1. [Heteroskedasticity] There are different ways to combine features of the Breusch-Pagan and White tests for heteroskedasticity. One possibility that we did not cover is to run the regression:

on x1i, x2i, x3i, . . . ., xki, and 

where the are the squares of the OLS residuals (estimated under homoskedasticty) and the are the squares of the predicted values from the OLS regression.

1. (6pts) How would you construct the LM statistic for this test? What are its degrees of freedom?
2. (5pts) How will the R2 from this regression compare with the R2 from the usual BP test and the White test?
3. (5pts) Does this imply that the new test statistic always delivers a smaller/larger p-value than the BP or White tests? Explain.
4. (5pts) Suppose someone suggested also adding to this regression. What do you think of this idea?
5. [Serial correlation] One type of partial adjustment model is

y\*t = γ0 + γ1xt+et

yt-yt-1 = λ(y\*t-yt-1) + ut

where y\*t is the desired or optimal level of y and yt is the actual observed level. For example, y\*t is the desired growth in firm inventories and xt is growth in firm sales. The parameter γ1 measures the effect of xt on y\*t. The second equation describes how the actual y depends on the relationship between the desired y in time t and the actual y in time t-1. The parameter λ measures the speed of adjustment and satisfies 0<λ<1.

1. Show that we can rewrite the model as a regression of yt on xt and yt-1. Show how the coefficients in this model relate to the parameters in the two equation model above, and show how the error term relates to the error terms above.
2. If E(et|xt, yt-1, xt-1, yt-2, . . . .) = E(ut|xt, yt-1, xt-1, yt-2, . . . .) = 0 and all series are weakly dependent, if we estimate the model you specified in (a) by OLS will we get consistent estimates? Explain why or why not.
3. Will the errors in your model in (a) be serially correlated? Explain why or why not.
4. [Time series models: tests, interpretation] Let hy6t denote the three-month holding yield (in percent) from buying a 6 month T-bill at time (t-1) and selling it at time t (three months later) as a three month T bill.Let hy3t-1 be the three month holding yield (in percent) from buying a 3 month T bill at time (t-1). At time (t-1), then, hy3t-1 is known, whereas hy6t is unknown because the price in time t of three month T-bills is unknown at time t-1.

The expectations hypothesis says that, of course, these two different three month investments should be the same, on average. In other words,

E(hy6t|all information up to time t-1) = hy3t-1

Suppose you were to estimate the model

hy6t = β0 + β1 hy3t-1 + ut

1. 4 pts How would you test the expectations hypothesis using this model? What is the null hypothesis?
2. 4pts Suppose your estimates of that equation produced the following

^hy6t = -.058 + 1.104 hy3t-1 + ut

 (.070) (.039)

N=123, R2 = .866

Do you reject the test in (a) at the 5% significance level?

1. 4pts Another implication of the expectations hypothesis is that no other variables dated as t-1 or earlier should explain hy6t after controlling for hy3t-1. Suppose you were to test this implication by estimating the following equation, which includes the lagged spread between the 6 and 3 month T-bill rates:

^hy6t = -.123+ 1.053 hy3t-1 + .480(r6t-1 – r3t-1)

 (.067) (.039) (.109)

According to this, is the lagged spread term significant at the 5% level? What do the results imply about whether you should invest in 6-month or 3-month T bills if at time t-1, r6 is greater than r3?

1. 3pts Conduct the test you specified in (a) using the results in (c). Do you conclude anything different from your results in (b)?
2. 6pts The sample correlation between hy3t and hy3t-1 is .914. Does this raise any concerns with your previous analysis? Be specific.
3. [Measurement error] (15 points) Suppose one is interested in estimating a linear model of the form y=xβ+ε but the outcome of interest (y) is measured with error. Suppose the true value of the ith observation for y is *y\*i* and the measured value is yi = yi\* + ei, where ei is an i.i.d. error. All other usual Gauss-Markov assumptions hold.

Let $\hat{β} $be the OLS estimate for β using the mismeasured value of y and $\hat{β^{\*}}$ be the estimate if y\* was available. Show that in the presence of measurement error, $\hat{β} $is still an unbiased estimate but var($\hat{β})>var(\hat{β^{\*}}$ ).

1. [Panel Model specification—general principles ] For one semester, you collect data on a random sample of college juniors and senior for each class taken. In other words, you have multiple observations for each student, with a different observation for each class the student takes. This information includes measures of a standardized final exam score, percentage of lectures attended, a dummy variable indicating whether the class is within the student’s major, cumulative GPA prior to the semester, and SAT score.
2. (6pts) Write a model for final exam performance in terms of attendance and other characteristics. Use subscripts for students and classes.
3. (6pts) Suppose that your main interest is the effect of attendance on final exam performance. If you pool all of the data and use OLS, what are you assuming? Explain this in the context of the other variables in your model.
4. (10 pts) What other models (other than OLS) might you consider using? Carefully describe under what circumstances would you use those models?
5. [Panel Model specification—general principles ] Suppose an author has data on outcomes (Y) for two periods, 1 and 2, for a number of cross-sectional units (e.g., firms). Suppose also that a subgroup of firms were hit with an intervention (e.g., minimum wage hike) sometime between period 1 and 2.
	1. 6 pts Write an equation for the difference-in-difference estimate (e.g., one that controls for group and time effects) that provides a consistent estimate of the intervention by using the first differences in outcomes between two periods. That is, the dependent variable would be ΔYi = Y2i-Y1i.
	2. 5 pts What is the advantage of the difference-in difference approach in this context?
	3. 5pts Are there potential problems with inference that the difference-in-difference approach does not resolve?
6. [Panel Model specification—tests] You have estimated the effect of X on Y using a number of different models:

Parameter Estimates and Standard Errors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Pooled | Between | First Difference | Fixed | Random |
| Xit | 0.3201(0.0842) | 0.1201(0.0542) | 0.0543(0.0321) | 0.0271(0.0242) | 0.0735(0.0223) |

1. What is the Hausman test statistic for the null hypothesis that ui and Xit are uncorrelated?
2. Can you reject the null?
3. What does that mean about which estimator you would prefer to use?
4. Explain the intuition for this test. [Why does the answer to (b) imply the estimator you provided in (c)?]
5. [Panel model specification—interpretation] In 1992, there was an increase in the (state) minimum wage in one U.S. state (New Jersey) but not in a neighboring location (eastern Pennsylvania). The study provides you with the following information,

|  |  |  |
| --- | --- | --- |
|  | *PA* | *NJ* |
| *FTE Employment before* | 23.33 | 20.44 |
| *FTE Employment after* | 21.17 | 21.03 |

Where *FTE* is “full time equivalent” and the table reports the average *FTE*s per fast food restaurant.

1. Calculate the change in the treatment group, the change in the control group, and finally 
2. Since minimum wages represent a price floor, did you expect to be positive or negative?
3. The standard error for is 1.36. Test whether or not the coefficient is

statistically significant, given that there are 410 observations.

1. If you believed that the benefit from small minimum wage increases outweighed the cost in terms of employment loss, would finding that this coefficient was not statistically significant discourage you?
2. Do you have any concerns about forming a conclusion about the effect of the minimum wage law using this approach? (Stronger answers will make specific reference to the information presented.)
3. [Panel model specification—interpretation] You are interested in the effect of airline concentration on prices. You have annual data for 1997 through 2000 for 1,149 different routes (id), log of the fares on those routes (lfare), the log of distance of the route (ldist) and distance squared (dist 2) along with the fraction of the market held by the carrier on that route (concen).

Use the following output to discuss your preferred specification to model the effect of concentration on price.

For all parts:

Consult and interpret the provided output and test statistics. If there are other test statistics that you would construct to determine the appropriate model, briefly describe your null hypothesis, what you would do to construct the test statistic and how you would interpret it.

Just in case you’ve forgotten, Stata reports the Root Mean Squared Error, which is the same thing as $\sqrt{\frac{SSR}{dof}}$ where SSR is the sum of squared residuals (or sum of squared predicted errors).

1. (10 points) Consider first whether it is appropriate to pool the data and estimate by OLS or whether separate annual regressions are more appropriate. Which model would you prefer?
2. (10 points) Using the pooled model, explain how you would test the hypothesis that the coefficients on ldist and ldist2 are zero. Can you carry out this test using the fixed effect results? Explain.
3. (10 points) Is your preferred model the annual regressions, pooled OLS, the fixed effect model, or the random effect model? Provide any test statistics that you need to make this decision and interpret them.

**SEPARATE ANNUAL REGRESSIONS**

. by year: reg lfare concen ldist dist2, robust

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

-> year = 1997

Linear regression Number of obs = 1149

 F( 3, 1145) = 290.12

 Prob > F = 0.0000

 R-squared = 0.4076

 Root MSE = .356

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

 | Robust

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

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

 concen | .3950364 .067515 5.85 0.000 .2625694 .5275035

 ldist | -.9360734 .2985159 -3.14 0.002 -1.521773 -.3503738

 dist2 | .10807 .0221572 4.88 0.000 .0645968 .1515431

 \_cons | 6.190051 .9996943 6.19 0.000 4.228613 8.151489

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

-> year = 1998

Linear regression Number of obs = 1149

 F( 3, 1145) = 327.91

 Prob > F = 0.0000

 R-squared = 0.4318

 Root MSE = .32214

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

 | Robust

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

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

 concen | .3758987 .0613662 6.13 0.000 .2554959 .4963016

 ldist | -.9088614 .2576948 -3.53 0.000 -1.414468 -.4032543

 dist2 | .1041423 .01911 5.45 0.000 .0666478 .1416368

 \_cons | 6.218367 .8629141 7.21 0.000 4.525296 7.911437

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

-> year = 1999

Linear regression Number of obs = 1149

 F( 3, 1145) = 274.08

 Prob > F = 0.0000

 R-squared = 0.4065

 Root MSE = .3343

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

 | Robust

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

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

 concen | .3889226 .0633928 6.14 0.000 .2645436 .5133016

 ldist | -.8048083 .2775762 -2.90 0.004 -1.349423 -.2601933

 dist2 | .0962123 .0206322 4.66 0.000 .0557312 .1366934

 \_cons | 5.889738 .9285482 6.34 0.000 4.067891 7.711584

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

-> year = 2000

Linear regression Number of obs = 1149

 F( 3, 1145) = 228.75

 Prob > F = 0.0000

 R-squared = 0.3665

 Root MSE = .33238

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

 | Robust

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

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

 concen | .2813454 .0614613 4.58 0.000 .160756 .4019348

 ldist | -.9681315 .2864295 -3.38 0.001 -1.530117 -.4061459

 dist2 | .1045345 .0212688 4.91 0.000 .0628042 .1462647

 \_cons | 6.733451 .9591077 7.02 0.000 4.851646 8.615257

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

**POOLED OLS MODELS**

. reg lfare concen y98 y99 y00, robust

Linear regression Number of obs = 4596

 F( 4, 4591) = 70.67

 Prob > F = 0.0000

 R-squared = 0.0546

 Root MSE = .42449

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

 | Robust

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

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

 concen | -.4875843 .030758 -15.85 0.000 -.5478847 -.427284

 y98 | .0286962 .0181277 1.58 0.113 -.0068428 .0642352

 y99 | .0313444 .0182368 1.72 0.086 -.0044085 .0670973

 y00 | .0905699 .0179255 5.05 0.000 .0554272 .1257125

 \_cons | 5.35543 .0218546 245.05 0.000 5.312585 5.398276

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

. reg lfare concen ldist dist2 y98 y99 y00, robust

Linear regression Number of obs = 4596

 F( 6, 4589) = 558.39

 Prob > F = 0.0000

 R-squared = 0.4062

 Root MSE = .33651

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

 | Robust

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

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

 concen | .3601203 .0318147 11.32 0.000 .2977482 .4224925

 ldist | -.9016004 .1406543 -6.41 0.000 -1.177351 -.6258503

 dist2 | .1030196 .0104402 9.87 0.000 .0825518 .1234875

 y98 | .0211244 .0141734 1.49 0.136 -.0066623 .048911

 y99 | .0378496 .0144012 2.63 0.009 .0096162 .0660829

 y00 | .09987 .0143821 6.94 0.000 .0716742 .1280658

 \_cons | 6.209258 .4711359 13.18 0.000 5.285605 7.132911

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

. reg lfare concen ldist dist2 y98 y99 y00, cluster(id)

Linear regression Number of obs = 4596

 F( 6, 1148) = 205.63

 Prob > F = 0.0000

 R-squared = 0.4062

 Root MSE = .33651

 (Std. Err. adjusted for 1149 clusters in id)

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

 | Robust

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

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

 concen | .3601203 .058556 6.15 0.000 .2452315 .4750092

 ldist | -.9016004 .2719464 -3.32 0.001 -1.435168 -.3680328

 dist2 | .1030196 .0201602 5.11 0.000 .0634647 .1425745

 y98 | .0211244 .0041474 5.09 0.000 .0129871 .0292617

 y99 | .0378496 .0051795 7.31 0.000 .0276872 .048012

 y00 | .09987 .0056469 17.69 0.000 .0887906 .1109493

 \_cons | 6.209258 .9117551 6.81 0.000 4.420364 7.998151

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

**RANDOM EFFECTS MODEL**

. xtreg lfare concen ldist dist2 y98 y99 y00, re

Random-effects GLS regression Number of obs = 4596

Group variable: id Number of groups = 1149

R-sq: within = 0.1348 Obs per group: min = 4

 between = 0.4176 avg = 4.0

 overall = 0.4030 max = 4

 Wald chi2(6) = 1360.42

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

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

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

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

 concen | .2089935 .0265297 7.88 0.000 .1569962 .2609907

 ldist | -.8520921 .2464836 -3.46 0.001 -1.335191 -.3689931

 dist2 | .0974604 .0186358 5.23 0.000 .0609348 .133986

 y98 | .0224743 .0044544 5.05 0.000 .0137438 .0312047

 y99 | .0366898 .0044528 8.24 0.000 .0279626 .0454171

 y00 | .098212 .0044576 22.03 0.000 .0894752 .1069487

 \_cons | 6.222005 .8099666 7.68 0.000 4.6345 7.80951

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

 sigma\_u | .31933841

 sigma\_e | .10651186

 rho | .89988885 (fraction of variance due to u\_i)

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

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

 lfare[id,t] = Xb + u[id] + e[id,t]

 Estimated results:

 | Var sd = sqrt(Var)

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

 lfare | .1904449 .4363999

 e | .0113448 .1065119

 u | .101977 .3193384

 Test: Var(u) = 0

 chibar2(01) = 5566.11

 Prob > chibar2 = 0.0000

**FIXED EFFECT MODEL**

. xtreg lfare concen ldist dist2 y98 y99 y00, fe

note: ldist omitted because of collinearity

note: dist2 omitted because of collinearity

Fixed-effects (within) regression Number of obs = 4596

Group variable: id Number of groups = 1149

R-sq: within = 0.1352 Obs per group: min = 4

 between = 0.0576 avg = 4.0

 overall = 0.0083 max = 4

 F(4,3443) = 134.61

corr(u\_i, Xb) = -0.2033 Prob > F = 0.0000

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

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

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

 concen | .168859 .0294101 5.74 0.000 .1111959 .226522

 ldist | 0 (omitted)

 dist2 | 0 (omitted)

 y98 | .0228328 .0044515 5.13 0.000 .0141048 .0315607

 y99 | .0363819 .0044495 8.18 0.000 .0276579 .0451058

 y00 | .0977717 .0044555 21.94 0.000 .089036 .1065073

 \_cons | 4.953331 .0182869 270.87 0.000 4.917476 4.989185

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

 sigma\_u | .43389176

 sigma\_e | .10651186

 rho | .94316439 (fraction of variance due to u\_i)

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

F test that all u\_i=0: F(1148, 3443) = 60.52 Prob > F = 0.0000

hausman fe re

 ---- Coefficients ----

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

 | fe re Difference S.E.

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

 concen | .168859 .2089935 -.0401345 .0126937

 y98 | .0228328 .0224743 .0003585 .

 y99 | .0363819 .0366898 -.000308 .

 y00 | .0977717 .098212 -.0004403 .

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

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

 = 10.00

. hausman re fe

 ---- Coefficients ----

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

 | re fe Difference S.E.

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

 concen | .2089935 .168859 .0401345 .

 y98 | .0224743 .0228328 -.0003585 .0001597

 y99 | .0366898 .0363819 .000308 .0001699

 y00 | .098212 .0977717 .0004403 .0001377

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

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

 = -10.00