 | .0801679 .0171204 4.68 0.000 .0457806 .1145551

 \_Istate\_12 | .2475448 .025829 9.58 0.000 .1956658 .2994239

 \_Istate\_13 | .3825208 .0458277 8.35 0.000 .2904732 .4745685

 \_Istate\_14 | .4211545 .0355765 11.84 0.000 .3496969 .4926121

 \_Istate\_15 | .3137652 .0302577 10.37 0.000 .2529908 .3745397

Etc.

*By adding state fixed effects, this model is now essentially a difference in difference model. The identification is coming from the change in the percent never married in states that passed waivers compared to the change in states that did not. This accounts for the fact that states that passed waivers may always have had percent never married that differed from states that did not pass waivers. It also accounts for the fact that the percent never married was rising over time in all states. The omitted variables here may be anything that both caused high percent never married and a high probability of passing a waiver—that would have led the coefficients to be positively biased. Indeed, we see here that accounting for this leads to much more negative coefficients.*

1. Add state specific time trends to the model:

NEVERMARi,s,t = α +γIMPDUMi,s,t + βX i,s,t + ΘsTime + ξs + ηt + εi,s,t

How do the results for IMPDUM change with the time trends? Explain why the results changed.

To do this, use this command:

 xi: reg *yvar xvars* i.year i.state\*year, cluster(state)

. xi: reg nevermar impdum age2634 age3544 age4554 msadum urate empgrat rmaxpay i.year i.state\*year, cluster(state)

i.year \_Iyear\_89-96 (naturally coded; \_Iyear\_89 omitted)

i.state \_Istate\_11-95 (naturally coded; \_Istate\_11 omitted)

i.state\*year \_IstaXyear\_# (coded as above)

Linear regression Number of obs = 11367

 F( 13, 50) = .

 Prob > F = .

 R-squared = 0.3557

 Root MSE = .40341

 (Std. Err. adjusted for 51 clusters in state)

------------------------------------------------------------------------------

 | Robust

 nevermar | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 impdum | -.0663318 .0218197 -3.04 0.004 -.110158 -.0225056

 age2634 | -.3766153 .0129324 -29.12 0.000 -.4025909 -.3506398

 age3544 | -.6157429 .0125436 -49.09 0.000 -.6409375 -.5905482

 age4554 | -.7467554 .0135146 -55.26 0.000 -.7739003 -.7196104

 msadum | .0107566 .0146224 0.74 0.465 -.0186134 .0401266

 urate | -.0064583 .0095549 -0.68 0.502 -.0256499 .0127333

 empgrat | .0061329 .0058943 1.04 0.303 -.0057061 .017972

 rmaxpay | .0228221 .040529 0.56 0.576 -.0585828 .1042271

 \_Iyear\_90 | .0122398 .0134809 0.91 0.368 -.0148375 .039317

 \_Iyear\_91 | .0432842 .0246012 1.76 0.085 -.0061288 .0926972

 \_Iyear\_92 | .0240937 .0193269 1.25 0.218 -.0147255 .0629128

 \_Iyear\_93 | .0032504 .0194948 0.17 0.868 -.0359061 .0424069

 \_Iyear\_94 | .0026042 .0194623 0.13 0.894 -.036487 .0416954

 \_Iyear\_95 | .0104986 .0134653 0.78 0.439 -.0165472 .0375445

 \_Iyear\_96 | (dropped)

 \_Istate\_12 | 2.314271 1.211019 1.91 0.062 -.1181317 4.746674

 \_Istate\_13 | 17.93038 1.383517 12.96 0.000 15.1515 20.70925

 \_Istate\_14 | 14.01565 .4043179 34.66 0.000 13.20355 14.82774

 \_Istate\_15 | 12.41331 .4210346 29.48 0.000 11.56764 13.25898

 \_Istate\_16 | 14.56202 1.062809 13.70 0.000 12.42731 16.69674

 \_Istate\_21 | 12.27712 .3500845 35.07 0.000 11.57396 12.98029

etc

 \_Istate\_94 | 9.16446 .5937657 15.43 0.000 7.971847 10.35707

 \_Istate\_95 | 11.31737 1.033968 10.95 0.000 9.24058 13.39415

 year | .1432159 .013527 10.59 0.000 .1160461 .1703857

\_IstaXyea~12 | -.022273 .0133652 -1.67 0.102 -.0491179 .0045718

\_IstaXyea~13 | -.1904142 .0140535 -13.55 0.000 -.2186414 -.1621869

Etc

*After adding time trends, the results are even larger—a waiver is associated with a 6.6 percentage point reduction in the percent never married. This specification controls for the fact that states may have had different trends in the rates of never married. Like the last specification, this is essentially a difference in difference specification. Instead of essentially just comparing the average difference in never married before and after the policy waiver, this specification compares the difference relative to the average trend for that state.*