Measuring Cross-Country Differences in Misallocation

Martin Rotemberg[⋆]
T. Kirk White[†]

[⋆]NYU
[†]Center for Economic Studies, U.S. Census Bureau

July 2017

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Disclaimer

The research in this presentation was conducted while the second author was an employee of Census Bureau. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

Within-Industry TFP Dispersion is Important

- Dispersion in firm outcomes is important for a lot of models
  - Determines responsiveness to a variety of shocks, such as trade liberalization (e.g., Melitz 2003)
  - Importance of management / R&D / investments (e.g., Bartelsman and Doms 2000)
  - Evidence of Misallocation (e.g. Hsieh Klenow 2009)
- What we're doing:
  - Dispersion vs measurement error
    - Focus on Hsieh Klenow misallocation measure
    - Use a new editing/imputation method – consistent cross-country comparisons

Measurement (1): Data

1. Census data (in most/all countries) tends to be self-reported
   1.1 US Census Bureau does a lot of editing and imputation of data (and pushes forward the frontier of knowledge on these topics).
      1.1.1 Other countries (especially developing countries) do not do this
2. Two (potential) major sources of changes to raw data
   2.1 Edits/imputation to raw data. Fill in missing or faulty data using imputation models and other survey data
   2.2 Linking to administrative records, e.g. access to tax records for payroll

Measurement (2): Theory (Intuition)

- Productivity growth from reallocation: reallocate inputs from plants with low marginal products to those with high ones
  - Hsieh and Klenow (2009): plants with large (small) distortions have high (low) marginal products
    - Remove distortions —> markets reallocation resources –> get aggregate TFP growth
- Using the Hsieh and Klenow model to quantify misallocation, we focus on the role of measurement:
  - How much does data cleaning affect measured misallocation (and thus measured potential for TFP growth from reallocation)?

Main results

- Census Bureau's data cleaning has an enormous effect on HK measure of within-industry misallocation:
  - *Five times* more measured misallocation in raw vs. cleaned U.S. data.
- Cross-country differences in data cleaning have a large effect:
  -
    -
    -
    -

Main results

- Census Bureau's data cleaning has an enormous effect on HK measure of within-industry misallocation:
  - *Five times* more measured misallocation in raw vs. cleaned U.S. data.
- Cross-country differences in data cleaning have a large effect:
  - E.g, set misallocation in India to same level in U.S.:
    - ▸
    - ▸
    - ▸

Main results

- Census Bureau's data cleaning has an enormous effect on HK measure of within-industry misallocation:
  - *Five times* more measured misallocation in raw vs. cleaned U.S. data.
- Cross-country differences in data cleaning have a large effect:
  - E.g, set misallocation in India to same level in U.S.:
    - Raw Indian data vs. Raw U.S. data: Indian TFP would *decrease* by about 56%

Main results

- Census Bureau's data cleaning has an enormous effect on HK measure of within-industry misallocation:
  - *Five times* more measured misallocation in raw vs. cleaned U.S. data.
- Cross-country differences in data cleaning have a large effect:
  - E.g, set misallocation in India to same level in U.S.:
    - Raw Indian data vs. Raw U.S. data: Indian TFP would *decrease* by about 56%
    - Raw Indian data vs. US *Census-cleaned* data: Indian TFP would *increase* by 32%
    -

Main results

- Census Bureau's data cleaning has an enormous effect on HK measure of within-industry misallocation:
  - *Five times* more measured misallocation in raw vs. cleaned U.S. data.
- Cross-country differences in data cleaning have a large effect:
  - E.g, set misallocation in India to same level in U.S.:
    - Raw Indian data vs. Raw U.S. data: Indian TFP would *decrease* by about 56%
    - Raw Indian data vs. US *Census-cleaned* data: Indian TFP would *increase* by 32%
    - Use common cleaning method in both countries: little difference in measured misallocation.

Outline

Hsieh and Klenow Set-Up

- Each intermediate good producer i producing in sector s has Cobb-Douglas production function
- Each producer faces idiosyncratic distortions on their prices of capital $\left(\tau_{k_i}\right)$ and output $\left(\tau_{Y_i}\right)$
- Producers face CES demand

Distortions

- HK insight: in the model with no distortions, $TFPR_{si} = \overline{TFPR}_s$
- Can measure the distortions from observed within-industry $TFPR_{si}$ dispersion
- Given the assumed CES demand structure, can back out $TFPQ_{si}$ from measured $TFPR_{si}$
- HK derive expression for aggregate TFP losses from misallocation (due to within-industry distortions)

Misallocation vs. Measurement Error

$$\frac{(1+\tau_{Ksi})}{(1-\tau_{Ysi})} = \frac{\alpha_s Y_{si}}{K_{si}}$$

- What could lead to $\frac{(1+\tau_{Ksi})}{(1-\tau_{Ysi})} \neq 1$?
  - Actual distortions: markups, taxes, different interest rates, ...
  - Measurement error.
    - Firm has undistorted optimal capital/labor, but reports the wrong thing
    - Firm reports optimal capital/labor, but Census edits change reported values
    - Firm doesn't report fully and Census imputes its values

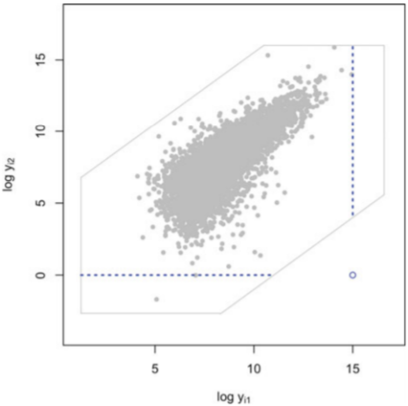
Outline

1. Static Misallocation
2. Editing in the US
3. Imputation in the US
4. Data Cleaning
5. Wrap-Up

Important Types of Editing in US Census of Manufactures

- Logical edits – Example: TVS
- Units errors (a.k.a. "Rounding" Edits)
- Analyst corrections
- Check against administrative records
- Ratio edits
  - based on specific industry knowledge, IQR, etc.

Combination of Edit Rules results in Feasible Region $\mathcal{D}$

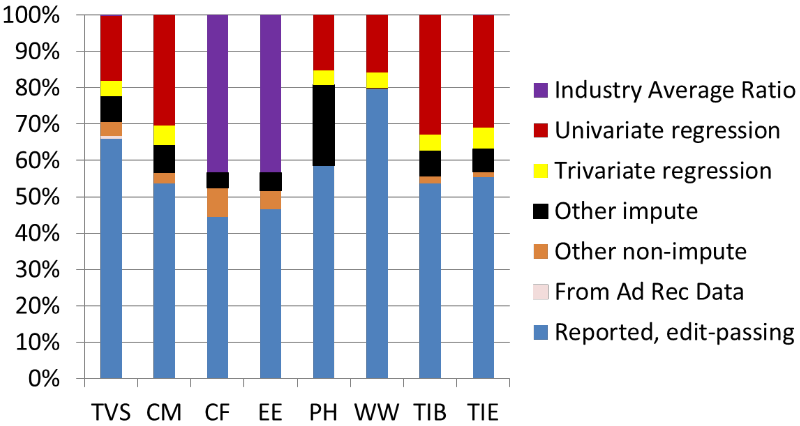


Outline

1. Static Misallocation
2. Editing in the US
3. Imputation in the US
4. Data Cleaning
5. Wrap-Up

Frequencies of Editing/Imputation

2007 Census of Manufactures. Note: Swiss Cheese Missingness



Quantifying effect of editing on HK measure of gains from reallocation

U.S. Census of Manufactures

Data	Trimming 1%
US Census-Cleaned	62%
US Raw	
India Raw	

Quantifying effect of editing on HK measure of gains from reallocation

U.S. Census of Manufactures

Data	Trimming 1%
US Census-Cleaned	62%
US Raw	371%
India Raw	

Quantifying effect of editing on HK measure of gains from reallocation

U.S. Census of Manufactures

Data	Trimming 1%
US Census-Cleaned	62%
US Raw	371%
India Raw	91%

Quantifying effect of editing on HK measure of gains from reallocation

U.S. Census of Manufactures

Data	Trimming		
	0%	1%	2%
US Census-Cleaned		62%	43%
US Raw		371%	263%
India Raw		91%	76%

Quantifying effect of editing on HK measure of gains from reallocation

U.S. Census of Manufactures

Data	Trimming		
	0%	1%	2%
US Census-Cleaned	165%	62%	43%
US Raw	4293%	371%	263%
India Raw	147%	91%	76%

Quantifying effect of editing on HK measure of gains from reallocation

U.S. Census of Manufactures

Data	Trimming 0%	1%	2%
US Census-Cleaned	165%	62%	43%
US Raw	4293%	371%	263%
India Raw	147%	91%	76%

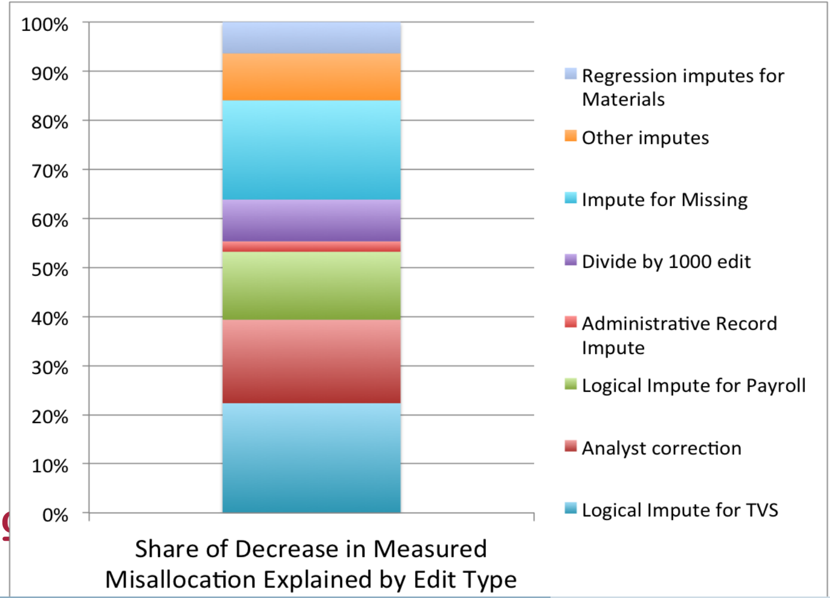
- Takeaway: US Census Bureau data cleaning has HUGE effect on measured misallocation!
- What would measured misallocation be in Indian data if we cleaned it using the same methods?

From 371% to 62%

- For cross-country comparisons, we would like to use same data cleaning methods as in U.S.
  - Problem: Census has an entire staff cleaning the data for months
  - Can we replicate just the important parts of what U.S. Census Bureau does?
  - Which U.S. Census edits have big impact on measured misallocation?

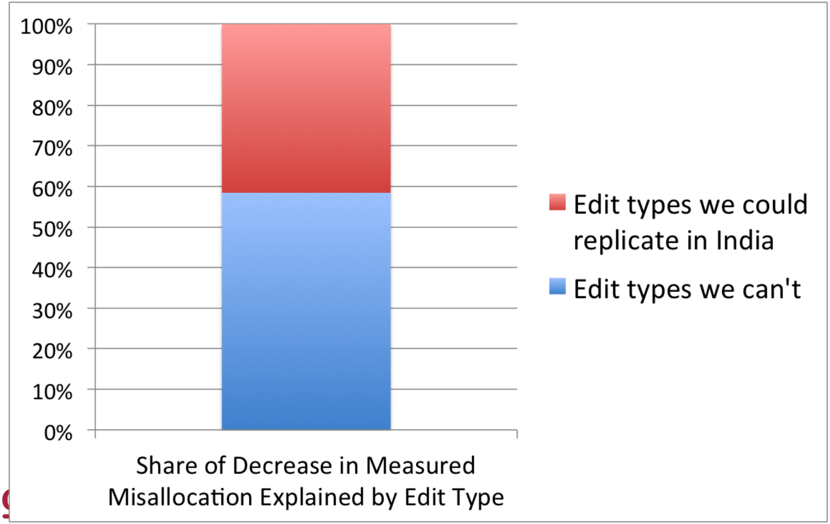
Effect of Census Bureau Edits (Shapley Shares)

on Measured Misallocation in U.S. data, 1% trimming



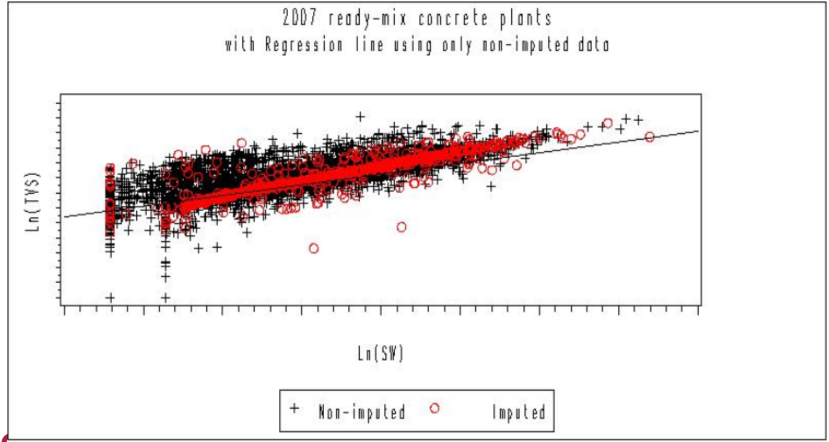
Effect of Census Bureau Edits (Shapley Shares)

on Measured Misallocation in U.S. data, 1% trimming



Census Bureau imputation methods are not designed for microdata research

From White et al. 2015



Outline

1. Static Misallocation
2. Editing in the US
3. Imputation in the US
4. Data Cleaning
5. Wrap-Up

Motivation for doing our own data cleaning

- For cross-country comparison of misallocation want to clean firm-level data in India like the U.S. data
- Problem:
  - Not feasible for us to replicate US Census Bureau’s data cleaning in India
- So...try a fully data-driven approach, from Kim et al. (2015, *JASA*)

Kim et al. (2015): Simultaneous Editing and Imputation

- Build statistical models for:
  - Unobserved true firm behavior – based on edit-rule-passing data
  - Indicators for which variable(s) is/are in error when a record fails an edit-rule
  - Reporting error

Advantages of Kim et al. (2015) Method

- Imputation model approximates the joint distribution of the edit-rule-passing data
- Imputes automatically satisfy all the edit rules
- Can estimate uncertainty of misallocation estimates due to editing/imputation (although we don't do this yet)
- Allows us to do cross-country comparisons using a common data cleaning method

Common Data Cleaning for US and India

- Based on the reported data and the edit rules, for edit-failing records, impute final values that
  - Are likely under the model for reporting error
  - Are likely under the model for error indicators
  - Are likely under the model for the underlying data
  - Satisfy all the edit rules
- We apply this method to clean the the raw data for India and the US for every manufacturing industry

New measures of misallocation, Commonly Cleaned Data

	Measured Misallocation for Our Cleaned Data		
	Trimming		
Country	0%	1%	2%
U.S.	65%		
India	63%		

New measures of misallocation, Commonly Cleaned Data

	Measured Misallocation for Our Cleaned Data		
	Trimming		
Country	0%	1%	2%
U.S.	65%	48%	
India	63%	58%	

New measures of misallocation, Commonly Cleaned Data

	Measured Misallocation for Our Cleaned Data		
	Trimming		
Country	0%	1%	2%
U.S.	65%	48%	40%
India	63%	58%	53%

Outline

1. Static Misallocation
2. Editing in the US
3. Imputation in the US
4. Data Cleaning
5. Wrap-Up

Conclusions

- Data cleaning done by Census Bureau has huge effect on measured misallocation in Census of Manufactures
- Cross-country differences in data cleaning by statistical agencies also have huge effect on cross-country comparisons of measured misallocation
- When we apply the same data cleaning methods to raw manufacturing data from both India and the US, we find little evidence that misallocation is significantly higher in India than in the US