Global Firm Dynamic, Productivity, (Mis)Allocations

Sebnem Kalemli-Ozcan, Professor, University of Maryland, NBER, CEPR

Real and Financial Linkages at the Firm Level: Data Sources and Identification

Friday, August 27, 2021
REAL AND FINANCIAL LINKAGES AT THE FIRM LEVEL: DATA SOURCES AND IDENTIFICATION

Şebnem Kalemli-Özcan

August 2021
Stanford Big Data
What are the data sources that links firms’ real decisions (investment, sales, employment) to their financing (bonds, assets, equity, loans)?

Start with US (most problematic) and then talk about other countries.
Firm Level Data Sources: US

- Census-LBD (no finance, no capital, all sectors)
- Census-ASM (no finance, capital, only manufacturing)
**Firm Level Data Sources for US with Firms Financing Info: All suffer from selection problems**

1. **Compustat**: Only listed firms.

2. **Capital IQ**: Mostly listed firms. Very few large private firms after late 1990s with assets over 10 million and more than 500 shareholders (around 7000)

3. **VentureX**: 10,000 VC-financed firms. 80% of the firms are younger than 3 with employment less than 20. Panel over time since 1980s.


7. **Sageworks**: Audited firms; sample is of similar to Orbis (around 200,000 firms). Firms are anonymized.

8. **Orbis**: Similar sample to Sageworks from audits and reports to local chambers of commerce. As firm ID is known, can be matched to Census and selection can be corrected.
New Data for US: Y-14 Supervisory Data on Bank Lending

- Collected as part of the FR Capital Assessments and Stress Testing (CCAR) exercise for all bank holding companies with total consolidated assets above $50 bil ($100 bil in 2019).

- Firm-bank-loan-quarter level with a reporting threshold of $1 million.

- Contractual terms and firm balance sheet items.

- 2012Q3–2019Q4, all sectors. Almost 4 million loan-level observations for 150,000+ corporations, of which 60,000+ have assets less than 10 million.
New Data for US: Y-14 Supervisory Data on Bank Lending

- Collected as part of the FR Capital Assessments and Stress Testing (CCAR) exercise for all bank holding companies with total consolidated assets above $50 bil ($100 bil in 2019).

- Firm-bank-loan-quarter level with a reporting threshold of $1 million.

- Contractual terms and firm balance sheet items.

- 2012Q3–2019Q4, all sectors. Almost 4 million loan-level observations for 150,000+ corporations, of which 60,000+ have assets less than 10 million.

Coverage:

- The banks subject to CCAR account for over 85% of the total assets in the banking sector and provide around 70% of all commercial and industrial lending.

- Supervisory data on private firms’ financing: representative relative to Compustat, QFR, Dealscan, CapitalIQ, SBFS,...

- Y14 firms account 65% of U.S. corporate sector debt and 78% of aggregate U.S. gross output.
New Data for US: Y-14 Supervisory Data on Bank Lending

- Collected as part of the FR Capital Assessments and Stress Testing (CCAR) exercise for all bank holding companies with total consolidated assets above $50 bil ($100 bil in 2019).
- Firm-bank-loan-quarter level with a reporting threshold of $1 million.
- Contractual terms and firm balance sheet items.
- 2012Q3–2019Q4, all sectors. Almost 4 million loan-level observations for 150,000+ corporations, of which 60,000+ have assets less than 10 million.

Coverage:

- The banks subject to CCAR account for over 85% of the total assets in the banking sector and provide around 70% of all commercial and industrial lending.

Supervisory data on private firms’ financing: representative relative to Compustat, QFR, Dealscan, CapitalIQ, SBFS,...

- Y14 firms account 65% of U.S. corporate sector debt and 78% of aggregate U.S. gross output.

SBA/Census definition: SMEs: <500 employee
⇒ 52% of private sector employment, 48% of private sector output
Firm Level Data Sources: World

1. Orbis: most comprehensive coverage, can be made representative for several European countries given regulation (listed firms is 1 percent of the sample, all private)

2. Worldscope: only listed firms

3. Compustat global: only listed firms
State of the Art Data

Credit Registry Matched to Firm and Bank Balance Sheets and Census

- Can only be obtained per country from confidential administrative sources

- Credit Registry will give you the details of the financial transaction between borrower-lender with the identity of the borrower and lender (loan, bond, equity contract. Amount, price, maturity, collateral..)

- Can control for demand for credit and supply of credit separately

- Matching to firm balance sheets give you characteristics of the borrower and real outcomes such as investment, sales, profits..

- Matching to bank balance sheets (or other lenders) give you characteristics of the lender and info on lenders financial strength, portfolio..

- Matching to census gives you link to employment and firm entry-exit.
How to Handle Any Firm Level Data?

- Cleaning
- Selection
- Coverage
Cleaning will differ for all firms and manufacturing firms (dropping or not negative equity (capital) firms)

Cleaning might also differ based on the goals

After cleaning, before empirical analysis, winsorize at the 1 and the 99 percentile at the very least.
Basic Cleaning (Needed for Reporting Mistakes and Data Consistency)

1. Drop firm-year observations that have missing information on total assets and operating revenues and sales and employment.

2. Drop firms if total assets are negative in any year, or if employment is negative or greater than 2 millions in any year, or if sales are negative in any year, or if tangible fixed assets are negative in any year.

3. Drop firm-year observations with missing, zero, or negative values for materials, operating revenue, and total assets.

4. Drop firm-year observations with missing information regarding their industry of activity.

5. These sums should not be extreme, so drop observations below 0.1 and above 99.9 percentile:
   - The sum of tangible fixed assets, intangible fixed assets, and other fixed assets as a ratio of total fixed assets.
   - The sum of stocks, debtors, and other current assets as a ratio of total current assets.
   - The sum of fixed assets and current assets as a ratio of total assets.
   - The sum of capital and other shareholder funds as a ratio of total shareholder funds.
   - The sum of long term debt and other non-current liabilities as a ratio of total non-current liabilities.
   - The sum of loans, creditors, and other current liabilities as a ratio of total current liabilities.
   - The sum of non-current liabilities, current liabilities, and shareholder funds as a ratio of the variable that reports the sum of shareholder funds and total liabilities.
Selection

- Select firms report financials, serious problem in US

- Selection might vary over time.

- Estimate logistic regressions to get propensity weights that will take care of selection into reporting financials in matched data set:

  **Example: Orbis-US Census (LOCUS)**

  **Continuers, Entrants, Exiters:**

  \[ R_{it} = \alpha + \gamma_1 \log(\text{emp}_{it}) + \gamma_2 \text{age}_{it} + \gamma_3 G_{it} + \text{ind}_i + \text{ifo}_{it} + \varepsilon_{it} \]
Addressing Selection: Age

- Decrease in the observable differences between reporting and non-reporting privately-held firms.
Decrease in the observable differences between reporting and non-reporting privately-held firms.
Coverage Statistics

If you cannot match to census and correct for selection, but argue your data is almost representative, then you need to report coverage statistics.
### Orbis Coverage Relative to Eurostat (Wage Bill)

<table>
<thead>
<tr>
<th>Year</th>
<th>Spain</th>
<th>Italy</th>
<th>Portugal</th>
<th>Germany</th>
<th>France</th>
<th>Norway</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>0.69</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.71</td>
<td>0.63</td>
<td></td>
<td></td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>0.73</td>
<td>0.62</td>
<td></td>
<td></td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>0.75</td>
<td>0.69</td>
<td></td>
<td></td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>0.74</td>
<td>0.68</td>
<td></td>
<td></td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>0.75</td>
<td>0.71</td>
<td></td>
<td></td>
<td>0.71</td>
<td>0.66</td>
</tr>
<tr>
<td>2005</td>
<td>0.74</td>
<td>0.72</td>
<td></td>
<td></td>
<td>0.71</td>
<td>0.67</td>
</tr>
<tr>
<td>2006</td>
<td>0.74</td>
<td>0.73</td>
<td>0.91</td>
<td>0.34</td>
<td>0.72</td>
<td>0.71</td>
</tr>
<tr>
<td>2007</td>
<td>0.74</td>
<td>0.73</td>
<td>0.94</td>
<td>0.34</td>
<td>0.73</td>
<td>0.73</td>
</tr>
<tr>
<td>2008</td>
<td>0.72</td>
<td>0.84</td>
<td>0.97</td>
<td>0.28</td>
<td>N/A</td>
<td>0.65</td>
</tr>
<tr>
<td>2009</td>
<td>0.72</td>
<td>0.81</td>
<td>0.96</td>
<td>0.28</td>
<td>0.71</td>
<td>0.85</td>
</tr>
<tr>
<td>2010</td>
<td>0.73</td>
<td>0.83</td>
<td>0.96</td>
<td>0.30</td>
<td>0.73</td>
<td>0.82</td>
</tr>
<tr>
<td>2011</td>
<td>0.74</td>
<td>0.86</td>
<td>0.97</td>
<td>0.78</td>
<td>0.75</td>
<td>0.82</td>
</tr>
<tr>
<td>2012</td>
<td>0.75</td>
<td>0.86</td>
<td>0.96</td>
<td>0.75</td>
<td>0.74</td>
<td>0.87</td>
</tr>
<tr>
<td>2013</td>
<td>0.74</td>
<td>0.85</td>
<td>0.97</td>
<td>0.76</td>
<td>0.74</td>
<td>0.88</td>
</tr>
<tr>
<td>2014</td>
<td>0.76</td>
<td>0.87</td>
<td>0.97</td>
<td>0.75</td>
<td>0.76</td>
<td>0.88</td>
</tr>
<tr>
<td>2015</td>
<td>0.76</td>
<td>0.88</td>
<td>0.98</td>
<td>0.79</td>
<td>0.77</td>
<td>0.89</td>
</tr>
</tbody>
</table>
## Orbis: Share of Total Wage Bill by Size Class

<table>
<thead>
<tr>
<th></th>
<th>Spain</th>
<th>Italy</th>
<th>Portugal</th>
<th>Germany</th>
<th>France</th>
<th>Norway</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ORBIS-AMADEUS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-19 employees</td>
<td>0.19</td>
<td>0.11</td>
<td>0.18</td>
<td>0.01</td>
<td>0.08</td>
<td>0.14</td>
</tr>
<tr>
<td>20-249 employees</td>
<td>0.47</td>
<td>0.53</td>
<td>0.50</td>
<td>0.33</td>
<td>0.30</td>
<td>0.43</td>
</tr>
<tr>
<td>250+ employees</td>
<td>0.34</td>
<td>0.36</td>
<td>0.32</td>
<td>0.67</td>
<td>0.61</td>
<td>0.43</td>
</tr>
<tr>
<td><strong>Eurostat (SBS)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-19 employees</td>
<td>0.20</td>
<td>0.22</td>
<td>0.21</td>
<td>0.07</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td>20-249 employees</td>
<td>0.43</td>
<td>0.44</td>
<td>0.49</td>
<td>0.26</td>
<td>0.31</td>
<td>0.41</td>
</tr>
<tr>
<td>250+ employees</td>
<td>0.37</td>
<td>0.34</td>
<td>0.30</td>
<td>0.67</td>
<td>0.55</td>
<td>0.44</td>
</tr>
</tbody>
</table>
IDENTIFICATION WITH FIXED EFFECTS IN MICRO DATA

Some Examples
Difference-in-Differences (DID) is the benchmark methodology:

**Advantages:** get at causality (if shock is exogenous); test the microfoundation directly

**Disadvantages:** Hard to quantify macro effect

\[
\log Y_{f,b,l,q} = \alpha_{f,b} + \alpha_{f,q} + \kappa(Bank_b \times \log \text{Shock}_{q-1}) + \delta_2 FX_{f,b,l,q} \\
+ \Theta_1 Bank_{b,q-1} + \vartheta_{f,b,l,q}
\]

- **Bank:** dummy based on any bank balance sheet variable
- **\( \alpha_{f,q} \):** firm×quarter fixed effects, which controls for unobserved time-varying firm characteristics (e.g., credit risk and credit demand)
Also possible to control for non-random matching of borrowers and lenders with detailed transaction level data:

$$\log Y_{f,b,l,q} = \phi_{f,b,q} + \beta_1 \text{Collateral}_{f,b,l,q} + \beta_2 (\text{Collateral}_{f,b,l,q} \times \log \text{Shock}_{q-1})$$
$$+ \beta_3 \text{FX}_{f,b,l,q} + e_{f,b,l,q}$$
Conclusion

Micro data on firms real and financial linkages can help to answer many macro questions with robust identification.

It takes a lot of effort, time and care to work with such data.
Data References:

Programs: http://econweb.umd.edu/kalemli/programs.html