Leveraging Establishment Databases for Economic Research

Examples from Sustained Growth Studies

Gary Kunkle, PhD

BDRC – Research Fellow
CREC – Research Collaborator
Outlier LLC - Founder
Establishment Datasets – The Basics

• Concept: Business Census Data
  >60 million establishments over three decades
  Public, Private, Non-Profit, Government agencies

• Examples:
  – National Establishment Time Series (NETS)
  – InfoGroup’s Historical Data Files

• Types of Variables:
  – Company Identifier -- Contact Info -- Executives
  – Location(s) -- Relocation -- Long/Lat
  – Establishment Type -- Parent/Children -- Ownership Structure
  – Employment -- Sales
  – Industry (SICs, NAICS, Alpha) -- Primary, Secondary, Tertiary
  – Miscellaneous (Minority/Women Owned, Government Contracts, Exporting)

• Cost:
  – State/Region: +/- $10,000 annual
  – National: +/- $200,000 annual
Strengths / Advantages – Empowered Insights

- Only publically available datasets that are similar to a census
- Many **time-series** variables to play with (employees, sales, locations, parents, industries)
- Lively data (not smoothed and averaged)
- Disaggregated to establishment level
  - So, countless ways to create **sub-groups**
- Allows researchers to measure and test relationships far beyond the worn-out heuristics of industry / location / size  ---  Example: FIRM PERFORMANCE
Weaknesses / Disadvantages – Mind the Gap

- Dated, although improving (and still fine for most applications)
- **Variance** (up-dating frequency, reporting accuracy, bias, data holes), but most are predictable and manageable
- Many relationships are **curvilinear** (need other tools than just linear regression)
- Large set – need suitable tools (hard and soft)
- **Merging** with other sets can be challenge
- Some usage restrictions
Analyses Examples – Tools of the Trade

• Descriptive – Frequencies, growth by groups (e.g. size, industry, location)

• Grouping/Profiling – T-Tests, Hierarchical Cluster Analysis (Euclidean Distance)

• Performance, Location Tracking – Longitudinal data

• Spatial Relationships – GIS, Geo Regressions

• Cause/Effect Tests – Lead/Lag, Regressions (caution: many relationships *not* linear)

• Target Outreach, Survey Support – Purposeful sampling, list creation
NETS vs. CES (ES-202) – Total Employment and By Sector

### Alternative Types of Growth

<table>
<thead>
<tr>
<th>Growth Variables</th>
<th>Units of Growth</th>
<th>Changes the odds of this outcome by this amount...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Future Growth and Survival</td>
</tr>
<tr>
<td>Past Growth (2000-2005)</td>
<td></td>
<td>Survival</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Survive</td>
</tr>
<tr>
<td>Sustained</td>
<td>Years of Growth</td>
<td>1.504</td>
</tr>
<tr>
<td>Absolute</td>
<td>Employee Change</td>
<td>1.000</td>
</tr>
<tr>
<td>Relative</td>
<td>% Employee Change</td>
<td>0.997</td>
</tr>
</tbody>
</table>

Based on logistics regression. N=12,933,411. All models are statistically significant; all variables are significant except *.

“In other words, **the more years a business establishment grew in the past, the more likely they are to survive, grow, and achieve exceptional growth in the future**.”

---

**Source:** Kunkle; “Building Scale and Sustaining Growth”; IEGC, 2013; BDRC website.
Laplace (‘Power Law’) Distribution of Sustained Growth

Sustained Growth Distribution

Impact of Sustained Growers - Virginia

<table>
<thead>
<tr>
<th></th>
<th>Establishments</th>
<th>Jobs</th>
<th>Avg. Jobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>All in Virginia (2014)</td>
<td>613,657</td>
<td>5,212,134</td>
<td>8.5</td>
</tr>
<tr>
<td>Survivors (2009-14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>306,770</td>
<td>2,837,341</td>
<td>9.2</td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td>2,741,516</td>
<td>8.9</td>
</tr>
<tr>
<td>Change: # of New Jobs</td>
<td></td>
<td></td>
<td>95,825</td>
</tr>
<tr>
<td>Sustained (2009-14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>3,019</td>
<td>93,413</td>
<td>30.9</td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td>35,299</td>
<td>11.7</td>
</tr>
<tr>
<td>Change: # of New Jobs</td>
<td></td>
<td></td>
<td>58,114</td>
</tr>
<tr>
<td>Change: % Growth</td>
<td></td>
<td></td>
<td>165%</td>
</tr>
<tr>
<td>% of Survivors &amp; of Job Change</td>
<td>1.0%</td>
<td>61%</td>
<td></td>
</tr>
</tbody>
</table>

Cumulative Growth – Learning, Improving, Repeating

Frequency and Amount of Growth

$R^2 = 0.9531$

Years of Net Sustained Growth (2005-2010)

Absolute Growth (or loss)

$n = 12,015,089$

Source: Kunkle; “Building Scale and Sustaining Growth”; IEGC, 2013; accessible via BDRC website.
## Impact of Employee Benefits on Sustained Growth

<table>
<thead>
<tr>
<th>Mixed: Savings and Incentives</th>
<th>Odds Change for Sustained Growth 2007-11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Binary Logistics Regression (Simplified Presentation)</strong></td>
<td><strong>Model 7</strong></td>
</tr>
<tr>
<td>Eligible Participation Growth Score, 2006</td>
<td>22%</td>
</tr>
<tr>
<td>Financial Asset Growth Score, 2006</td>
<td>16%</td>
</tr>
<tr>
<td>Red Flags, 2006</td>
<td>-11%</td>
</tr>
<tr>
<td>2J: 401(k), 2007</td>
<td>41%</td>
</tr>
<tr>
<td>2G: Total part. Directed, 2007</td>
<td>15%</td>
</tr>
<tr>
<td>2I: Stock Bonus, 2007</td>
<td>155%</td>
</tr>
<tr>
<td>2E: Profit Sharing, 2007</td>
<td>-11%</td>
</tr>
</tbody>
</table>

All variables and the models are significant at 0. \( n=351,713 \)

“Overall, these models tell us that the **optimal mix of employee benefit plans encourages savings for all employees** as well as incentives that are **linked to firm performance for management**.”

State-Level Defense Industry

**Average Employment**

- **Fast Growers**
- **Sustained**

![Graph showing average employment over years with two trends: Fast Growers and Sustained.]

**Average Sales**

- Average Sales (Thousands)
- $12,000
- $10,000
- $8,000
- $6,000
- $4,000
- $2,000
- $-

![Graph showing average sales over years with two trends: Fast Growers and Sustained.]

- **n** = 13,264
State-Level Defense Industry

**Average Employment Size**

- Employment size from 2010 to 2015.
- Data shows a steady increase with fluctuations.

**Average Sales Per Establishment**

- Sales growth from 2010 to 2015.
- Fast Growers and Sustained lines exhibit different trends.

**Average Sales Per Employee**

- Sales per employee from 2010 to 2015, with Fast Growers and Sustained lines.
- Variations in sales per employee across the years.
Sustained Growth Companies - Pennsylvania

Source: Kunkle and PA DCED; “Pennsylvania Sustained Growth Study”; PA DCED, 2012.
OLS Residual LISA Clusters

Source: Campbell, James, Kunkle “Sustained Firm Growth and Regional Income Convergence”; Papers in Applied Geography, March 2016.
Thank you!

Gary Kunkle, PhD
704-953-4331
gmkunkle@yahoo.com