Econometric Analysis and Endogeneity

Class 5

Econometric Analysis in ECO410H

- Econometrics used in competition policy to:
 - Estimate demand for use in merger simulation
 - A hard econometric task: available data are usually observational (endogeneity bias)
 - Examine consummated merger's impact on competition: merger retrospectives
 - Hard to estimate impact: mergers are not random
 - Study how market structure affects competition
 - Sutton's *Sunk Costs and Market Structure*; S-C-P Structure-Conduct-Performance literature

Today's Agenda

- A fast-paced review of inference with multiple regression (pre-requisite material)
 - Highlight issue of endogeneity and key terms
- Some econometric solutions to endogeneity
 - Difference-in-difference ("diff-in-diff") with specific merger retrospective application
 Also discuss basic panel data solutions: fixed effects
 - Mention instrumental variables, but specific discussion and applications in Class 6

1

2

Multiple Regression Model

- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$
 - $-y_i$: dependent variable, LHS variable, y-variable
 - $-x_{1i}, ..., x_{ki}$: independent variables, RHS variable, explanatory variables, x-variables
 - *i*: observation index (often *i* or *j* cross-sectional data; *t* time series; *it* or *jt* panel)
 - $-\beta_0$: intercept (constant) parameter
 - $-\beta_1, \dots, \beta_k$: slope parameters
 - $-\varepsilon_i$: error term, residual, disturbance

Sample
cross-
sectional
data
Sample
time
series
data

Cross-Sectional and Time Series

4

6

Sample Panel (Longitudinal) Data

area_id	week	p1	p2	р3	q1	q2	q2	inc	рор
1	1								
1	2								
1	Т								
2	1								
2	2								
2	Т								
К	Т								

Ordinary Least Squares (OLS)

- OLS is one method of estimating the parameters of linear regression model
 - OLS estimates solve: $\min_{b_0,\dots,b_k} \sum_{i=1}^n (y_i \widehat{y_i})^2$
 - Where $\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki}$
 - Use statistical software package (e.g. Stata)
 - Coefficient estimates and the standard errors of the coefficient estimates valid only if underlying assumptions met

Underlying Assumptions

7

8

9

- Linearity: each x linearly related to y (x variables and/or y variable can be non-linearly transformed)
- 2) Errors independent (common problem: autocorrelation in time series data)
- 3) Homoscedasticity (single variance) of errors
- 4) Normally distributed errors
- 5) Constant included (error has mean 0)
- Each x and error unrelated; i.e. x variables are exogenous (not endogenous), no unobserved/lurking/confounding/omitted variables

New Type of Demand Curve?

- Henry J. Moore, "father of economic statistics," conduced regressions for many industries in early 20th century
- In some regressions found negative demand elasticities, but in pig iron, for example, found positive demand elasticity and concluded he "had discovered a new type of demand curve with positive slope"

If estimate $Q_t = \alpha + \beta P_t + \varepsilon_t$ using prices and quantities demanded, P is endogenous: i.e. is related to the error term, which includes demand shifters



Direction of Bias

• In a simple regression (only one explanatory variable), can sign the direction of bias

 $y_i = \alpha + \beta x_i + \varepsilon_i$

- If x_i and ε_i are positively correlated then this causes an upward bias: $E[b] > \beta$
- If x_i and ε_i are negatively correlated then this causes an downward bias: $E[b] < \beta$
- Which is the case in the failed attempt to estimate the demand for pig iron?

Collins and Preston (1966)

Food manufacturing industry	Price-Cost Margin	CR4
meat packing	4.32	34
prepared meats	8.04	17
flavorings	39.71	55
cottonseed oil mills	4.61	42
soyabean oil meal	7.42	40
grease and tallow	15.35	23
macaroni and spaghetti	18.84	25

N.R. Collins and L.E. Preston. 1966. "Concentration and price-cost margins in food manufacturing industries." *The Journal of Industrial Economics* 14(3): 226-242 Which kind of data are these? 11

Simple Regression and Graph

- R²: 40% of variation in price-cost margins across industries explained by variation in the CR4
 - Interpret constant?
 - Interpret slope?
- Standard errors (s.e.) in parentheses below the point estimates



$$\hat{y}_j = 6.05 + 0.31 x_j$$

(3.19) (0.07)

13

Standard Error of OLS Slope

- SE(b₁) reflects size of sampling error and depends on:
 - 1) Sample size (n)
 - 2) Amount of scattering about line (s_e)

3) How much x-variable varies in the data (s_r)

•
$$SE(b_1) = s_{b_1} = \frac{s_e}{s_x \sqrt{n-1}}$$

For multiple regression, no simple formula (use software).



Testing Slope Coefficients

- $H_0: \beta_i = \beta_i^0$
- $H_1: \beta_j \neq \beta_j^0$ - Or $H_1: \beta_j > \beta_j^0$ - Or $H_1: \beta_j < \beta_j^0$
- Use *t* test statistic with v = n k 1:

$$t = \frac{b_j - \beta_j^0}{s_{b_j}}$$

- The standard error of the slope coefficient (s_{b_j}) – like the slope coefficient itself – is obtained w/ software
- "Test of statistical significance":

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Statistically Significant?

- Economists often use two tailed tests
 - Even when a directional test seems more obvious
 - It is conservative: fewer statistically significant results
- Informal rule of thumb: t test statistic >2 or <-2?
 - If yes, coefficient is "statistically significant" (reject H₀ in favor of H₁)





The equation obtained is as follows (t ratios of the coefficients of regression are shown in parentheses):



17

Collins and Preston (1966)

. regress gamma_1 x_1 x_1_sq x_2 x_3 x_4;

Source	SS	df	MS	Number of obs	= 32
Model Residual 	2765.78621 697.467573	5 553. 26 26.8	157242 256759 	Prob > F R-squared Adj R-squared	= 0.0000 $= 0.7986$ $= 0.7599$ $= 5.1704$
gamma_1	Coef.	Std. Err.	t P> t	[95% Conf.	Interval]
x_1 x_1_sq x_2 x_3 x_4 	3095445 .0079021 1352038 .2802349 000887 12.4046	.1962929 .0020352 .0265599 .0780401 .0106402 4.803231	-1.58 0.12 3.88 0.00 -5.09 0.00 3.59 0.00 -0.08 0.93 2.58 0.01	77130303 1 .0037187 01897985 1 .1198212 40227583 5 2.531418	.0939413 .0120855 0806091 .4406487 .0209843 22.27778

Can compare P-values with conventional significance levels of

 $\alpha = 0.01, \alpha = 0.05$ and $\alpha = 0.10$.

Pig Iron & Margins Analysis: Recap

- Both Moore and Collins and Preston relied on *observational data*, real market outcomes
 - The key x-variables price of pig iron or industry concentration – were *endogenous* (correlated with the error, violating the underlying assumptions), causing *endogeneity bias* of parameter estimates
 - In contrast, in *experimental data* the values of x are randomly set (i.e. x variable is *exogenous*) and OLS yields unbiased estimates of the *causal* relationship
 - But why? Lots of variables in ε with experimental data?

19

Specification and Misspecification

- Model specification includes choices about:
 - Functional form such as logarithms
 - Including dummy/indicator variables/fixed effects
 - Including interaction terms
 - Approaches to addressing endogeneity
- A misspecified model relies on faulty assumptions and yields biased estimates
 - Standard errors do not reflect these errors

20

Endogeneity: Some Solutions

- Collect data on things in the error (correlated with x variables) and include them as RHS variables
 - Conceptually simple, but often impossible
- Include fixed effects (requires panel data)
 - If omitted variables vary (only) over time then control for them with a full set of time fixed effects:

 $Y_{it} = \alpha + \beta x_{it} + \delta_t + \varepsilon_{it}$

• When would you use this: $Y_{it} = \alpha + \beta x_{it} + \delta_i + \varepsilon_{it}$?

- Difference-in-difference (requires panel data)
- Use instrumental variables (requires instruments)

Ashenfelter and Hosken (2010) "The Effect of Mergers on Consumer Prices: Evidence from Five Mergers on the Enforcement Margin" *Journal of Law and Economics*

Abstract: In this paper we propose a method to evaluate the effectiveness of U.S. horizontal merger policy and apply it to the study of five recently consummated consumer products mergers. We select the mergers from those that, from the public record, seem most likely to be problematic. Thus, we estimate an upper bound on the likely price effect of completed mergers. Our study employs retail scanner data and uses familiar panel data program evaluation procedures to measure price changes. Our results indicate that four of the five mergers resulted in some increases in consumer prices, while the fifth merger had little effect.

http://www.nber.org/papers/w13859

22

Ready to Eat Breakfast Cereal (RTE Cereal)

- Jan. 1997 General Mills acquired the branded cereal business of Ralcorp for \$570M
 - Ralcorp: Chex
 - General Mills: many RTE cereals including Cheerios and Wheaties
 - Other major firms: Post, Kellogg's, and Quaker





23

Scanner Data

- Weekly total revenue and unit sales for each UPC (Universal Product Code) for 64 metropolitan areas
 - Obtained from retail scanners
 - According to these data, the revenue-based market share of General Mills is 28% and Ralcorp 4%: post-merger HHI 2,357 with ΔHHI 238
 - Must aggregate because many package sizes
 - "Price" is defined as a weighted average

"Diff-in-Diff" Approach: Concept

- City A affected by a merger in period 1

 If observe higher prices in period 2, conclusion?
- City B (the control group) is similar to City A but not affected by the merger; Again, measure the price change from period 1 to 2
 - If the *change* in prices in City A is greater than the *change* in prices in City B, conclusion?
 - Common trends assumption: while the levels may differ, the trends in two cities same but for the merger

25

"Diff-in-Diff": Lots of Possibilities

- But what if all cities affected by the merger at the same time, like in RTE cereal case?
 - Product A affected by a merger
 - Product B (the control group) is similar to Product
 A but not affected by the merger
 - Which kinds of products would be in the control group?
 - Prices of each observed before and after merger
 - If the *change* in price of Product A is greater than the *change* in price of Product B, conclusion?

26

Preferred Control Group, p. 23

- Use the private label cereals in same product category as Chex, Wheaties and Cheerios
 - IRI (data source, paid): Groups cereals, e.g. "Adult Fruit and Nut" and "All Family Wholesome"
 - Branded products sell at a premium: e.g. Cheerios
 58% more expensive than the private label version
 - Input costs (except advertising) similar
 - Distant enough substitutes so that merger should have little effect on private label prices

Alternate Control Group, p. 24

- Branded RTE cereals in same IRI category (closest substitutes)
 - Advantage: control for shocks to both cost and demand; e.g. increase in income likely to boost branded sales at expense of private label sales
 - Disadvantage: the prices of close substitutes directly affected by the merger
 - Why?
 - May lead to an underestimate of the merger's effect

28

Empirical Specification

• For *each* of the merging parties products estimate:

$$p_{ijt} = \alpha_{ij} + \beta_1 P M_t + \beta_2 P M_t * M P P_i + \sum_{m=1}^{11} \delta_m M_{mt} + \varepsilon_{ijt}$$

- *i* indexes products, *j* cities (64 cities), *t* time (32 months)
- p_{ijt} is the natural log of price (weighted average price)
- α_{ii} is a region-specific, product-specific fixed effect
- $-PM_t = 1$ after merger consummated, 0 otherwise
- $-MPP_i = 1$ if product owned by merging firms, 0 otherwise
- $-M_{mt}$ are month-specific (e.g. March) fixed effects

"PM" = post-merger "MPP" = Merging parties' product

29

"Diff-in-Diff": Mechanics

• For *each* of the merging parties products estimate:

 $p_{ijt} = \alpha_{ij} + \beta_1 P M_t + \beta_2 P M_t * M P P_i + \sum_{m=1}^{11} \delta_m M_{mt} + \varepsilon_{ijt}$

- Includes one of the merging parties products and control group: e.g. i = 1 Cheerios; i = 2 Private label "Ohs"...

	Δ prices
Control group (e.g. comparable private labels)	eta_1
Merging parties' product (e.g. Cheerios)	$\beta_1 + \beta_2$

Hence, the difference in differences is β_2 : the increase in price of the merging parties' product caused by the merger

Semi-log functional form

- If estimate $\widehat{\ln(y_i)} = \hat{\alpha} + \hat{\beta}x_i$, how to interpret $\hat{\beta}$? E.g., suppose $\hat{\beta} = 0.02$?
 - When x is 1 unit higher, on average, y is 2% higher
 See "Logarithms in Regression Analysis"
- Recall, p_{iit} is the natural log of price

$$p_{ijt} = \alpha_{ij} + \beta_1 P M_t + \beta_2 P M_t * M P P_i + \sum_{m=1}^{11} \delta_m M_{mt} + \varepsilon_{ijt}$$

– Interpretation if $\hat{\beta}_2 = 0.044?$

31

Preferred Event Window, p. 22

- Drop 3 months before and after merger closes and use same number of months before/after
 - Compare fully-coordinated with not-at-allcoordinated pricing
 - Monthly data from 07/1995 to 08/1998 excluding 11/1996 to 04/1997, for a total of 32 months
 - 64 cities * 32 months * 5 to 6 products ≈ 11,000 to 13,500 observations (private label control group)
 - 64 cites * 32 months * 17 to 27 products ≈ 36,000 to 55,000 observations (branded control group)

32

Table 7: Estimated Price Effects of General Mills' Purchase of Ralcorp (s.e.'s in parentheses)

Product	Control Group: Other Brands	Control Group: Private Label
Cheerios	0.046 (0.011)	0.044 (0.010)
Cheerios Apple Cinn	0.138 (0.013)	0.067 (0.011)
Cheerios Honey Nut	0.105 (0.012)	0.035 (0.011)
Cheerios Multi-Grain	0.029 (0.009)	0.027 (0.007)
Corn Chex	0.005 (0.017)	0.003 (0.017)
Multi-Grain Chex	0.003 (0.014)	0.001 (0.014)
Rice Chex	0.007 (0.017)	0.005 (0.016)
Wheat Chex	0.008 (0.017)	0.006 (0.017)
Wheaties	0.027 (0.013)	0.026 (0.012)
Wheaties Crispy Raisin	0.090 (0.020)	0.028 (0.019)
Wheaties Honey Frosted	0.110 (0.017)	0.048 (0.014)

Looking Ahead

- TA Tutorial on Tuesday, 11:10 1:00 LA 211
 Econometrics review with Victor
- How to approach the required readings for our last lecture (Class 6):
 - Hausman and Leonard (2002) "The Competitive Effects of a New Product Introduction: A Case Study"
 - Hosken et al (2011) "Does Concentration Matter? Measurement of Petroleum Merger Price Effects,"

34