Research with private sector business microdata: The case of NETS/D&B

July 13, 2020

NBER Summer Institute - Entrepreneurship

Leland D. Crane
Ryan A. Decker
Federal Reserve Board

This presentation draws in part on joint work with Keith Barnatchez, without implication.
Opinions expressed herein are those of the authors alone and do not necessarily reflect the views of the Federal Reserve System or its staff.

The need for business microdata

- Demand for business microdata
  - Entrepreneurship
  - Innovation
  - Job flows
  - Monopoly/monopsony/market power
  - Trade
  - Urban economics and geography of firms
  - Business “dynamism”
- Census/BLS microdata are costly to access and use

A potential alternative: NETS/D&B

- **Dun & Bradstreet (D&B):** collects (or imputes…) establishment-level data on employment, sales, industry, and location for marketing, credit rating, and business intelligence purposes

A potential alternative: NETS/D&B

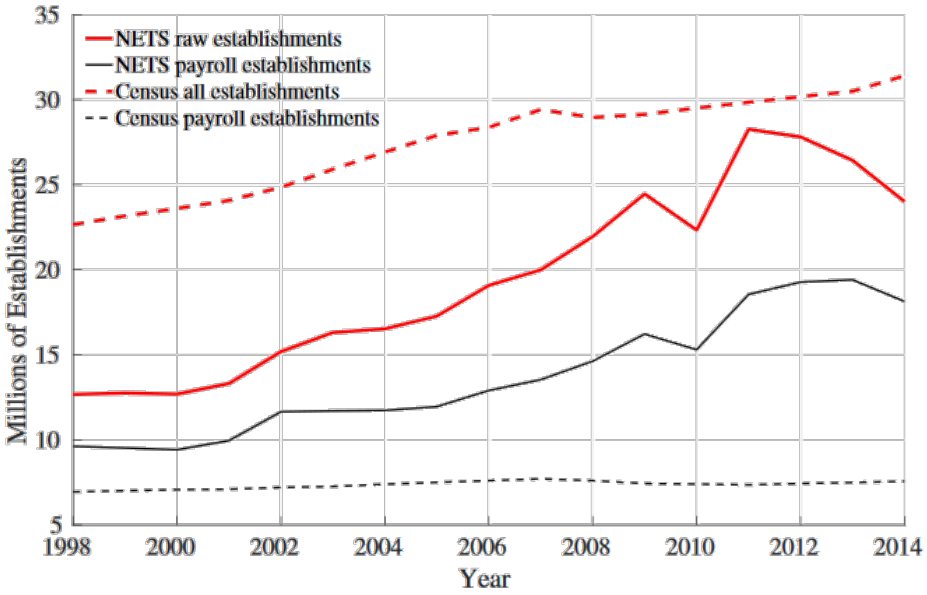
- **Dun & Bradstreet (D&B):** collects (or imputes…) establishment-level data on employment, sales, industry, and location for marketing, credit rating, and business intelligence purposes
- **National Establishment Time Series (NETS):** Walls & Associates: creates longitudinal (establishment) links, further imputation
  - Longitudinal establishment ID (dunsnumber)
  - Longitudinal firm ID (hqduns)

A potential alternative: NETS/D&B

- **Dun & Bradstreet (D&B):** collects (or imputes…) establishment-level data on employment, sales, industry, and location for marketing, credit rating, and business intelligence purposes
- **National Establishment Time Series (NETS):** Walls & Associates: creates longitudinal (establishment) links, further imputation
  - Longitudinal establishment ID (dunsnumber)
  - Longitudinal firm ID (hqduns)
- **What we do**
  - Compare NETS to Census Bureau data
  - Offer concrete suggestions for researchers

1. The NETS/D&B universe

- Two business universe concepts
  - Employers (have formal W-2 employees)
  - Nonemployers (business entity with no formal employees)

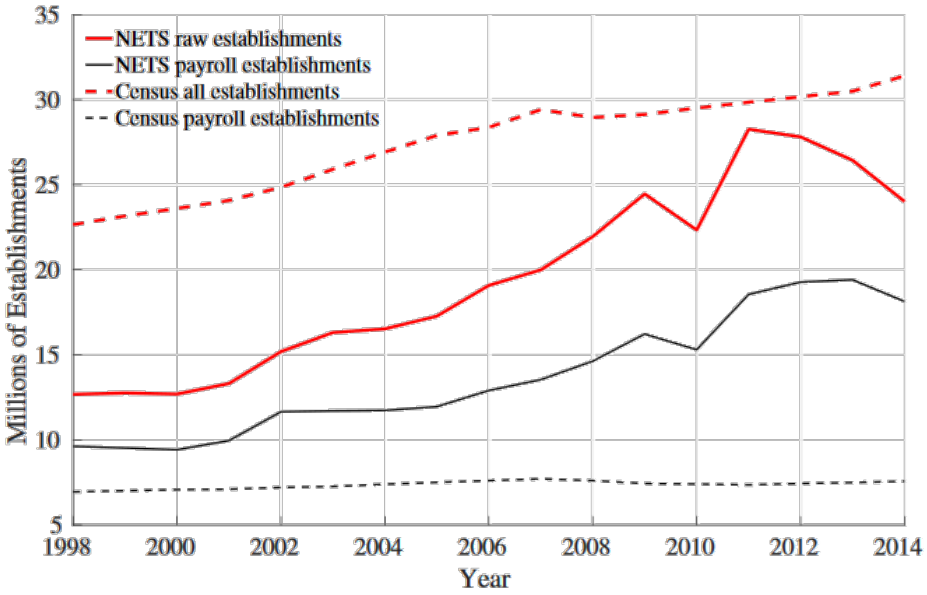


Source: NETS database, County Business Patterns, Census Nonemployer Statistics.
Note: NETS sample restricted to CBP industry scope.

NETS "payroll establishments": subtract one employee from each firm.

1. The NETS/D&B universe

- Two business universe concepts
  - Employers (have formal W-2 employees)
  - Nonemployers (business entity with no formal employees)
- Too many establishments to be the employer universe

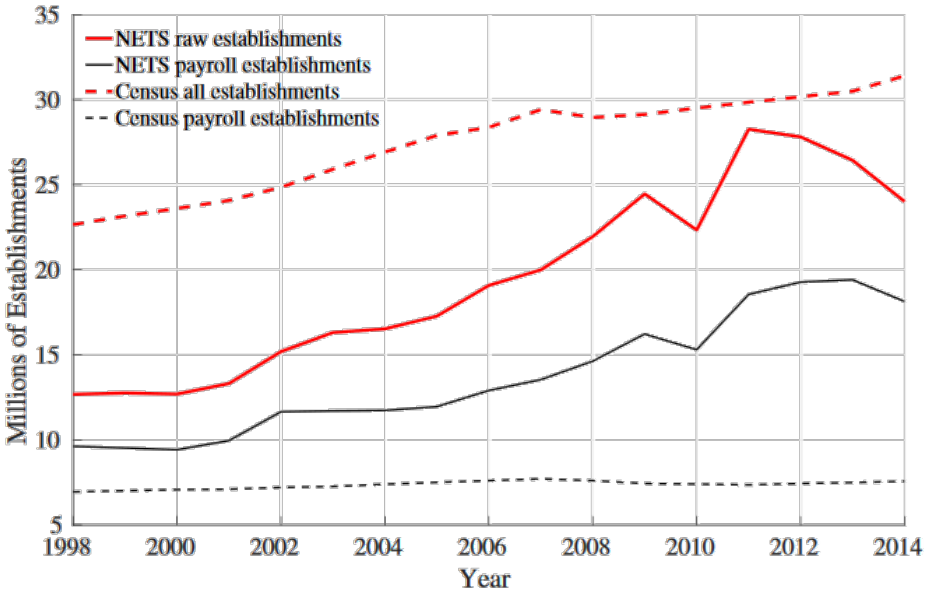


Source: NETS database, County Business Patterns, Census Nonemployer Statistics.
Note: NETS sample restricted to CBP industry scope.

NETS "payroll establishments": subtract one employee from each firm.

1. The NETS/D&B universe

- Two business universe concepts
  - Employers (have formal W-2 employees)
  - Nonemployers (business entity with no formal employees)
- Too many establishments to be the employer universe
- Too few establishments to be the total business universe

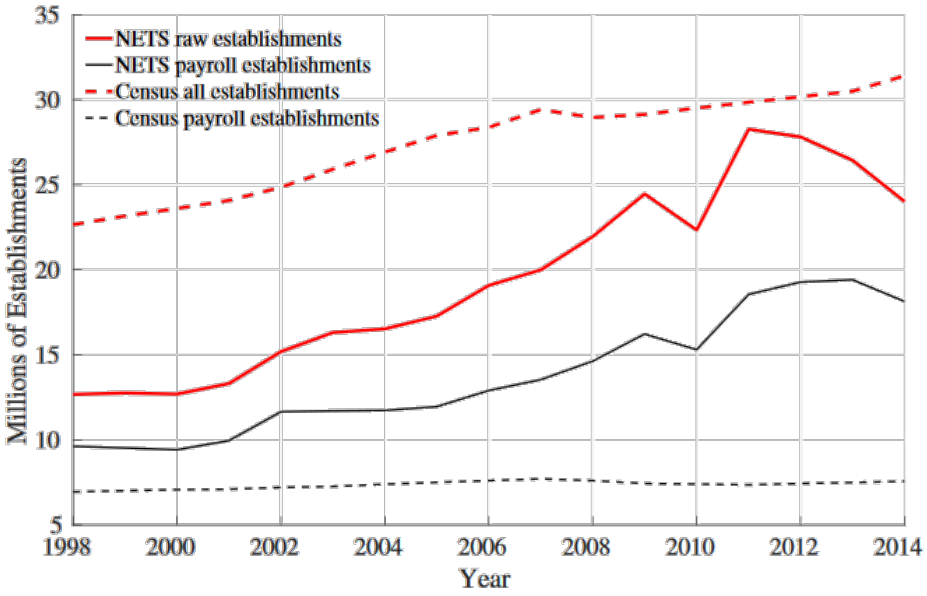


Source: NETS database, County Business Patterns, Census Nonemployer Statistics.
Note: NETS sample restricted to CBP industry scope.

NETS "payroll establishments": subtract one employee from each firm.

1. The NETS/D&B universe

- Two business universe concepts
  - Employers (have formal W-2 employees)
  - Nonemployers (business entity with no formal employees)
- Too many establishments to be the employer universe
- Too few establishments to be the total business universe



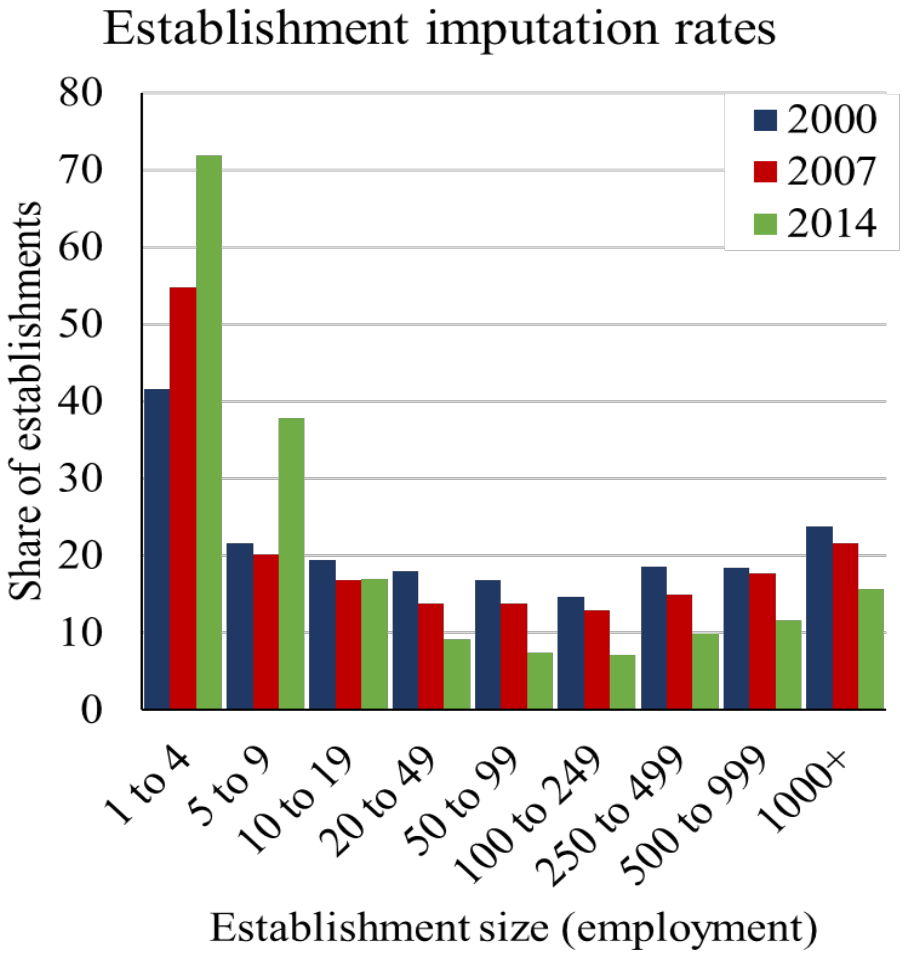
Source: NETS database, County Business Patterns, Census Nonemployer Statistics.
Note: NETS sample restricted to CBP industry scope.

NETS "payroll establishments": subtract one employee from each firm.

Suggestion 1: Use NETS/D&B as a supplement to other data, not as the definition of the business universe (e.g.: Ma, Murfin, Pratt 2020)

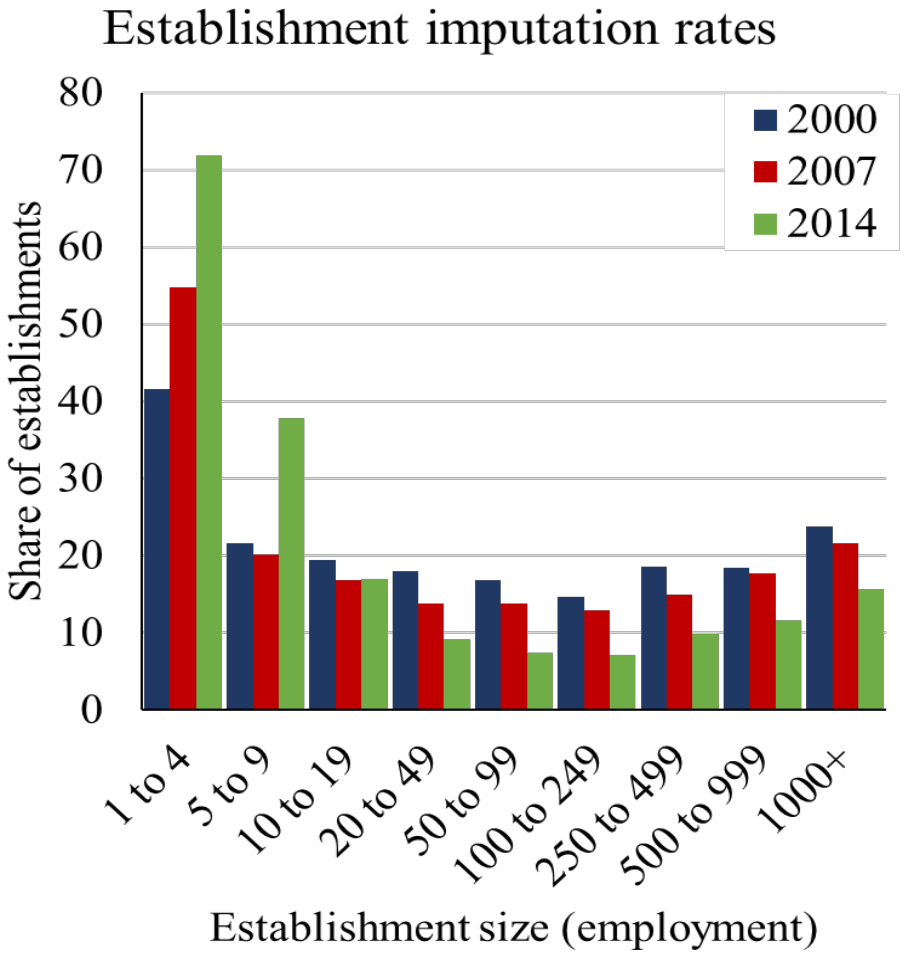
2. Employment measurement

- Imputation
  - 2/3 of smallest establishments
  - Half of firms



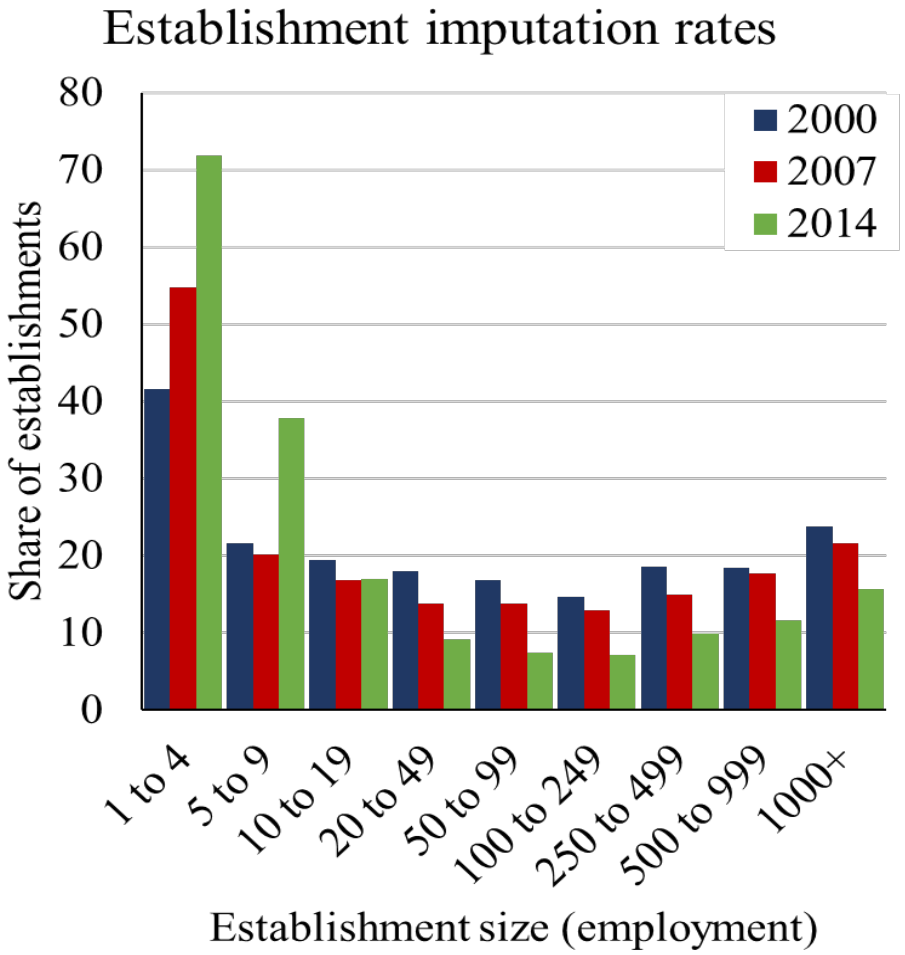
2. Employment measurement

- Imputation
  - 2/3 of smallest establishments
  - Half of firms
- Consecutive imputation
  - Median firm: 2 years
  - Top 10%: 7 years
  - 25% of employment at firms imputed 20+ years



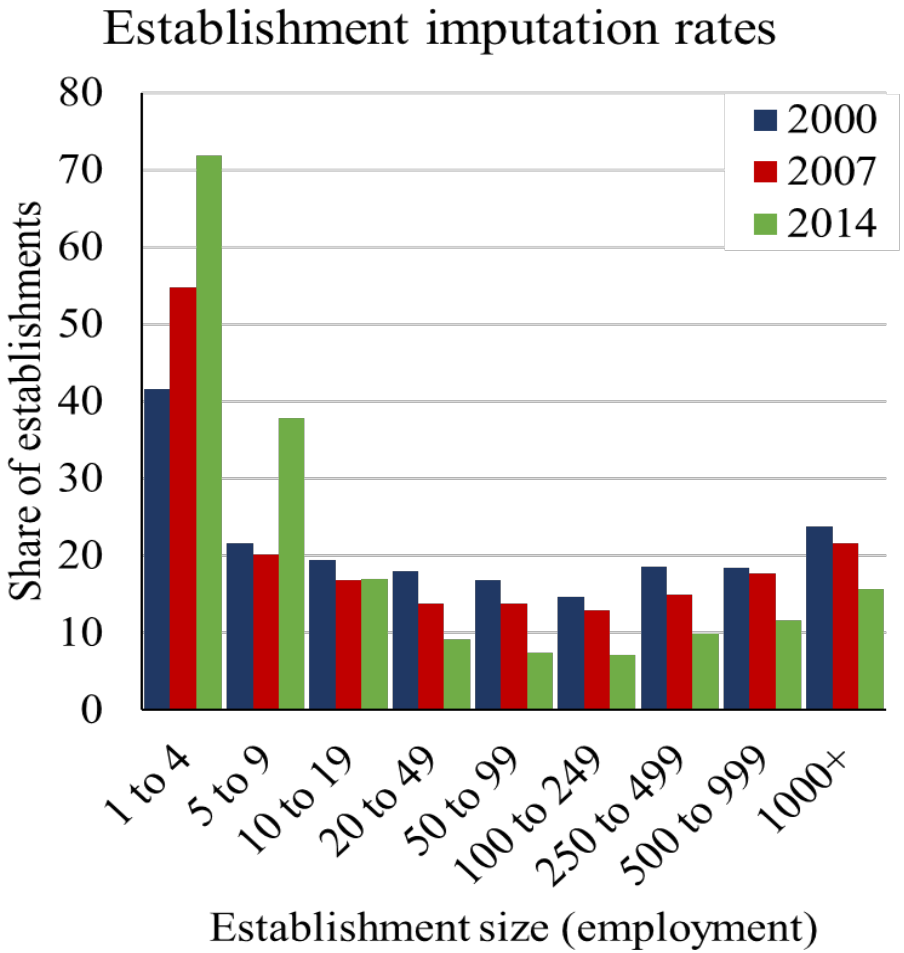
2. Employment measurement

- Imputation
  - 2/3 of smallest establishments
  - Half of firms
- Consecutive imputation
  - Median firm: 2 years
  - Top 10%: 7 years
  - 25% of employment at firms imputed 20+ years
- Rounding:
  - Nearest 5
  - Nearest 50
- Seasonality



2. Employment measurement

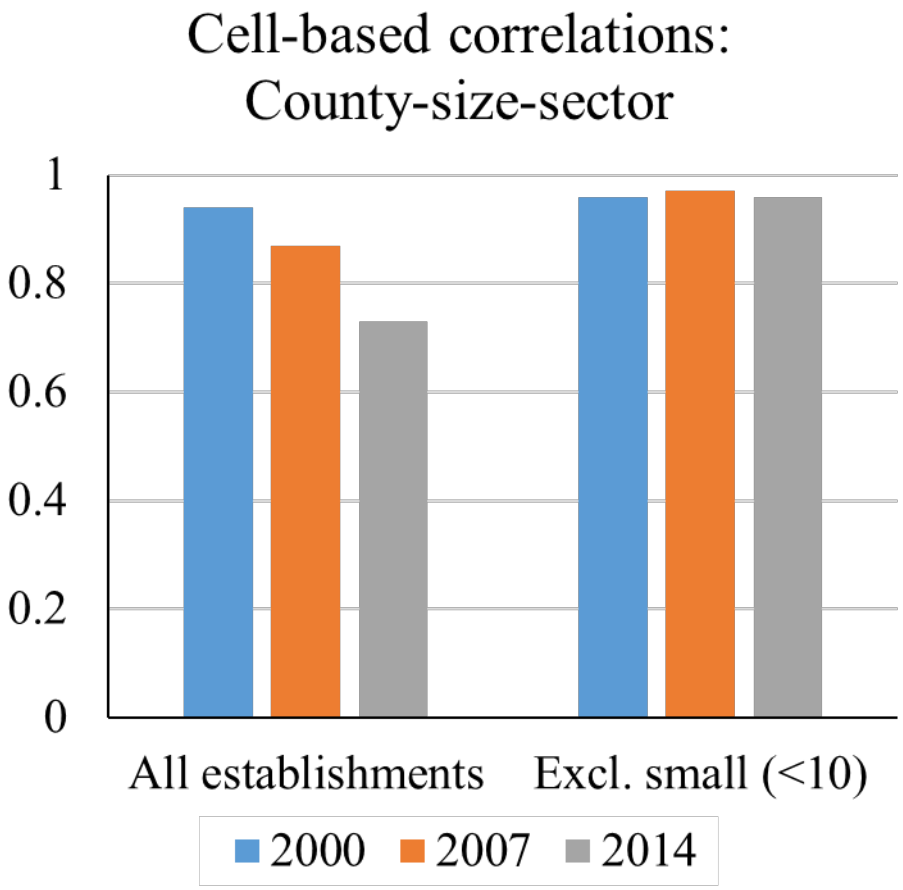
- Imputation
  - 2/3 of smallest establishments
  - Half of firms
- Consecutive imputation
  - Median firm: 2 years
  - Top 10%: 7 years
  - 25% of employment at firms imputed 20+ years
- Rounding:
  - Nearest 5
  - Nearest 50
- Seasonality



Suggestion 2: Measure employment in "bins" rather than relying on precise employment values, be transparent about imputation, and consider measurement error

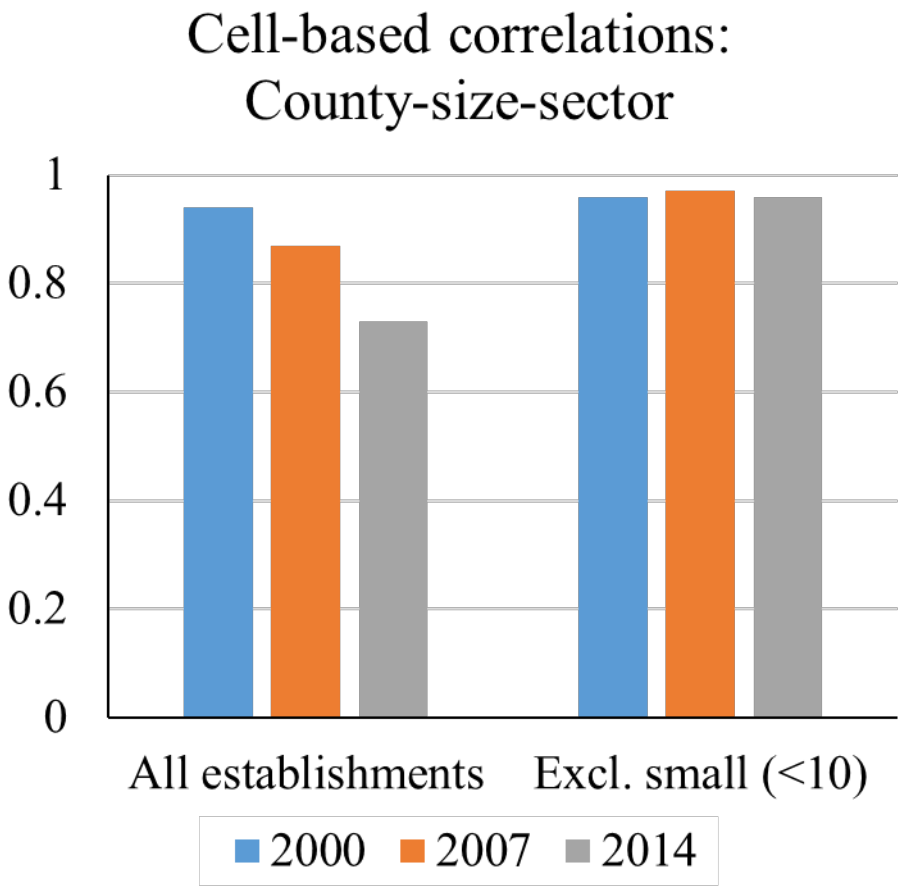
3. Cross-sectional correlations

- Geography-size-sector cells in NETS and CBP
- Establishment count correlations above 0.95 when small estabs omitted
  - Employment correlations 0.8-0.9
- Side note: NETS misses large part of 2000s manufacturing decline, shale oil boom



3. Cross-sectional correlations

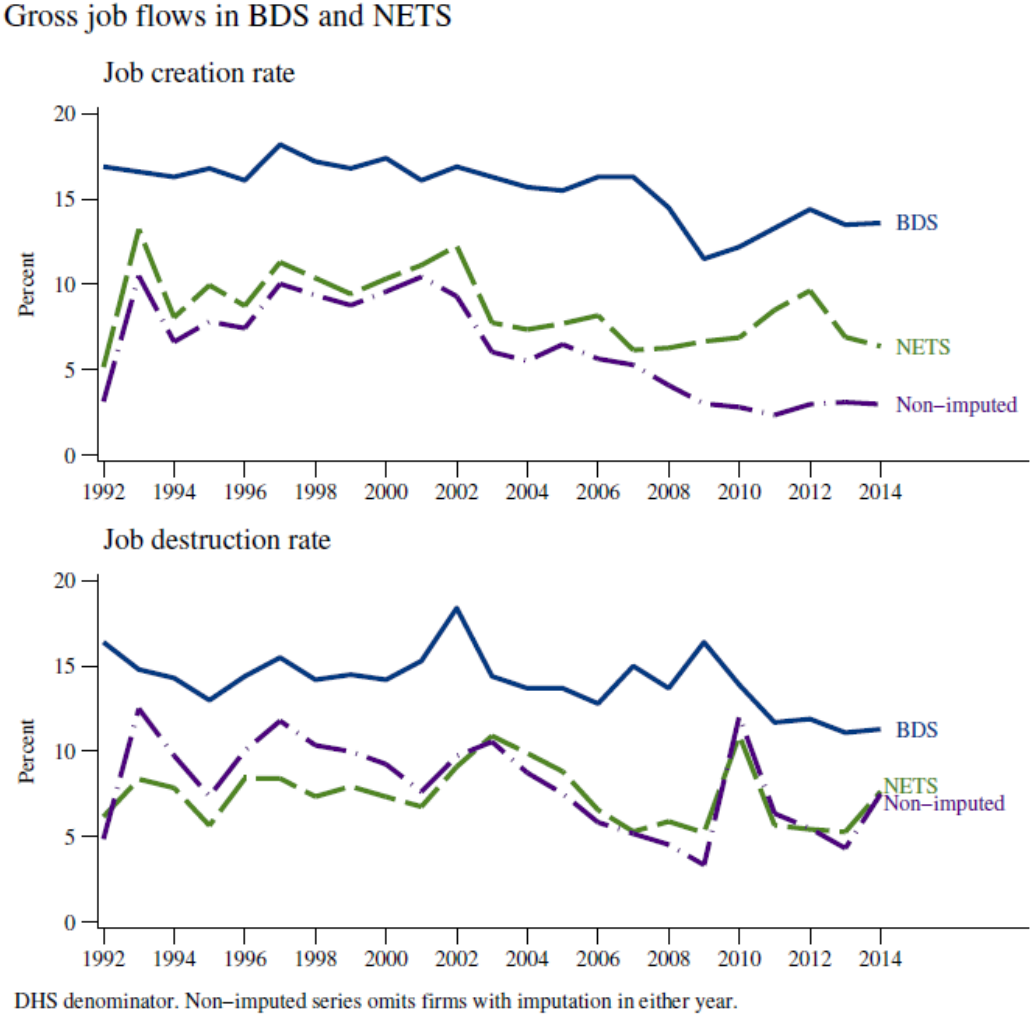
- Geography-size-sector cells in NETS and CBP
- Establishment count correlations above 0.95 when small estabs omitted
  - Employment correlations 0.8-0.9
- Side note: NETS misses large part of 2000s manufacturing decline, shale oil boom



Suggestion 3: Focus on cross-sectional moments, and omit small establishments

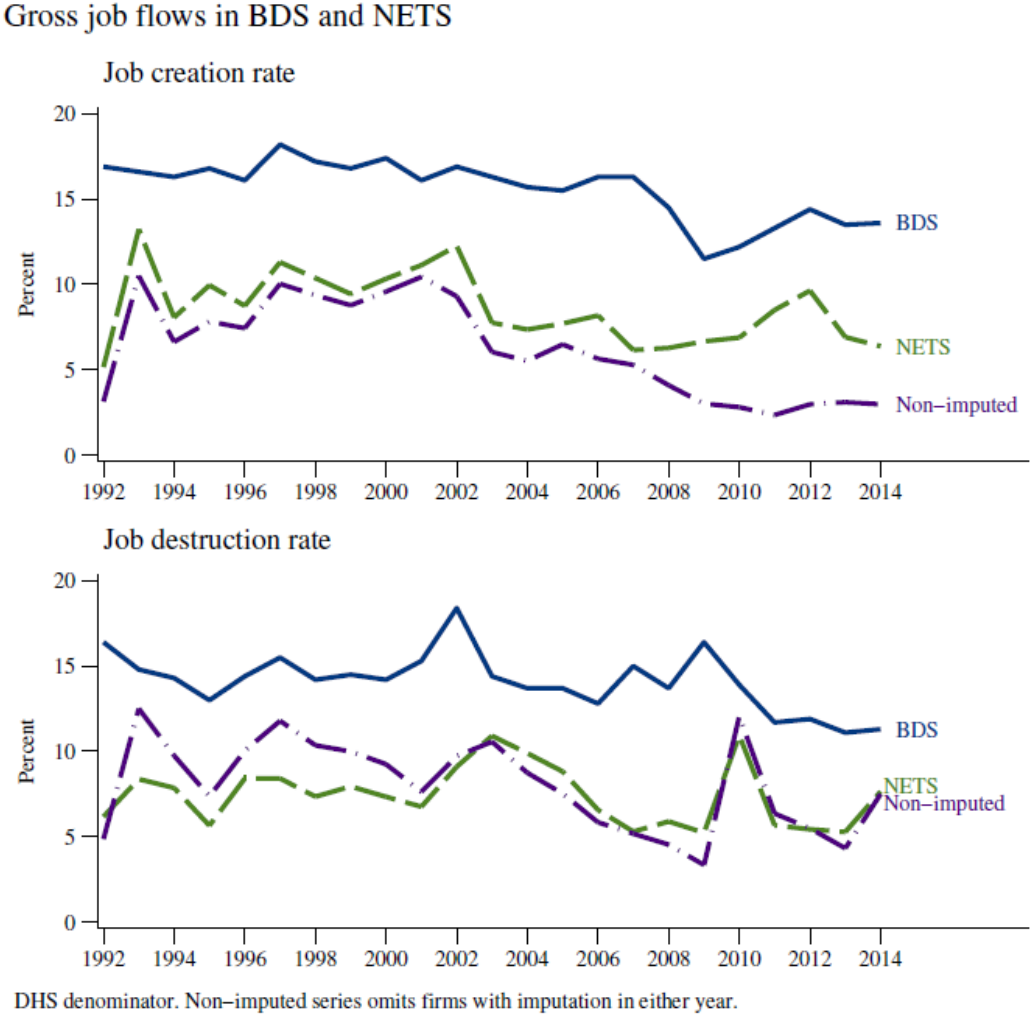
4. Muted business dynamics

- Not enough volatility
- Imputation not the problem



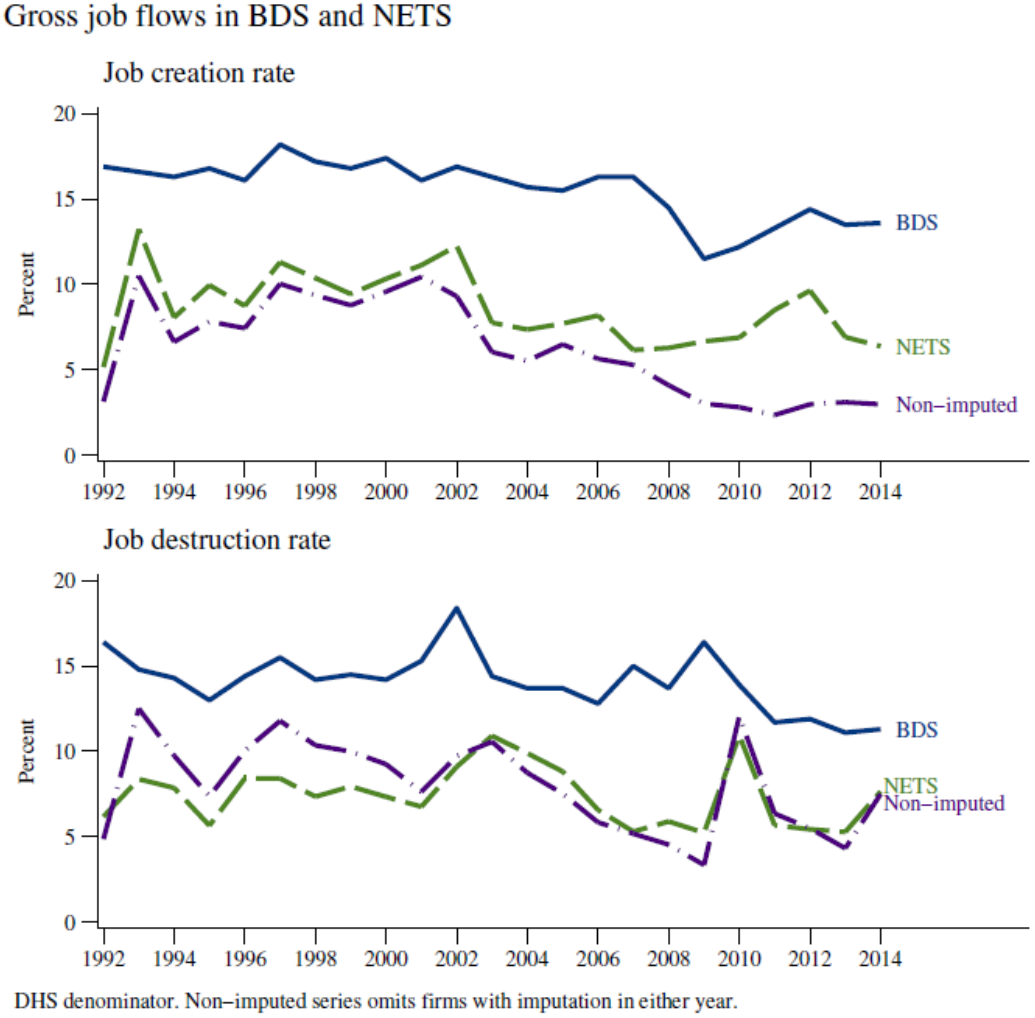
4. Muted business dynamics

- Not enough volatility
- Imputation not the problem
- Low cell-based correlations between job flow rates



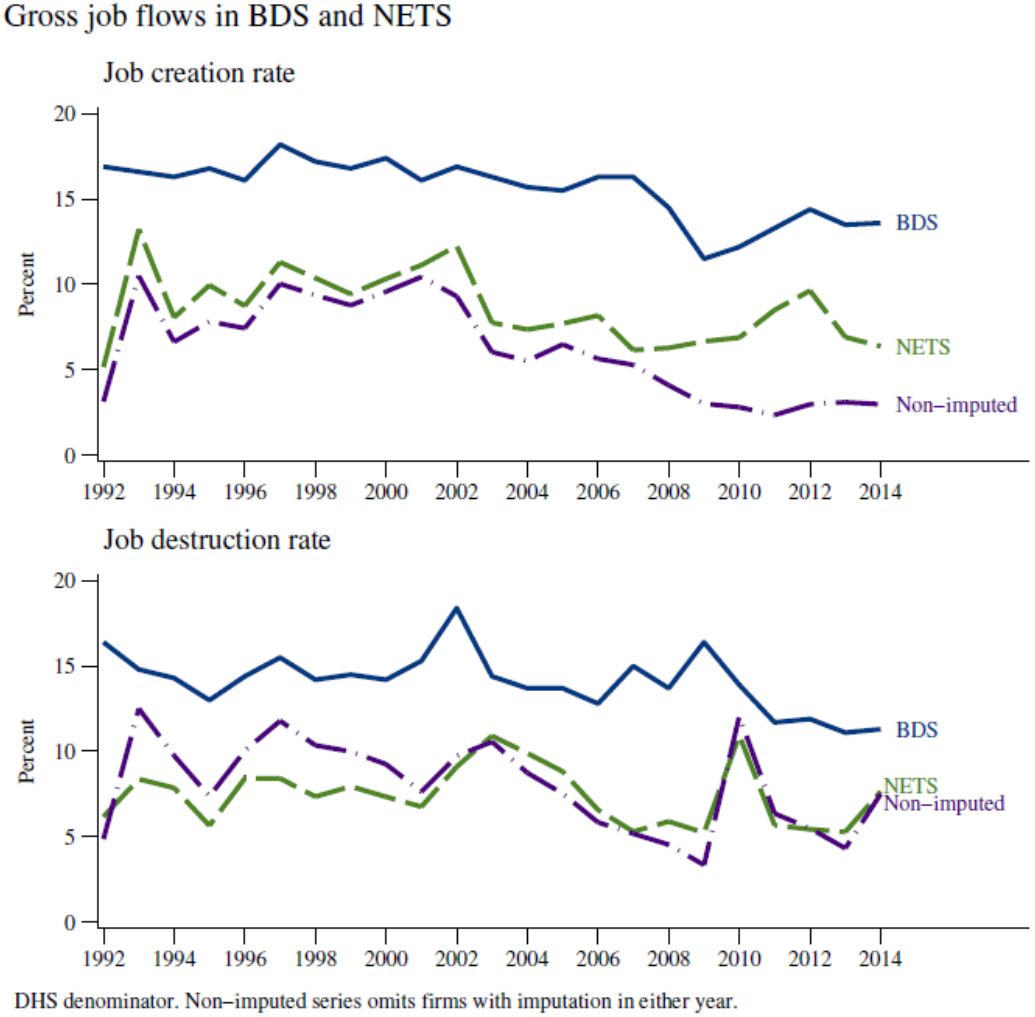
4. Muted business dynamics

- Not enough volatility
- Imputation not the problem
- Low cell-based correlations between job flow rates
- Exit rates off too
  - Falling exit in Great Recession



4. Muted business dynamics

- Not enough volatility
- Imputation not the problem
- Low cell-based correlations between job flow rates
- Exit rates off too
  - Falling exit in Great Recession



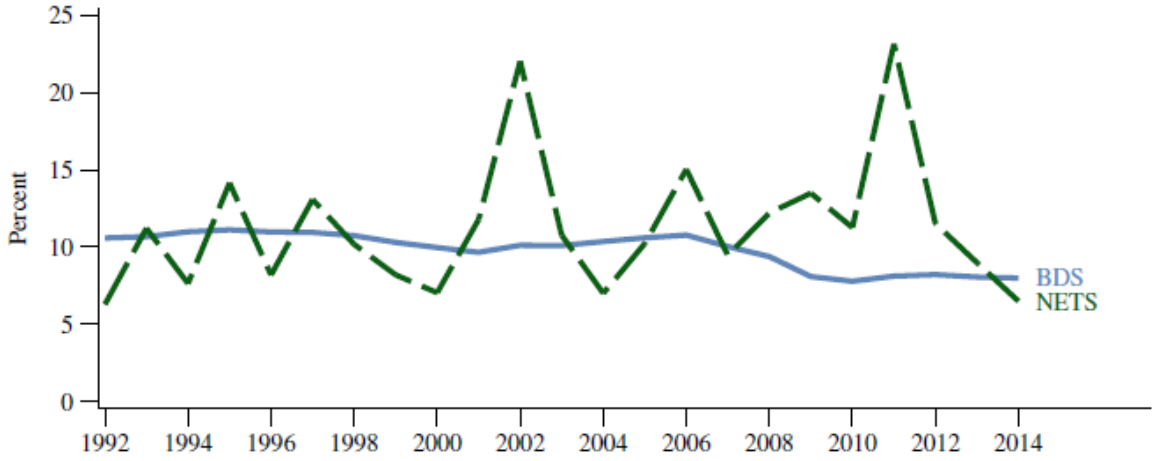
Suggestion: Be very cautious in studying business dynamics: firm/establishment growth, job flows, etc. Data are "stale" and insufficiently volatile.

5. Young firm mismeasurement

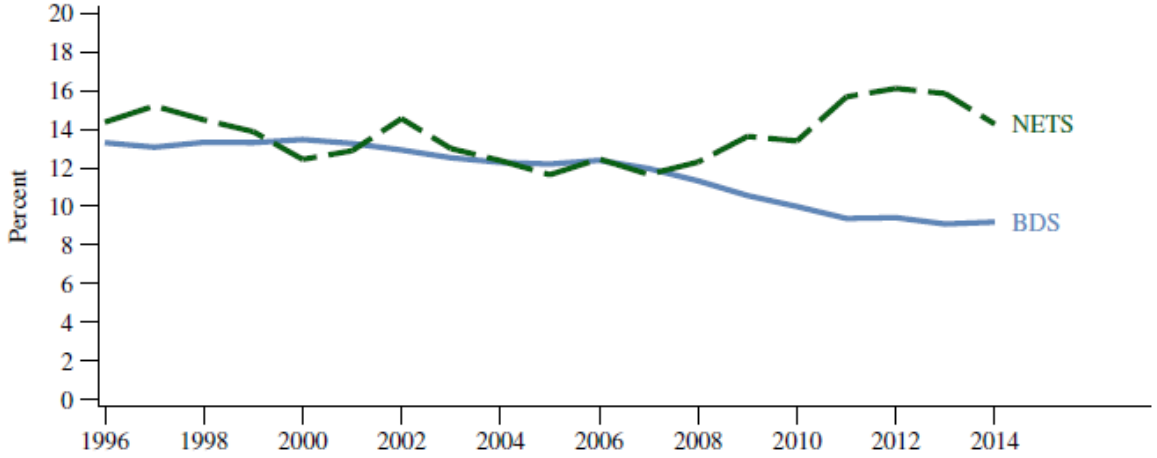
- Surge of spurious entry in 2000s
- No “declining dynamism”

Entrants and young firms in BDS and NETS

a. Entrants as share of all firms



b. Young firm employment as share of total



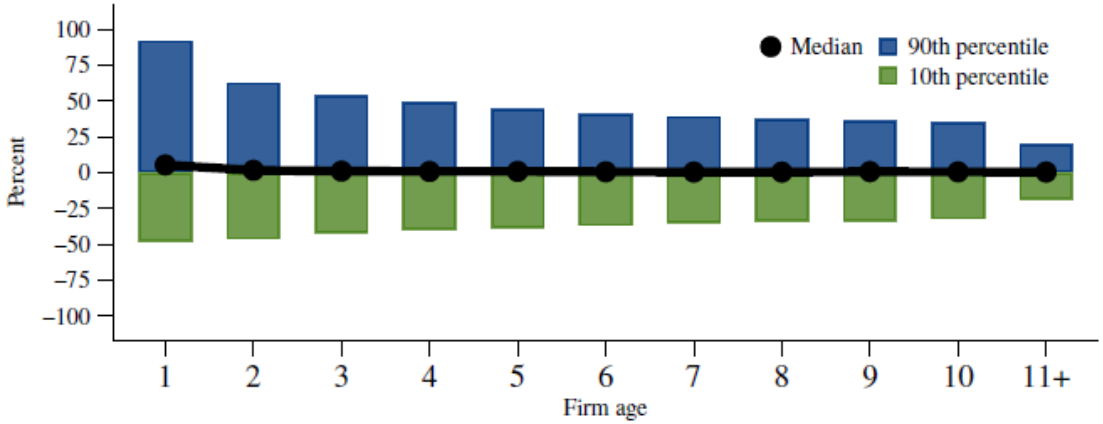
Entrants have age zero; young firms have age less than five.

Growth distribution over the lifecycle

- Key young firm traits: Skewness & dispersion

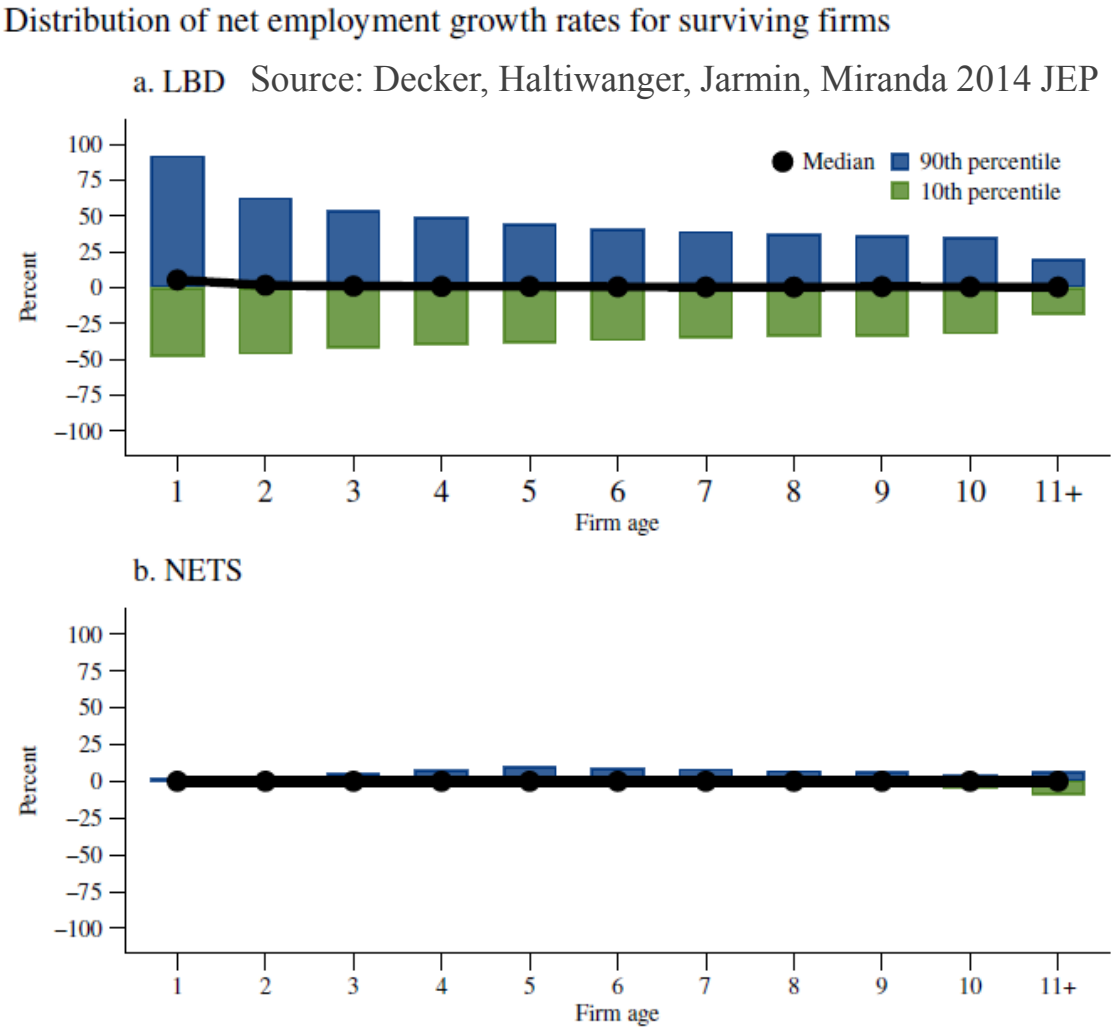
Distribution of net employment growth rates for surviving firms

a. LBD Source: Decker, Haltiwanger, Jarmin, Miranda 2014 JEP



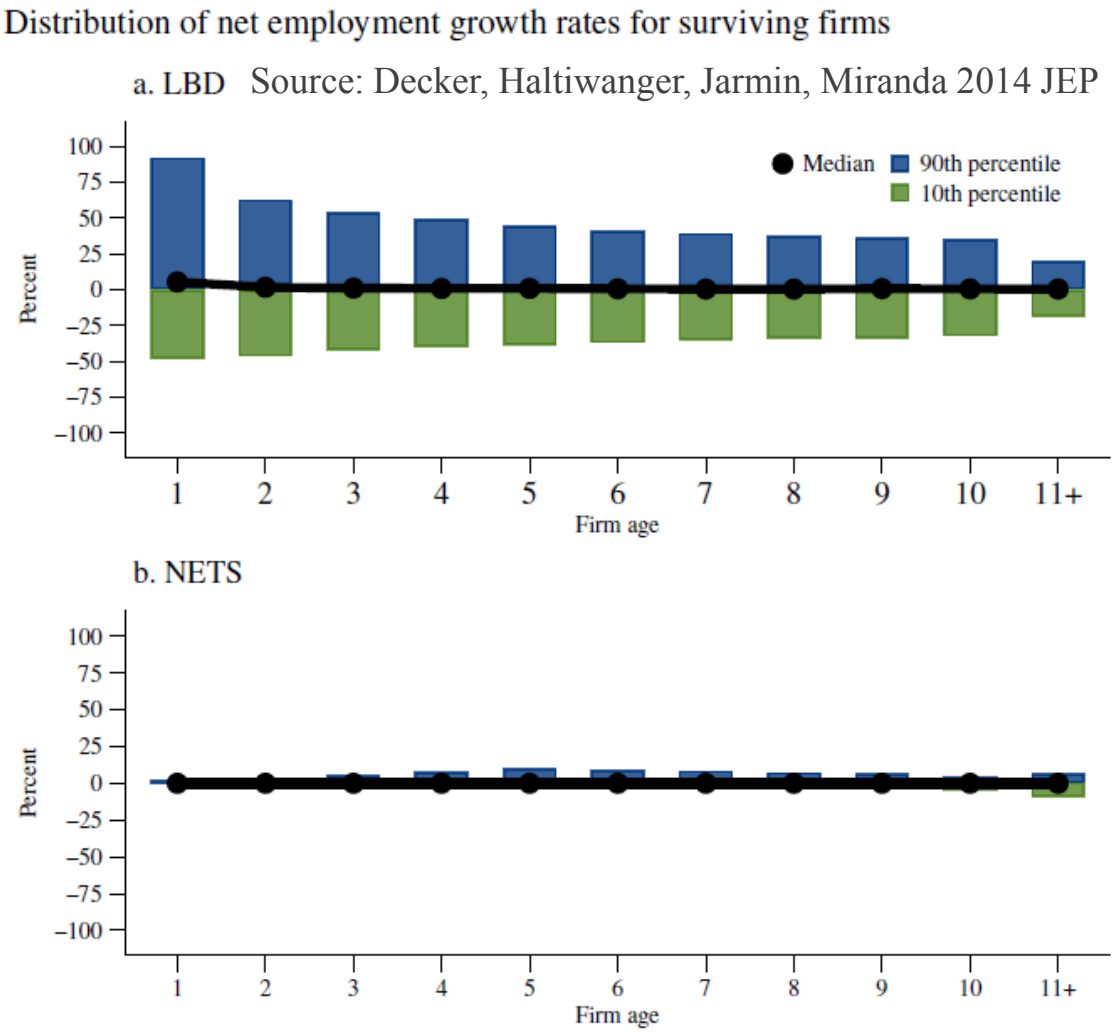
Growth distribution over the lifecycle

- Key young firm traits: Skewness & dispersion
- Also: “Up or out” dynamics muted in NETS



Growth distribution over the lifecycle

- Key young firm traits: Skewness & dispersion
- Also: “Up or out” dynamics muted in NETS



Suggestion 5: Be very cautious in studying young firms—key attributes of young firm lifecycle are missing from NETS

6. Imputed sales data

- Sales data are almost entirely imputed based on employment
  - 90% of Walmart establishments have identical sales per worker
- Counterfactually low dispersion of sales/worker

Imputation rates

Firm size (employees)	Year 2000	2014
1 to 4	80	80
5 to 9	78	85
10 to 19	77	82
20 to 49	79	84
50 to 99	85	88
100 to 249	89	91
250 to 499	93	94
500 to 999	94	94
1,000 to 2,499	93	93
2,500 to 4,999	95	92
5,000 to 9,999	95	94
10,000+	96	94

Source: NETS
Notes: Percent of firms with imputed establishment sales data.

6. Imputed sales data

- Sales data are almost entirely imputed based on employment
  - 90% of Walmart establishments have identical sales per worker
- Counterfactually low dispersion of sales/worker

Imputation rates

Firm size (employees)	Year 2000	2014
1 to 4	80	80
5 to 9	78	85
10 to 19	77	82
20 to 49	79	84
50 to 99	85	88
100 to 249	89	91
250 to 499	93	94
500 to 999	94	94
1,000 to 2,499	93	93
2,500 to 4,999	95	92
5,000 to 9,999	95	94
10,000+	96	94

Source: NETS
Notes: Percent of firms with imputed establishment sales data.

Suggestion 6: Ignore the sales data

Conclusion

- NETS may be useful for **static/cross sectional analysis**
  - Business location, industry, etc. may be reasonably useful
  - Imputation issues may be of minor importance for some questions
  - Can focus on establishments larger than 10 employees

Conclusion

- NETS may be useful for **static/cross sectional analysis**
  - Business location, industry, etc. may be reasonably useful
  - Imputation issues may be of minor importance for some questions
  - Can focus on establishments larger than 10 employees
- For **business dynamics** and **young firms**, large discrepancies between NETS and official data

Conclusion

- NETS may be useful for **static/cross sectional analysis**
  - Business location, industry, etc. may be reasonably useful
  - Imputation issues may be of minor importance for some questions
  - Can focus on establishments larger than 10 employees
- For **business dynamics** and **young firms**, large discrepancies between NETS and official data
- *Build your research design around the data limitations*

Conclusion

- NETS may be useful for **static/cross sectional analysis**
  - Business location, industry, etc. may be reasonably useful
  - Imputation issues may be of minor importance for some questions
  - Can focus on establishments larger than 10 employees
- For **business dynamics** and **young firms**, large discrepancies between NETS and official data
- *Build your research design around the data limitations*
- **The case for official data** (see section 6.2 of paper)
  - Official sources have well-defined “universe”
  - Official data rely heavily on administrative sources (IRS, SSA, etc.)
  - Official data collection focuses on consistency and scientific measurement

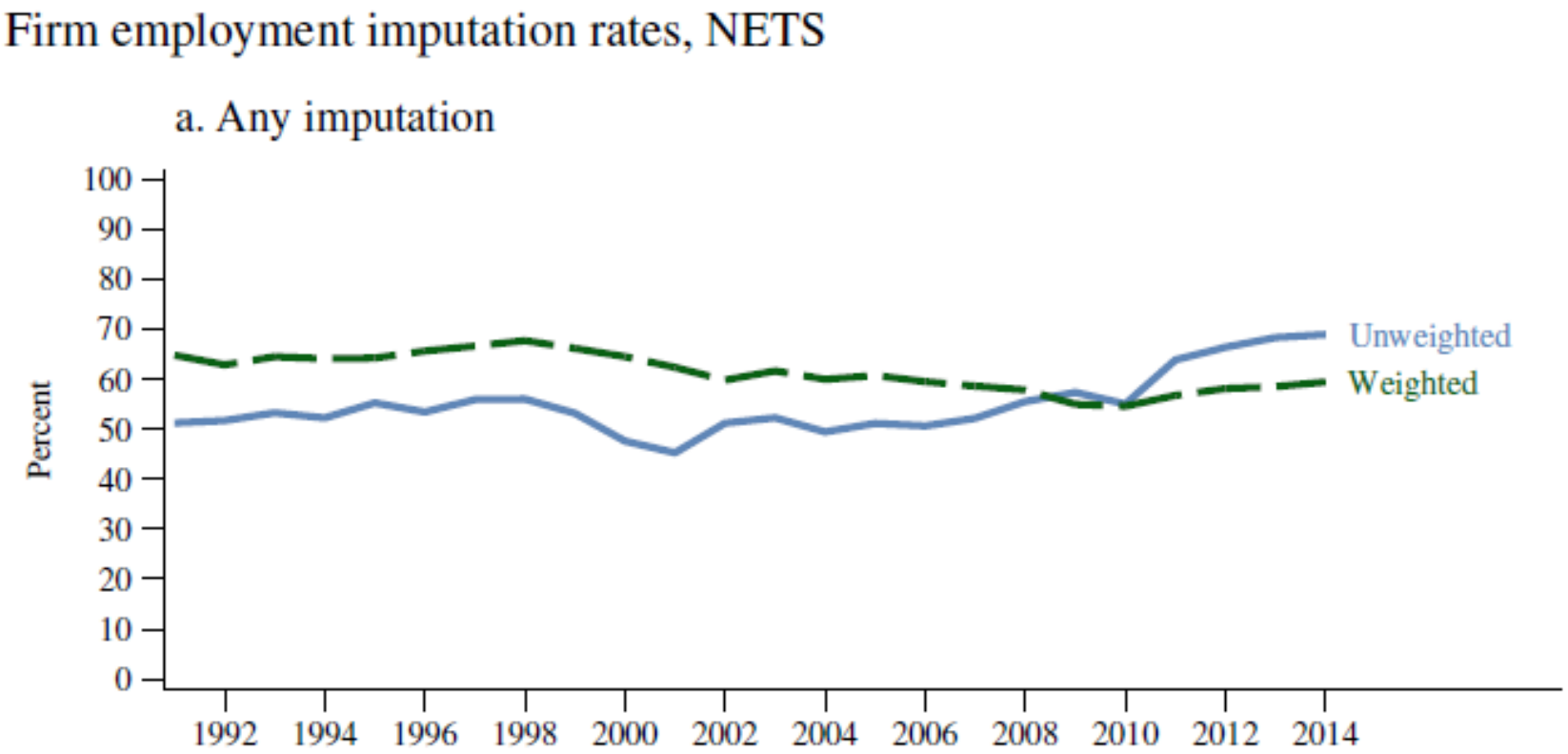
Our suggestions

1. Use NETS/D&B as a supplement to existing data, not as the definition of the business universe
2. Focus on employment *bins* instead of precise figures (and mind the imputation)
3. Static/cross-sectional aggregates appear reasonable, particularly without small establishments
4. Be *very* cautious about dynamics…
5. … and particularly the dynamics of young firms
6. Drop the sales data

Thanks

Extra slides

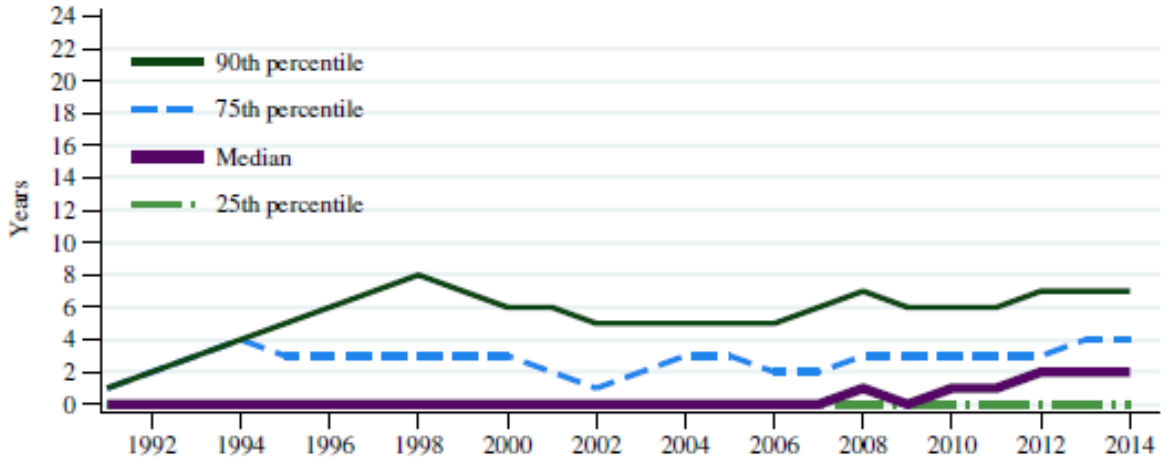
Firm imputation



Imputation spells

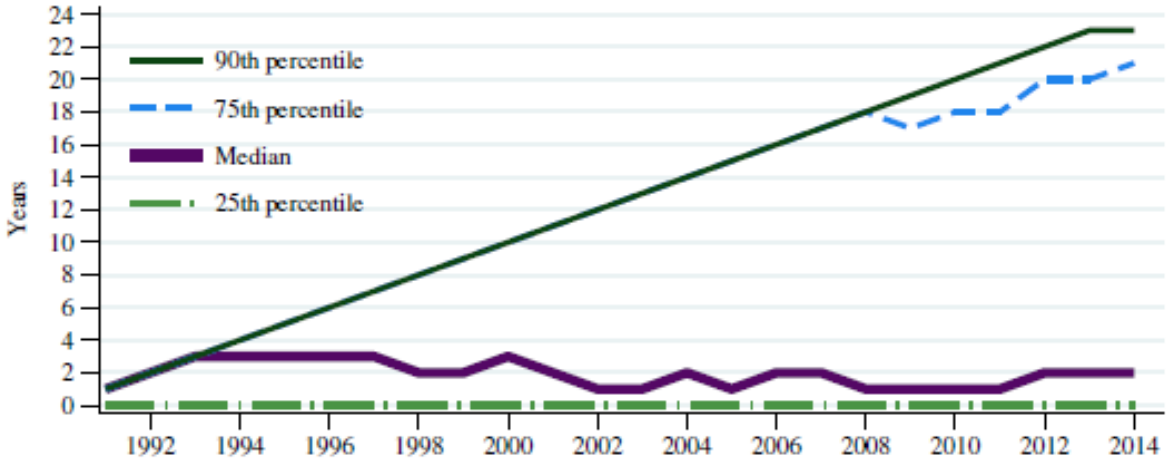
Length of imputation spells (unweighted)

a. All firms



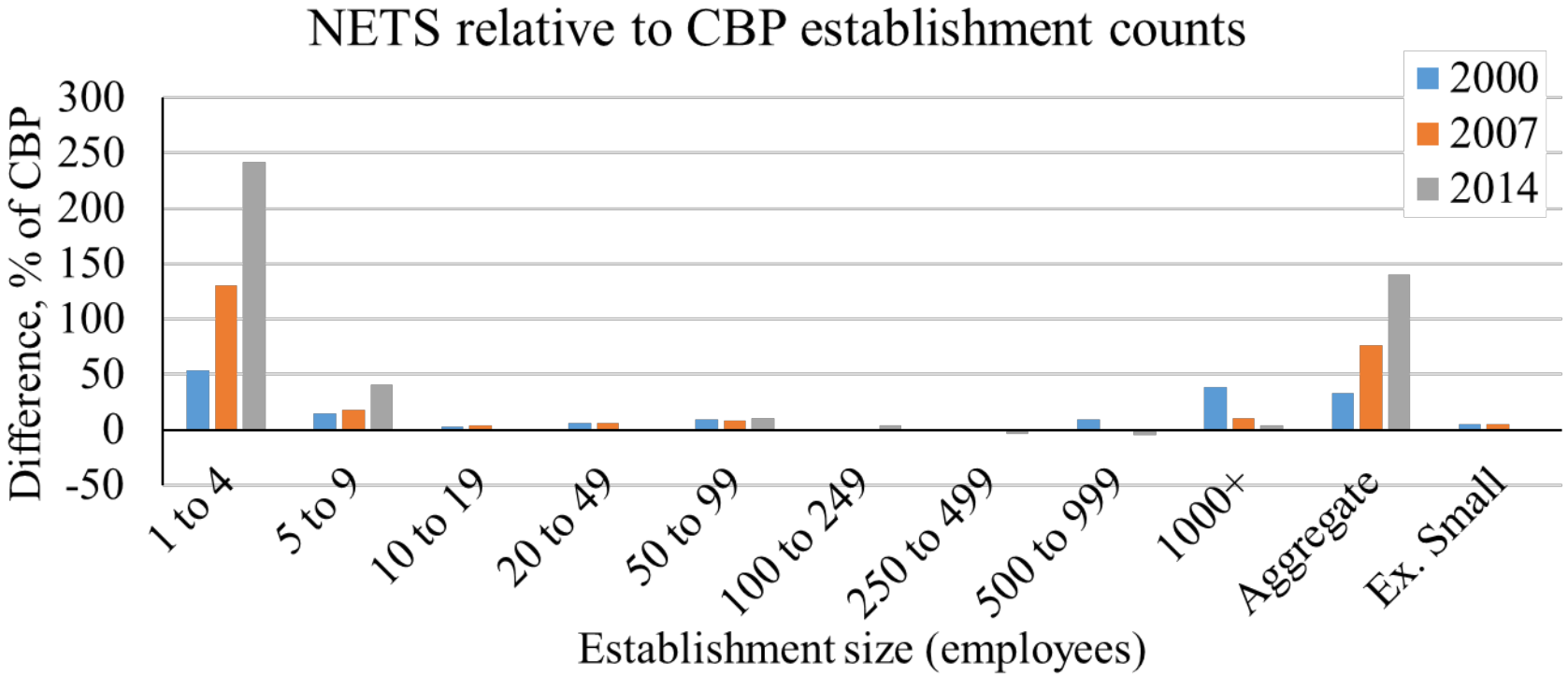
Length of imputation spells (employment weighted)

a. All firms



Size results

- NETS excess establishment count driven by small establishments (<10 employees)
- Close correspondence in most size classes



Business dynamics

- Make NETS as comparable to LBD as possible:
  - Raw data: each establishment records the establishment *to which they report* each year (hqduns)
  - How do deal with mergers, acquisiations, spinoffs, change of headquarters?
    - We link a firm from t-1 to the largest surviving fragment of t
    - For growth rate concepts, follow Haltiwanger, Jarmin, Miranda 2013 organic growth
  - Firm age: When a firm first appears, assign it the age of its oldest estab; age naturally thereafter
  - A lot of other detail cleaning up identifiers…
- DHS rates for job creation, job destruction, employment growth

“Up or out” dynamics

- NETS lacks young firm job creation
- … and destruction