Economics 220C
Empirical Methods in Industrial Organization
Bronwyn H. Hall
Economics, UC Berkeley
Spring 2005
Outline

• Testing Gibrat’s Law
  – Introduction to unbalanced panels
  – Selection

• SCP and some stylized facts

• Brief discussion of choosing a paper topic
  – Ideas?
  – Finding data
Gibrat’s law example

Ijiri-Simon (1964) assumptions (see also Sutton (1998):

1. Prob(next investment opportunity taken up by any particular firm) proportional to current size.
2. Prob(next investment opportunity taken up by new entrant) constant over time.

Generates log-normal size distribution and Gibrat’s law. How to test?

Gibrat’s law: firm growth is independent of size.

An example of a statistical model that predicts a conditional expectation (not a structural model).
Testing Gibrat’s law

$y_{it} = \log$ of employment for firm $i$ at time $t$.
$\Delta y_{it} = y_{it} - y_{i,t-1} = \text{growth of employment}$

The model is

$$E[\Delta y_{it} \mid y_{i,t-1}] \neq f(y_{i,t-1})$$

Consider the following regression model:

$$\Delta y_{it} = \alpha_t + \beta y_{i,t-1} + \epsilon_{it}$$

Gibrat’s law says the estimated slope should be zero. But to get consistent estimates we need

$$E[\epsilon_{it} \mid y_{i,t-1}] = 0$$
Hall (1987)

Why might $E[\varepsilon_{it} | y_{i,t-1}] = 0$ fail?

1. Measurement error in $y_{i,t-1}$ will be in the disturbance:
   - observed size: $w_{it} = y_{it} + \nu_{it}$
   - true regression: $\Delta y_{it} = \alpha + \varepsilon_{it}$

   \[
   E[\Delta w_{it} | w_{i,t-1}] = E[\Delta y_{it} + \nu_{it} - \nu_{i,t-1} | y_{i,t-1} + \nu_{i,t-1}]
   \]

   \[
   = \alpha + 0 - \frac{\sigma_\nu^2}{\sigma_w^2} y_{i,t-1}
   \]

2. If regression uses only data observed at $t$, we are computing

   \[
   E[\Delta y_{it} | y_{i,t-1}, \text{data observed at } t] = \alpha + E[\varepsilon_{it} | y_{i,t-1}, \text{data observed at } t]
   \]

   and this quantity may not have a zero expectation if the exit process is non-ignorable.
Panel exit and entry

Unbalanced panels (different numbers of time periods per firm) are usually due to
- Missing data
- Entry of new firms – birth, listing
- Exit of old firms – bankruptcy, liquidation, acquisition, etc.

Whether or not this feature of a panel causes difficulty for estimation depends on whether the process generating exit/entry is “ignorable.”
- Ignorable means that the disturbance in the equation of interest is independent of the selection rule.
Attrition can be a problem

- Computing the impact of policy change on profits or productivity of firms in an industry
  - Could cause exit or mergers, so we observe effects only on surviving firms
  - If surviving firms a nonrandom sample (likely), and we observe effects of change on them only, results are biased.

- Not a problem if we want net effects conditional on exit patterns induced by that policy change, but a problem for structural estimates
Estimation with unbalanced data

Process generating exit/entry/missing is ignorable:
- Most estimators are easily extended to accommodate unbalanced panels (sometimes with the use of a selection matrix containing ones and zeros). There is no need to model the process generating selection.

Process generating exit/entry/missing is non-ignorable:
- To obtain consistent estimates, need to build a model of the process and describe its relation to the equation of interest.
Example – unbalanced sample

<table>
<thead>
<tr>
<th>Year</th>
<th>Entrants</th>
<th>Exits</th>
<th>Missing</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td></td>
<td></td>
<td></td>
<td>1208</td>
</tr>
<tr>
<td>1986</td>
<td>139</td>
<td>93</td>
<td>0</td>
<td>1254</td>
</tr>
<tr>
<td>1987</td>
<td>150</td>
<td>77</td>
<td>0</td>
<td>1327</td>
</tr>
<tr>
<td>1988</td>
<td>99</td>
<td>115</td>
<td>0</td>
<td>1311</td>
</tr>
<tr>
<td>1989</td>
<td>63</td>
<td>110</td>
<td>2</td>
<td>1262</td>
</tr>
<tr>
<td>1990</td>
<td>78</td>
<td>76</td>
<td>2</td>
<td>1262</td>
</tr>
<tr>
<td>1991</td>
<td>112</td>
<td>57</td>
<td>2</td>
<td>1315</td>
</tr>
<tr>
<td>1992</td>
<td>144</td>
<td>63</td>
<td>0</td>
<td>1396</td>
</tr>
<tr>
<td>1993</td>
<td>144</td>
<td>65</td>
<td>3</td>
<td>1472</td>
</tr>
<tr>
<td>1994</td>
<td>132</td>
<td>100</td>
<td>2</td>
<td>1502</td>
</tr>
<tr>
<td>1995</td>
<td>113</td>
<td>230</td>
<td>1</td>
<td>1384</td>
</tr>
</tbody>
</table>
Size at entry

Histograms by Entrant

Log of employment

Entrant  Incumbent
Size at exit

Histograms by Exit status

Log of employment

Spring 2005 Economics 220C
Summary

<table>
<thead>
<tr>
<th></th>
<th>Entrants</th>
<th>Incumbents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geometric mean</td>
<td>202</td>
<td>638</td>
</tr>
<tr>
<td>Median</td>
<td>165</td>
<td>527</td>
</tr>
<tr>
<td>Exit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geometric mean</td>
<td>275</td>
<td>613</td>
</tr>
<tr>
<td>Median</td>
<td>228</td>
<td>502</td>
</tr>
</tbody>
</table>
Firm size and growth – 1988-93

Growth = \(-0.03\% - 1.41\% \log E(-5)\)  
\(s=15.0\%\)  
\((0.52\%)  (0.22\%)\)
Firm exit 1988-93

Exit: mean employment = 164
No exit: mean employment = 400
Is exit related to size-growth relation?

Log of employment
Histograms by sample
Sample selection example

• Simple two equation model of firm growth, allowing for exit from the sample. Introduction to
  – Modeling with QDV (qualitative dependent variables)
  – Maximum likelihood estimation

• Idea:
  – Firm growth a function of initial size (testing Gibrat’s law)
  – Exit from the sample not uncorrelated with initial size, so the regression is not unbiased (estimates conditional on observing growth are not the same as those that would be obtained if unconditional – exit process is not ignorable).
Parametric selection model

Data on variables $X$, $Z$ for $I = 1, \ldots, N$ individuals.
Observe $y_1$ for only a subset $N_1 = N - N_0$

Model:
\[ y_{1i} = X_i \beta + \nu_{1i} \quad \text{if} \quad y_{2i} > 0 \]
\[ y_{1i} \quad \text{not observed} \quad y_{2i} \leq 0 \]
\[ y_{2i} = Z_i \delta + \nu_{2i} \]
\[ D_{2i} = I (y_{2i} > 0) \]

Regression for observed data:
\[ E[y_1 \mid X, y_2 > 0] = X \beta + E[\nu_1 \mid \nu_2 > -Z \delta] = X \beta + \lambda(Z \delta; \theta) \]
Where $\theta$ are parameter(s) of the joint distribution of the $\nu$s.
Parametric selection model

\[ E[y_1 \mid X, y_2 > 0] = X\beta + E[\nu_1 \mid \nu_2 > -Z\delta] = X\beta + \lambda(Z\delta; \theta) \]

• Implication for estimated coefficients from a regression of \( y_1 \) on \( X \):
  – Intercept biased because mean of the disturbance is not zero.
  – If the \( X \)s and \( Z \)s are not independently distributed (i.e., they have variables in common, or are correlated), the estimated slope coefficients will be biased due to omitted variables.

• Well-known solutions to this problem under normality:
  – Heckit (Probit of \( y_2 \) on \( Z \), compute Mills’ ratio, include in regression of \( y_1 \) on \( X \)).
  – Maximum likelihood for bivariate distribution of \( (\nu_1, \nu_2) \).
    SAMPSEL in TSP; heckman in Stata.

• Later in course: semiparametric version.
Example

Probit (firm survives until 1993)
1201 firms, of which 281 exit, 920 survive.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>dF/dx</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log employment</td>
<td>-1.05</td>
<td>.033</td>
<td>.007</td>
</tr>
<tr>
<td>Cash flow/capital</td>
<td>0.164</td>
<td>.057</td>
<td>.011</td>
</tr>
<tr>
<td>Log Tobin’s q</td>
<td>0.744</td>
<td>.036</td>
<td>.014</td>
</tr>
<tr>
<td>Scaled R-squared</td>
<td>.065</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mills ratio for firm exit

US Hi-tech Firms 1988-1993

Spring 2005 Economics 220C
Sample selection results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Heckman</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.759 (.056)</td>
<td>.654 (.063)</td>
</tr>
<tr>
<td>Log employment</td>
<td>.109 (.023)</td>
<td>.093 (.022)</td>
</tr>
<tr>
<td>Cash flow/capital</td>
<td>.191 (.035)</td>
<td>.199 (.035)</td>
</tr>
<tr>
<td>Log Tobin’s q</td>
<td>.120 (.040)</td>
<td>.208 (.047)</td>
</tr>
<tr>
<td>Intercept</td>
<td>.075 (.016)</td>
<td>.032 (.009)</td>
</tr>
<tr>
<td>Log employment</td>
<td>-.024 (.003)</td>
<td>-.019 (.002)</td>
</tr>
<tr>
<td>Mills (rho)</td>
<td>-.227 (.044)</td>
<td>-.624 (.109)</td>
</tr>
<tr>
<td>Sigma</td>
<td>0.145</td>
<td>0.160 (.007)</td>
</tr>
</tbody>
</table>

- Negative (*conditional*) correlation between survival and growth disturbances (does this make sense?)
- Very slight reduction in size-growth relationship.
- LM Test for non-normality: $\chi^2(2) = 18.0$ (p-value=.000)
Traditional analysis relates concentration to profitability and productivity (see Salinger article, and Schmalensee survey in the Handbook)

– Given market shares $s_i$, concentration measured as
  - K-firm concentration ratio $CR_k = \sum_{i=1}^{k} s_i$
  - Herfindahl $H = \sum_{i=1}^{N} s_i^2 = \frac{1}{N} + N\text{Var}(s_i)$

– Profitability measured as
  - Accounting profits (intertemporal problems)
  - Tobin’s q, forward looking but volatile and omits intangibles
  - “Lerner” index or markup $= (P-MC)/P$
# Selected census data in 1992

<table>
<thead>
<tr>
<th>$H_{\text{decile}}$</th>
<th>$SIC$</th>
<th>Industry</th>
<th>$N$</th>
<th>Output ($M$)</th>
<th>CR4 (%)</th>
<th>$H_{*10000}$</th>
<th>$1/N_{*10000}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3544</td>
<td>Special dies, tools</td>
<td>7223</td>
<td>9295</td>
<td>3</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>3069</td>
<td>Fab. rubber products, nec</td>
<td>984</td>
<td>6937</td>
<td>18</td>
<td>129</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>3552</td>
<td>Textile machinery</td>
<td>480</td>
<td>1565</td>
<td>21</td>
<td>197</td>
<td>21</td>
</tr>
<tr>
<td>30</td>
<td>2821</td>
<td>Plastics materials</td>
<td>241</td>
<td>31557</td>
<td>24</td>
<td>284</td>
<td>42</td>
</tr>
<tr>
<td>40</td>
<td>2371</td>
<td>Fur goods</td>
<td>211</td>
<td>205</td>
<td>32</td>
<td>403</td>
<td>47</td>
</tr>
<tr>
<td>50</td>
<td>2086</td>
<td>Bottled &amp; canned softdrinks</td>
<td>637</td>
<td>25422</td>
<td>37</td>
<td>537</td>
<td>16</td>
</tr>
<tr>
<td>60</td>
<td>3274</td>
<td>Lime</td>
<td>57</td>
<td>904</td>
<td>46</td>
<td>693</td>
<td>175</td>
</tr>
<tr>
<td>70</td>
<td>3731</td>
<td>Ship building and repairing</td>
<td>562</td>
<td>10609</td>
<td>53</td>
<td>878</td>
<td>18</td>
</tr>
<tr>
<td>80</td>
<td>2047</td>
<td>Dog and cat food</td>
<td>102</td>
<td>7024</td>
<td>58</td>
<td>1229</td>
<td>98</td>
</tr>
<tr>
<td>90</td>
<td>2812</td>
<td>Alkalis and chlorine</td>
<td>34</td>
<td>2787</td>
<td>75</td>
<td>1994</td>
<td>294</td>
</tr>
<tr>
<td>100</td>
<td>2823</td>
<td>Cellulosic manmade fibers</td>
<td>5</td>
<td>1748</td>
<td>98</td>
<td>NA</td>
<td>2000</td>
</tr>
</tbody>
</table>
Some stylized facts (Schmalensee)

- Correlation among accounting RORs are high and results not sensitive to choice of measure. Correlation with PCM (Lerner) lower and with $q$ even lower, affecting results.
- Accounting profitability differences among firms tend to persist for long periods.
- At the firm level, industry characteristics account for about 10-25 per cent of cross section variation in accounting RORs.
- Measures of scale economies or capital intensity tend to be positively correlated with industry-level accounting profitability and negatively related to entry.
Some stylized facts (Schmalensee)

• The cross-section relation between concentration and profitability is weak statistically, usually small, and unstable over time and space. That between market share and profitability is somewhat stronger, across but not within industry.

• In manufacturing, both advertising and R&D tend to be positively related to profitability (and concentration), except possibly when concentration is very high.

• Ratio of imports to domestic consumption tends to be negatively correlated with domestic firm profitability, especially when concentration is high.
Overall conclusions

- Qualitatively, US industries can be described as imperfectly competitive, with markups or profits related to sunk costs of entry, and with a long run zero profit equilibrium (entry until fixed or sunk costs are covered, with prices above marginal or short run average cost).
- The typical US industry is heterogeneous, with a fairly skew firm size distribution.
Why is SCP no longer popular?

• Relating concentration and performance is a correlation exercise
  – of interest for descriptive purposes
  – tends to lead to unstable results
  – Not a structural model, although related to one. Cournot with \( \eta = \) demand elasticity

\[
\log \text{PCM}_i = \alpha + \beta \log H_i + \gamma \log \eta_i + \varepsilon_i
\]

• PCM measured using average cost
• Industry level (multi-product) demand elasticities hard to estimate
• An equilibrium relationship (not a regression)
  – Endogeneity of concentration (and other possible RHS variables)

• Later in course look at the NEIO structural approach to identifying oligopoly behavior.
Finding a paper topic

The challenge:

– Find an unanswered question
– Find an area you would like to know more about (exploratory research)
– And if you want it to lead into a thesis topic, finding something that is not too derivative of others’ work....
Where to start looking

• Survey articles often have lists of open research topics
  – JEL, JEP, Handbooks (recent)

• Theoretical articles or your own theory
  – is there a model that needs empirical testing?

• Applied journals
  – Are there unanswered questions?
  – Do you agree with the method used?
  – Can you apply their methodology to other data (as a starting point)?
  – Can you extend their methodology in some way?

• Newspapers/Economist
  – find a current policy question and collect articles on it, looking for some
    questions that need answering or things you do not understand
  – Is conventional wisdom on some topic correct?
Potential data sources - public

• public data bases on firms
  – usually a byproduct of securities regulation. e.g., Compustat, Datastream, Worldscope
  – financial markets data (transaction prices, e.g., CRSP)

• regulatory data (usually in public agencies)

• patent data (public because that's what it means)

• industry-level statistics published by government and quasi-governmental agencies
  – OECD/OCDE, US Census, UN, EU
  – national statistical offices in most countries.

• industry sources (trade publications)
Potential data sources - private

• surveys
• interviews
• personally collected transaction data
  – e.g., Graddy (fish markets), Genesove (used cars)
  – Internet transactions or prices
• various proprietary data sources
  – Derwent for detailed patent data
  – Lexis-Nexis for litigation data