The Effects of Editing and Imputation on Measured Plant-level and Aggregate Productivity Growth in a Panel of U.S. Manufacturing Plants

Hang J. Kim, University of Cincinnati

Martin Rotemberg, NYU

T. Kirk White, U.S. Census Bureau

(Preliminary and Incomplete)

FCSM Research and Policy Conference, Washington, DC

October 25, 2022

The research presented here was conducted while the first two authors were Special Sworn Status researchers and the third author was an employee of the U.S. Census Bureau. Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release (DRB approval number CBDRB-FY2022-CES020-003, DMS subproject number 7526913).

Plant-level Productivity and Data Cleaning

- Large economics literature uses Census Bureau's plant-level manufacturing data to study productivity and productivity dynamics
    - See, e.g., Bailey, Hulten, and Campbell (1992), Olley and Pakes (1992), Bartelsman and Doms (2000), Foster, Haltiwanger, and Syverson (2008), Syverson (2011)
- Literature mostly ignores fact that large % of Census's manufacturing data are *imputed*
- Data cleaning done by Census Bureau has huge impact on measured cross-sectional productivity dispersion
    - White, Reiter and Petrin (2018), Rotemberg and White (2020)
- How does Census data cleaning affect measured productivity growth?

United States® Census Bureau

What We Do in This Paper

- Start with unedited “captured” Annual Survey of Manufactures (ASM), 2009-2013
- Use Census Bureau’s edit rules
- Extend Bayesian simultaneous edit-imputation method of Kim, Cox, Kerr, Reiter, and Wang (2015) to panel data
- Compare estimates of aggregate productivity growth using:
  - Data edited-imputed by Census Bureau
  - Data edited-imputed by us using our Bayesian method

The Annual Survey of Manufactures

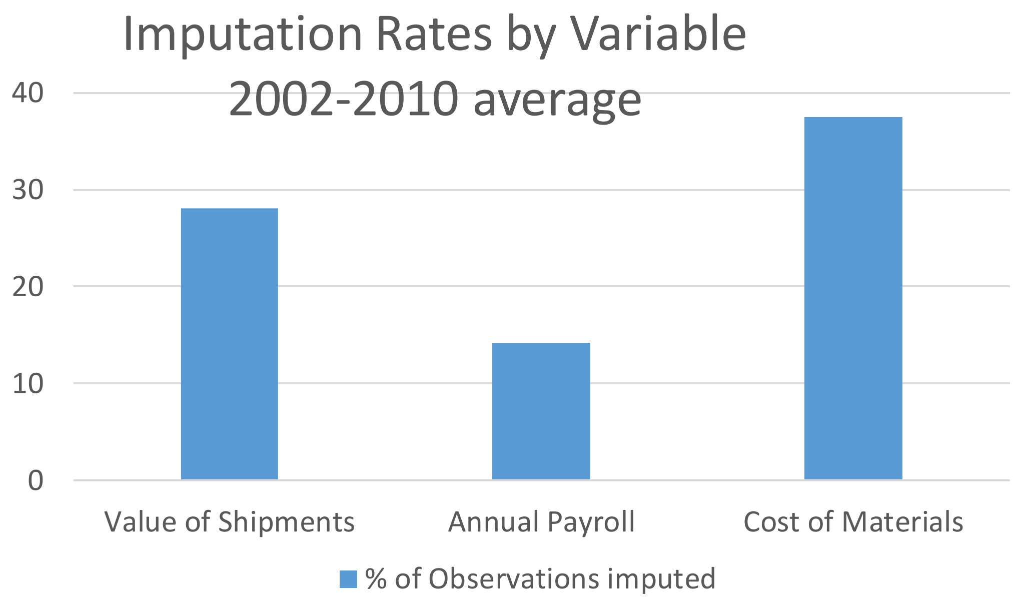
- Plant-level survey conducted annually*
- Sample size: about 50,000 plants
- Universe is U.S. manufacturing sector (about 300,000 plants)
- Certainty sample:
  - large plants, sampled every year
- Probability sample:
  - Small-to-medium plants, sampled with probability proportional to size
  - 5-year rotating panels beginning in years ending in ‘4’ or ‘9’

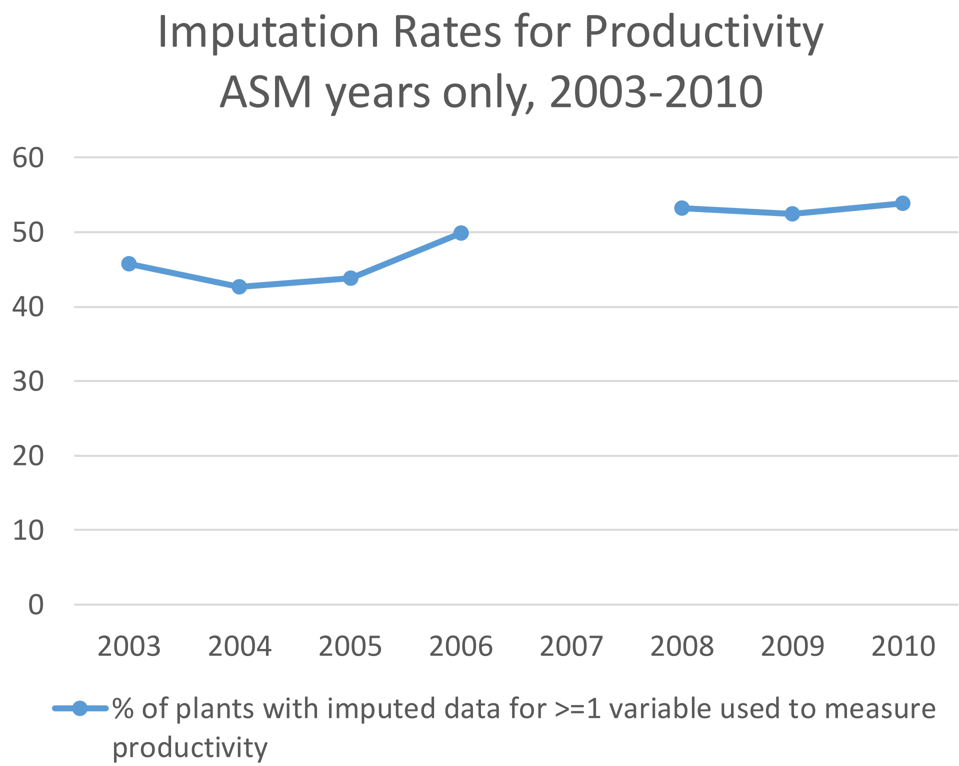
Variables in the ASM

- ASM
  - Collects (at plant-level) and publishes (at industry level) 55 plant-level variables
- This paper:
  - We focus on 5 variables used for measuring productivity:
    - Value of shipments
    - Payroll
    - Employment
    - Cost of materials
    - Capital stock*

United States® Census Bureau

Imputation Rates in the ASM





Source: Foster, Grim, Haltiwanger, and Wolf (2017), table A7

ASM Editing and Imputation Comparison

What Census Bureau does	What we do in this paper
Uses within-year ratio edit rules	Same
Uses year-to-year ratio-of-ratios edit rules	Same
Edits-imputes one year at a time (prior-year data is fixed)	Edit-impute 5-year panel simultaneously
Fellegi-Holt editing algorithm (minimizes # of edits)	Simultaneous Bayesian edit-imputation (chooses likeliest variable(s) to edit given the "good" data)
Uses hierarchy of imputation methods, e.g., • Administrative records substitution • Uni- or trivariate regression models • Analyst corrections	Use truncated Dirichlet Process mixture of normals to model joint distribution of 5x5 variables
Edit rules & imputation models same within NAICS6	Edit rules & imputation models same within NAICS4

United States Census Bureau

Types of ratio edit rules used in the ASM

- Cross-sectional:

$$XLB_{jt} \leq \frac{X_{ijt}}{Y_{ijt}} \leq XUB_{jt}$$

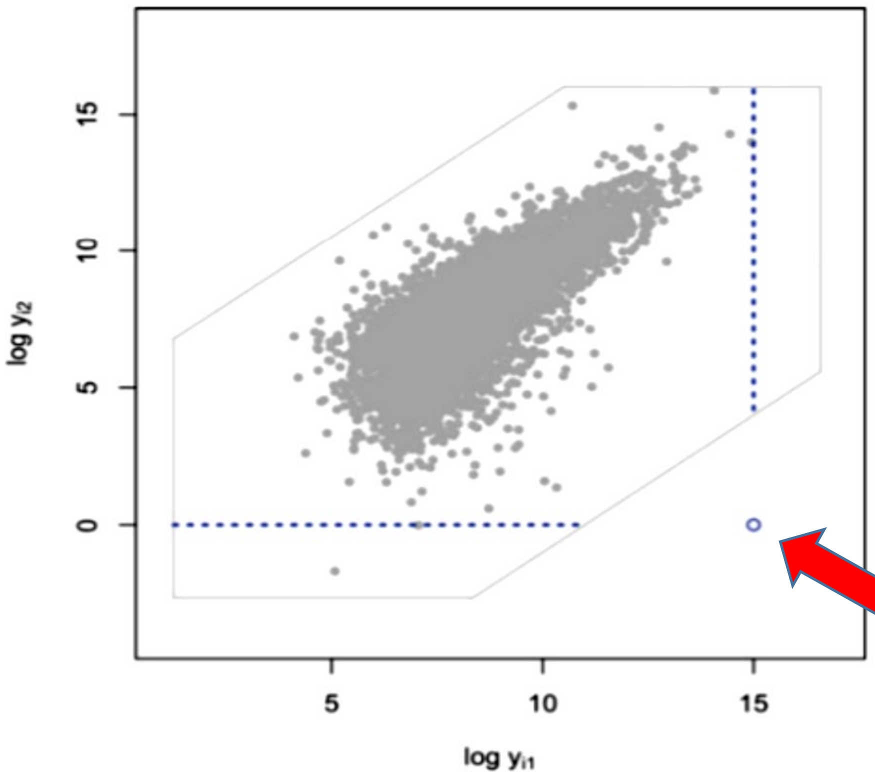
- Longitudinal :

$$LLB_j \leq \frac{\frac{X_{ijt}}{Y_{ijt}}}{\frac{X_{ij,t-1}}{Y_{ij,t-1}}} \leq LUB_j$$

United States® Census Bureau

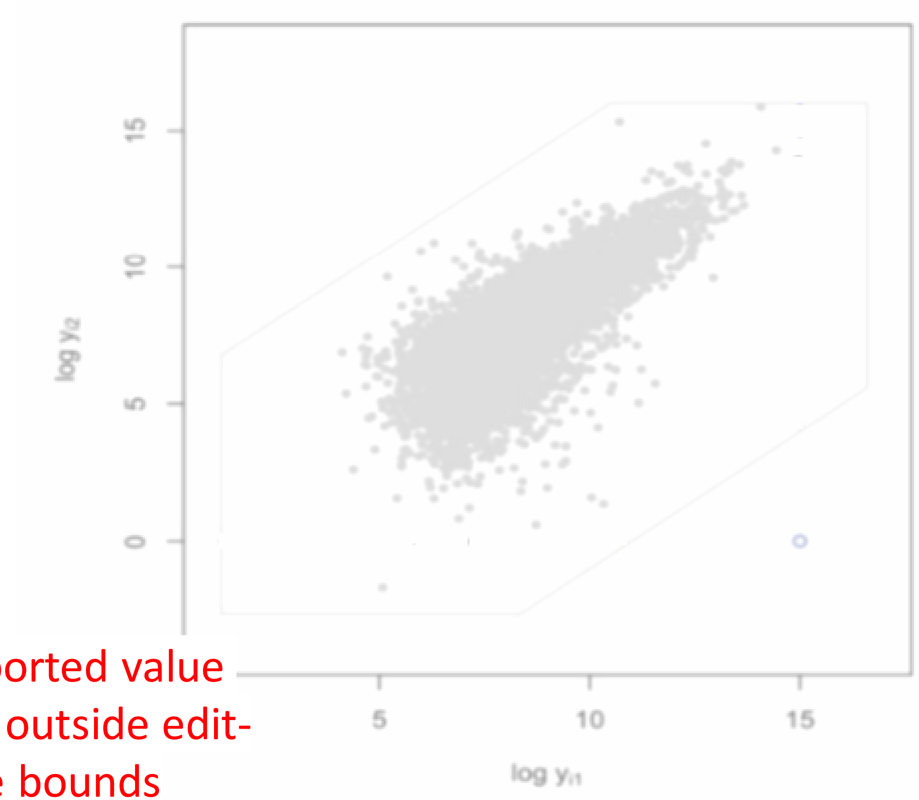
Census Editing vs. Bayesian Edit-Imputation
(Note: Scatterplots are from Kim et. al (2015) synthetic data)

What Census Bureau does



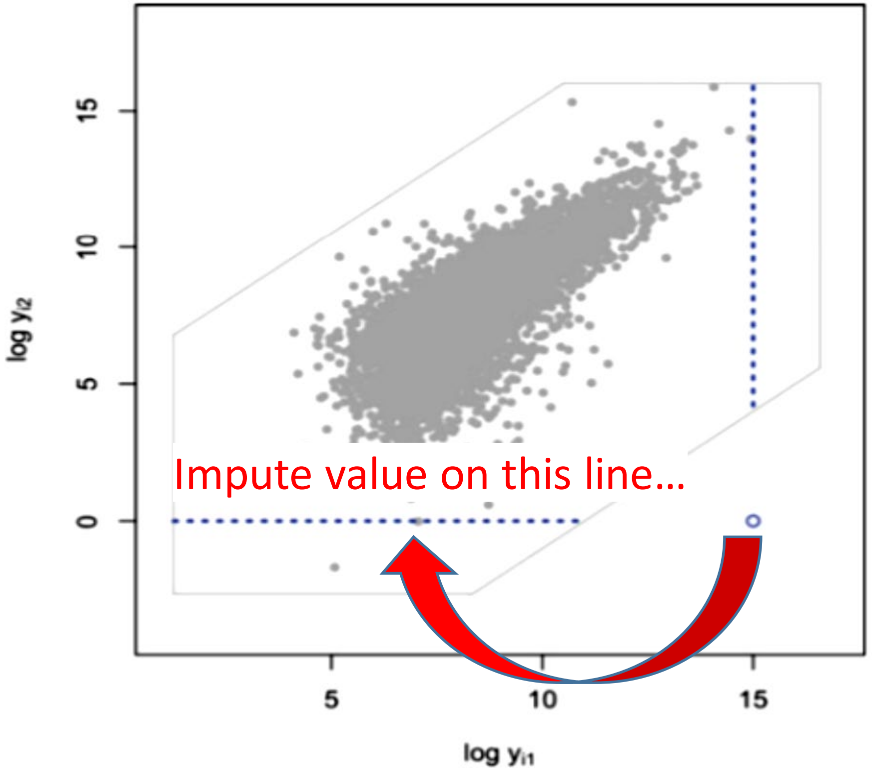
Reported value lies outside edit-rule bounds

What we do in this paper

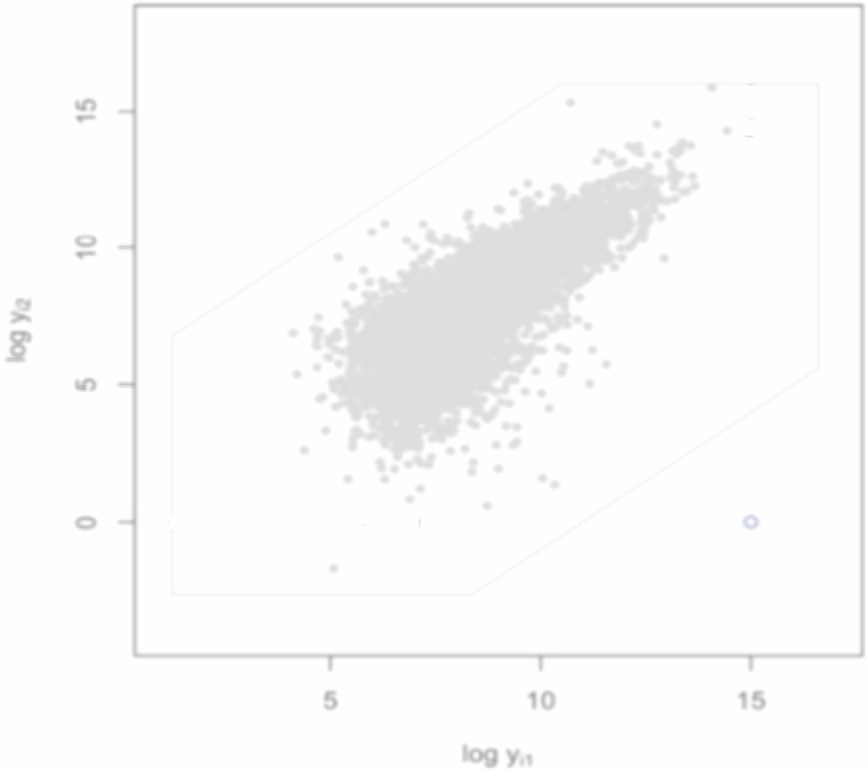


Census Editing vs. Bayesian Edit-Imputation

What Census Bureau does

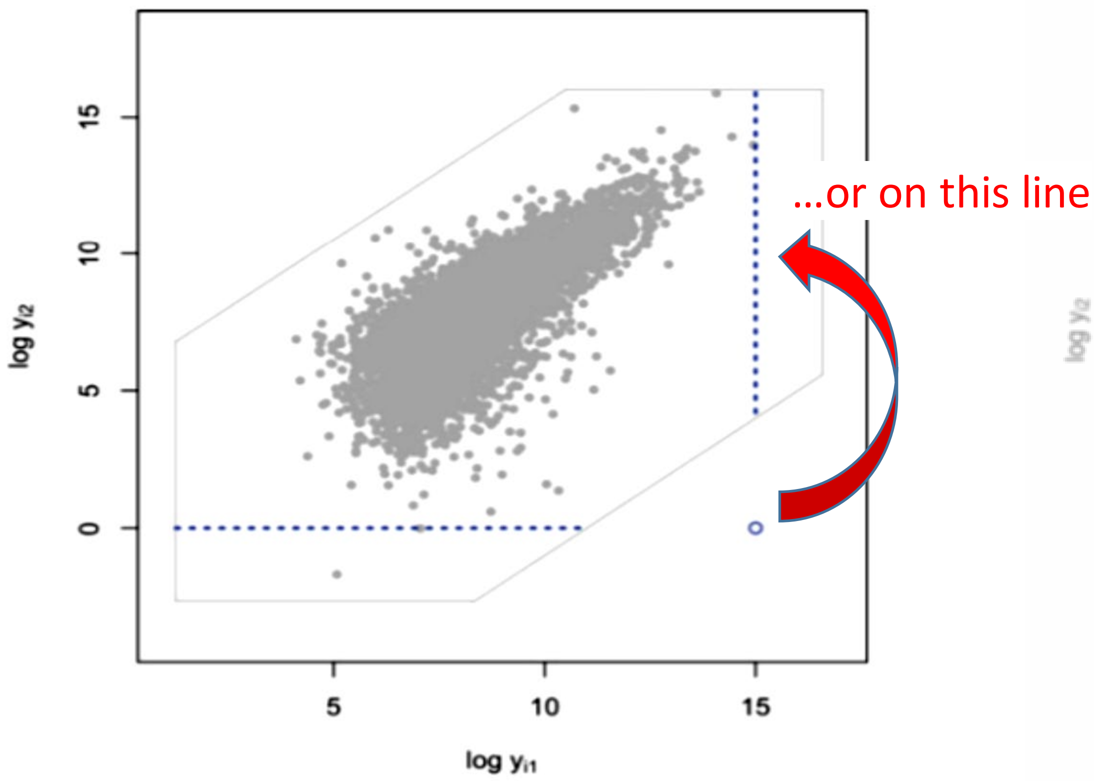


What we do in this paper

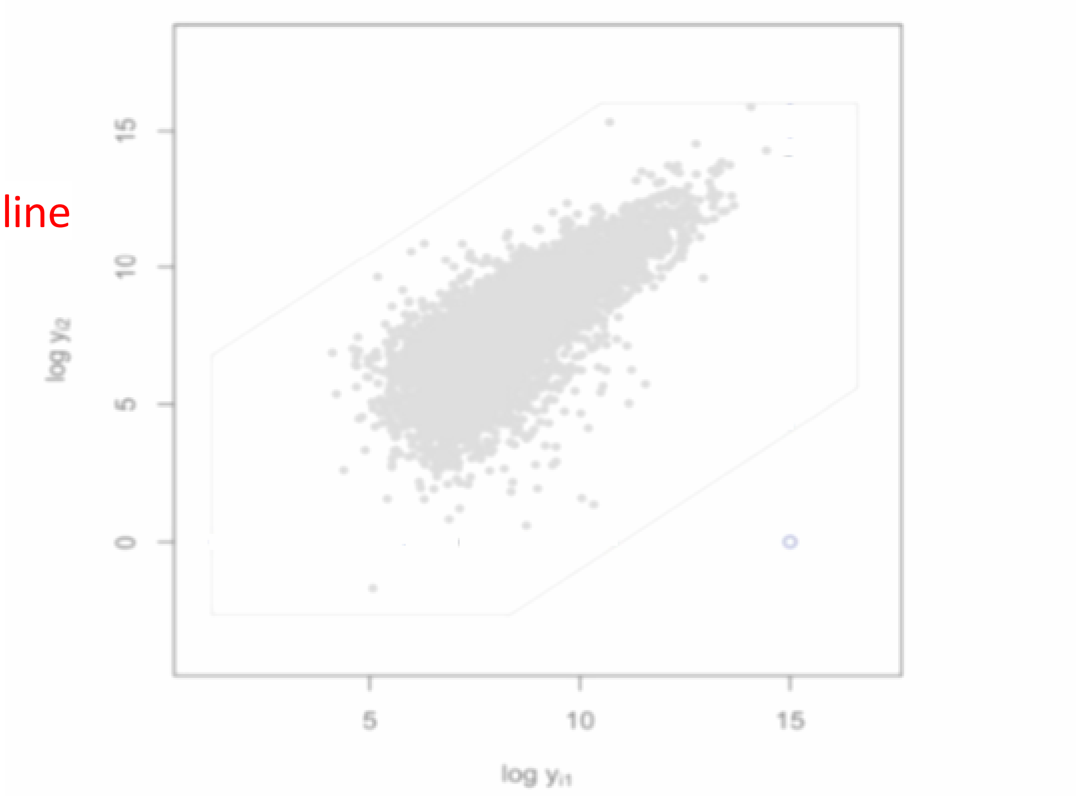


Census Editing vs. Bayesian Edit-Imputation

What Census Bureau does



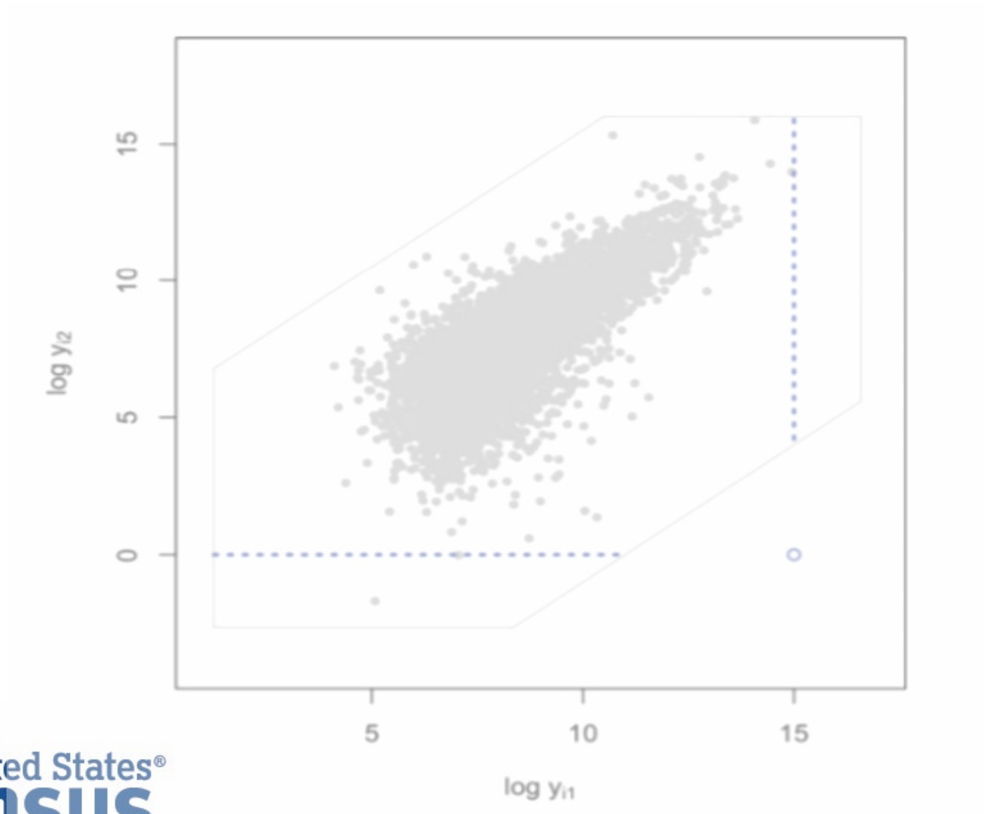
What we do in this paper



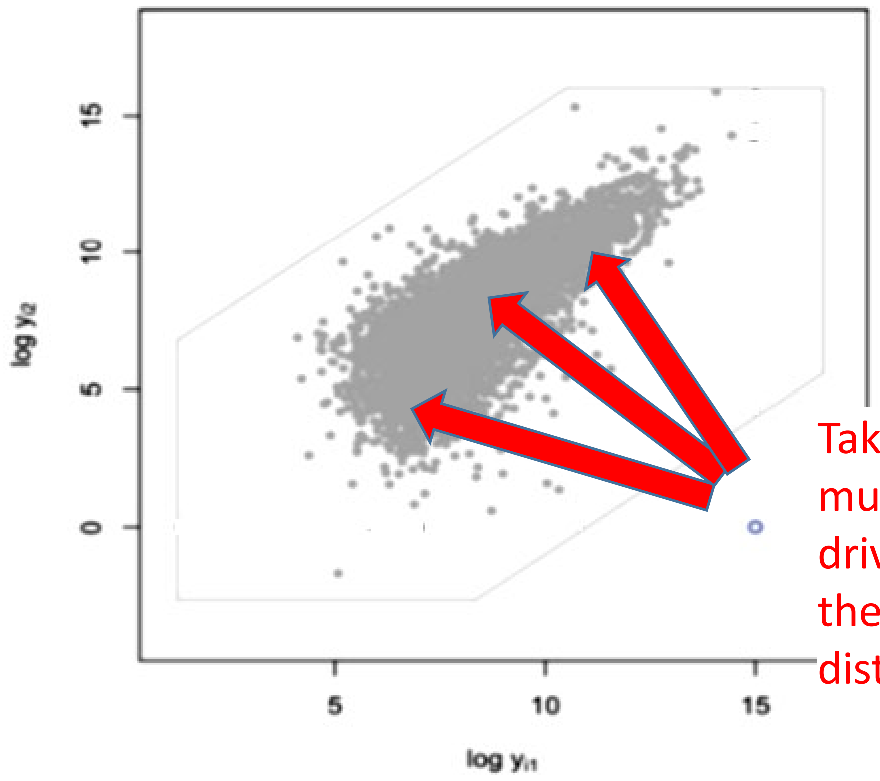
United States Census Bureau

Census Editing vs. Bayesian Edit-Imputation

What Census Bureau does

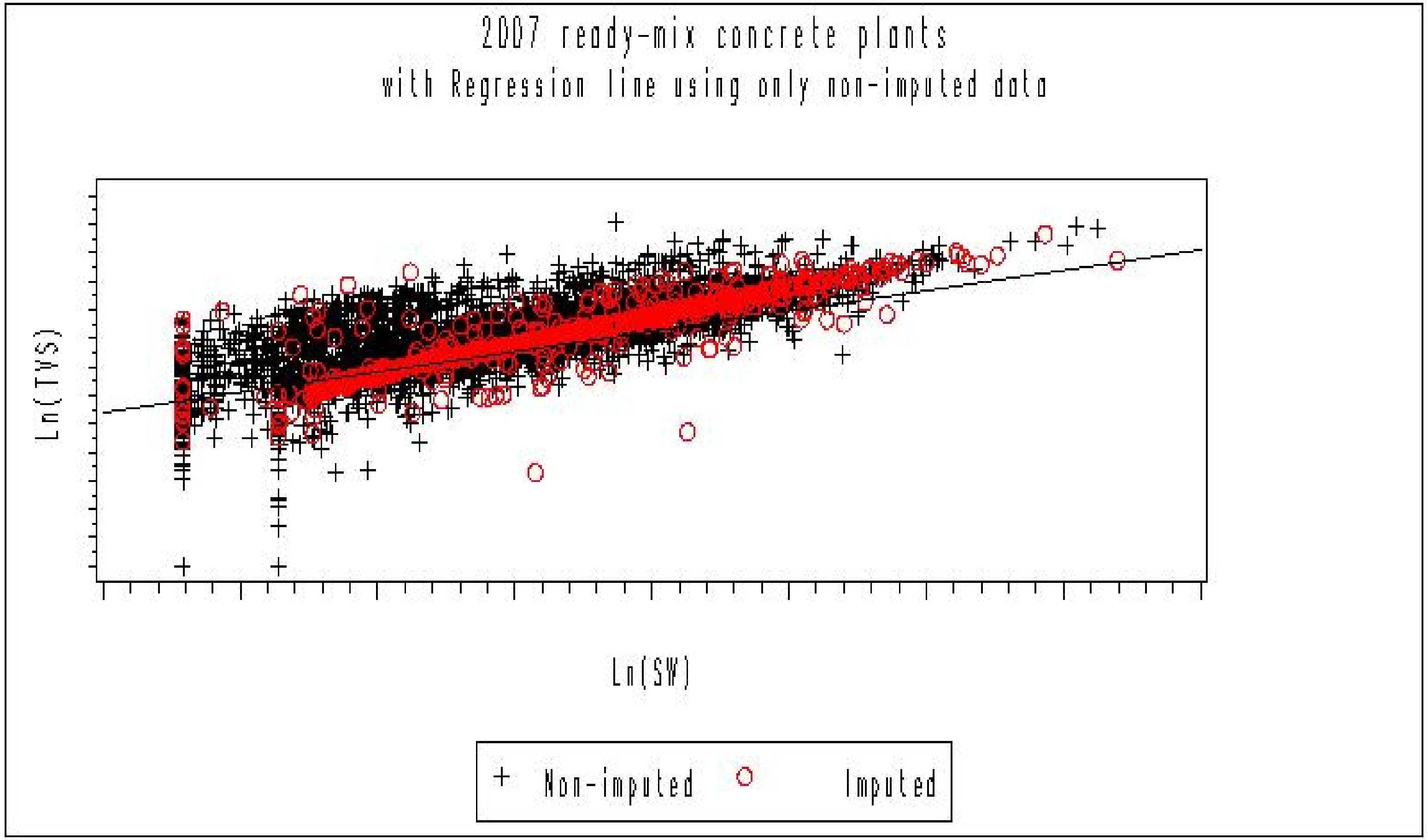


What we do in this paper



Take draws from multiple data-driven models of the joint distribution

Census Bureau Imputation
(includes imputation for non-response)



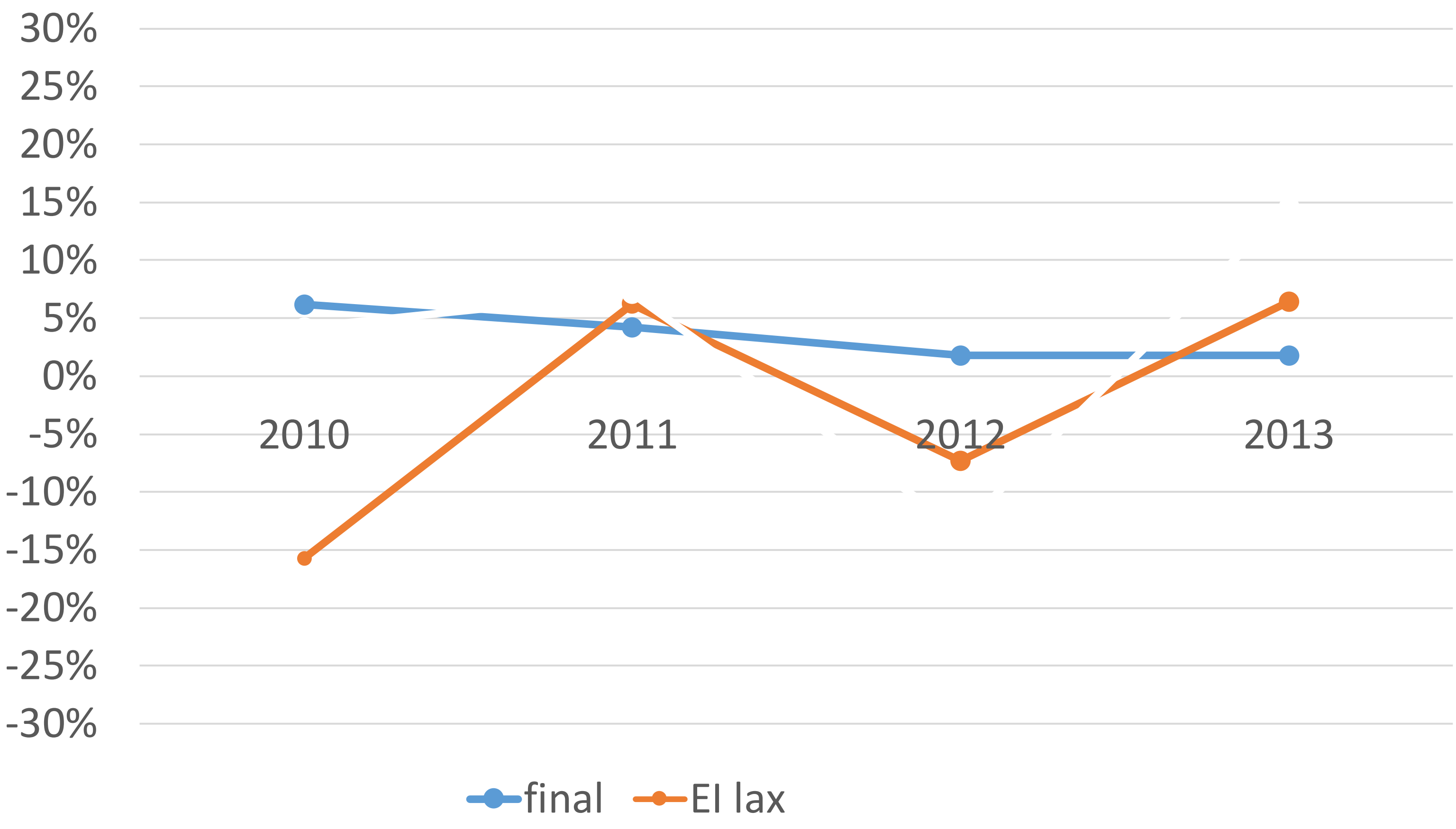
Three Versions of ASM Data

- One "final" dataset edited-imputed by Census
- Two versions of data multiply-edited-and-imputed by us:
  - "EI lax" data:
    - use the laxest of the ratio-edit bounds within each NAICS4 industry
  - "EI strict" data:
    - use the strictest of the ratio-edit bounds within each NAICS4 industry

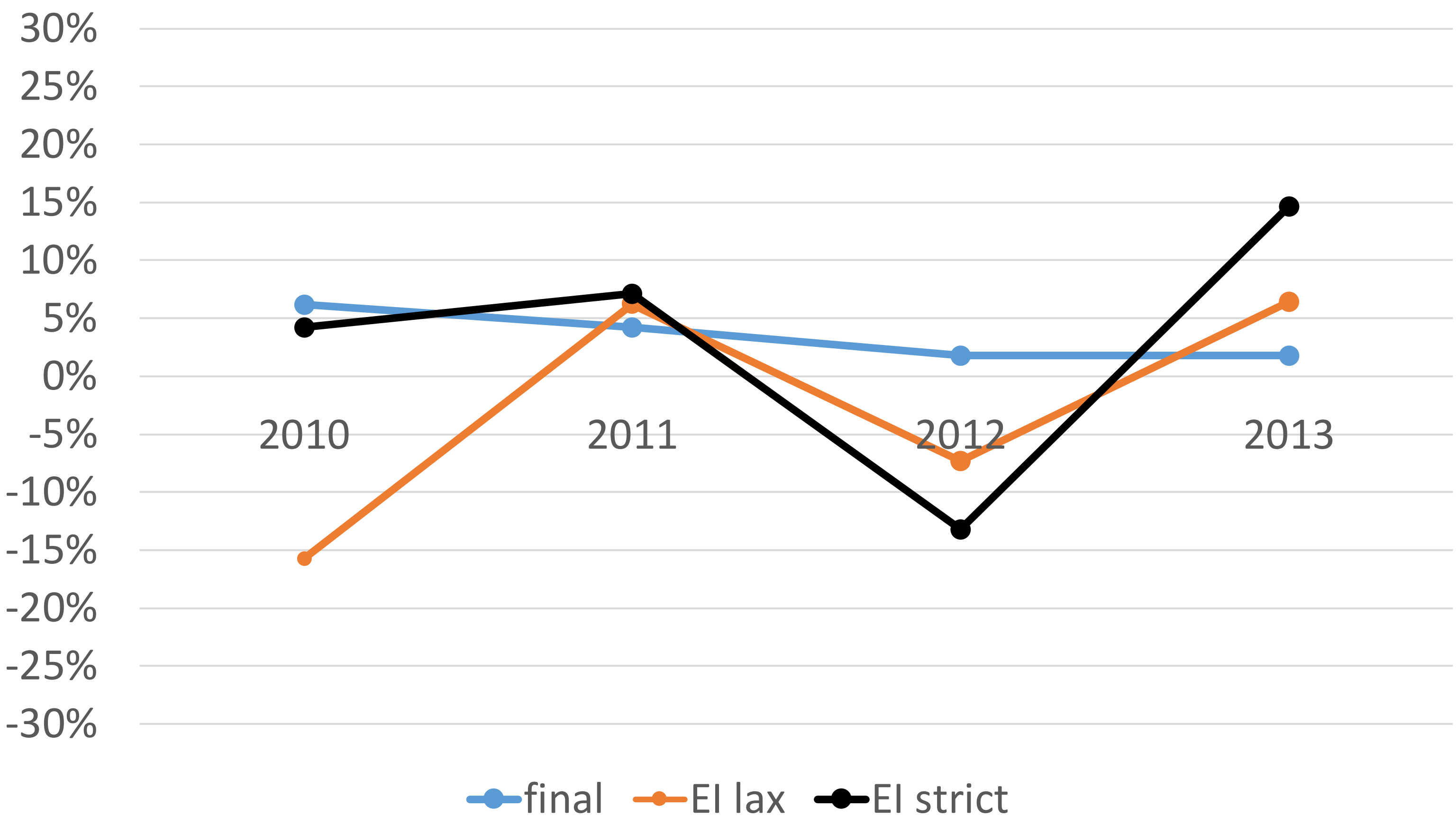
Estimates from multiply imputed data

- For each of our models (“lax” and “strict”) and within industry:
  - Run an MCMC with a burn-in of 3000 iterations
  - Keep 10 implicates with 300 iterations between implicates
- For each implicate-year, estimate aggregate productivity growth
- Compute means and standard errors across implicates

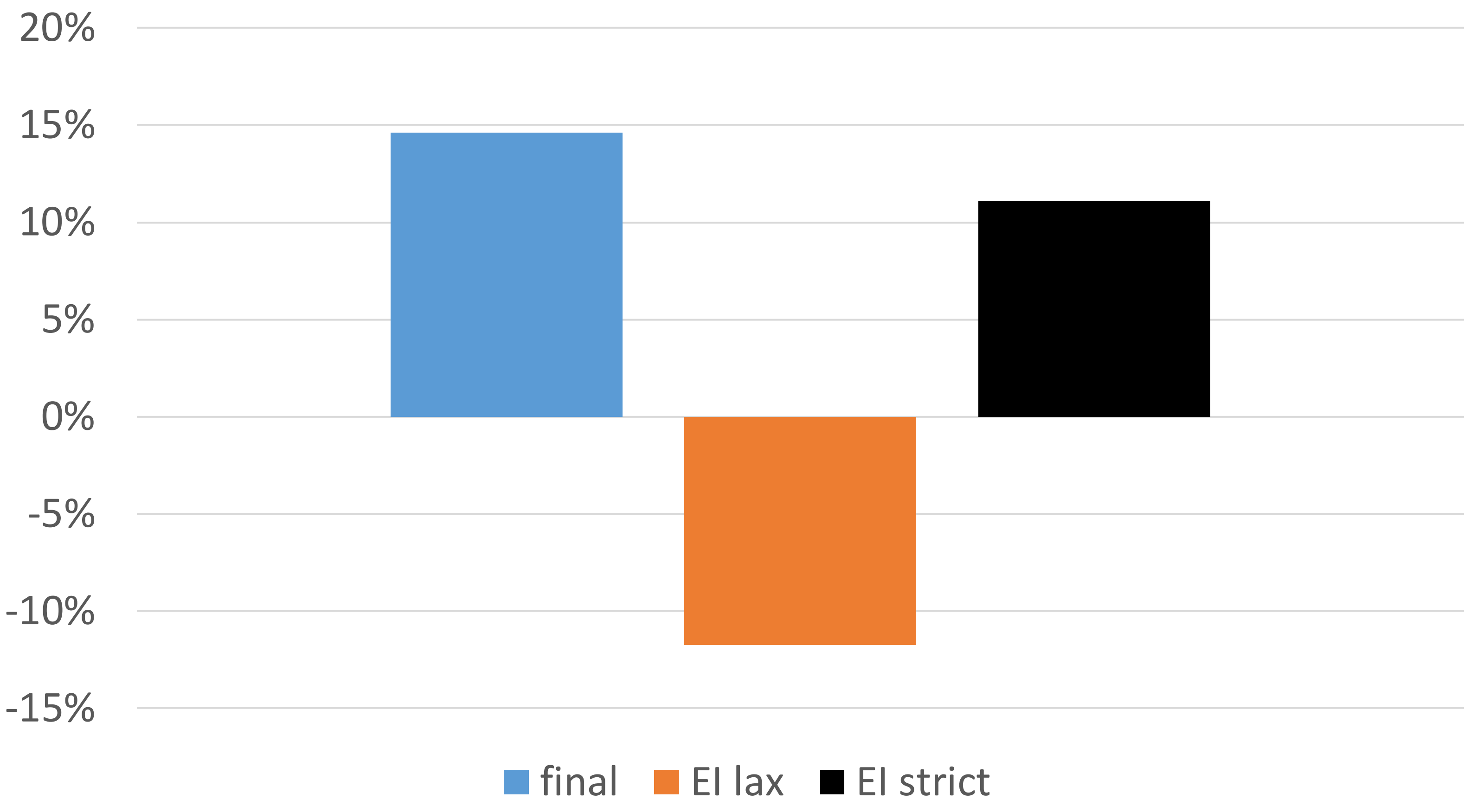
Aggregate Productivity Growth Rates (annual), 2009-2013 ASM, Census Final vs. EI “lax” mean



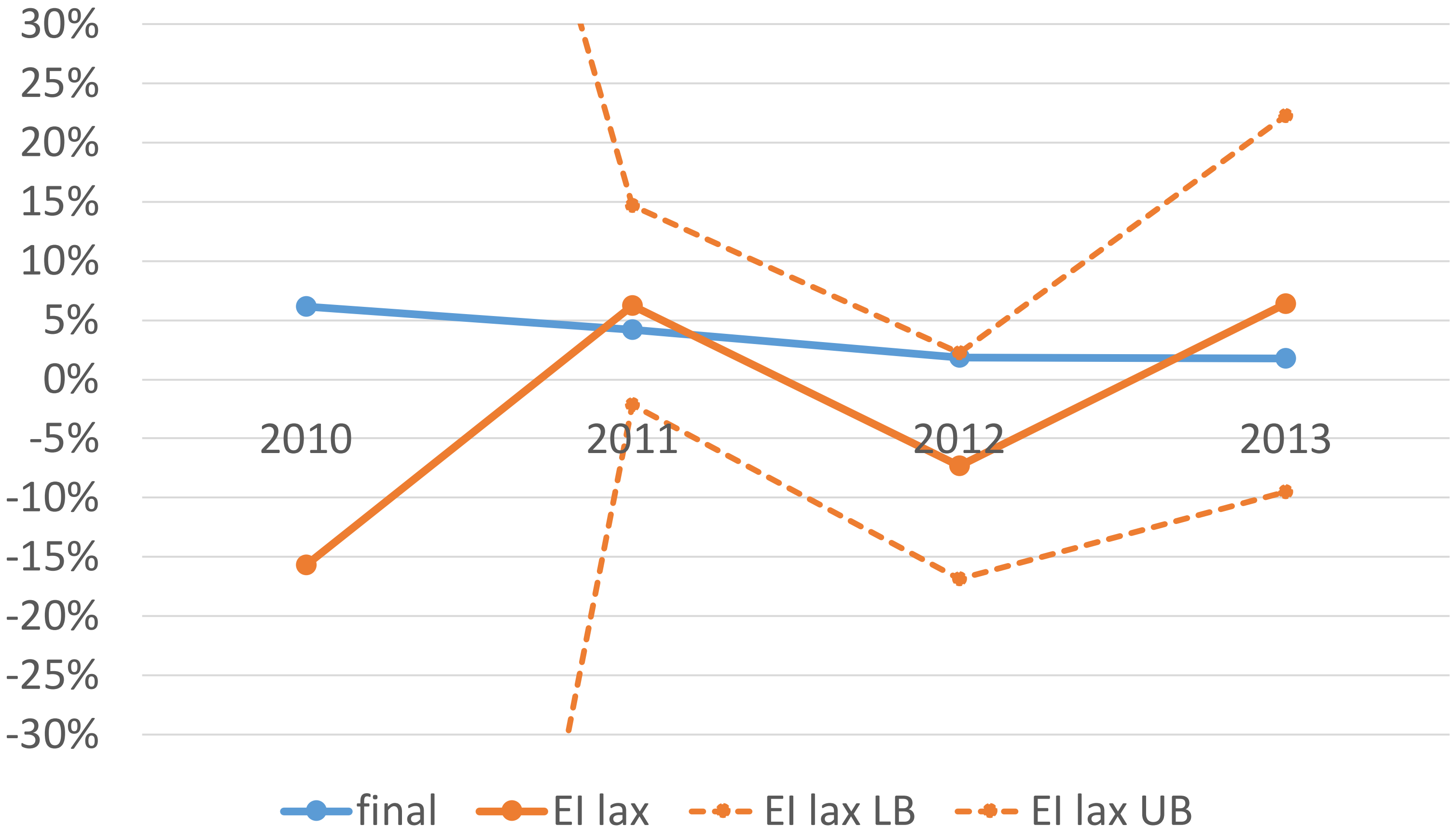
Aggregate Productivity Growth Rates (annual), 2009-2013 ASM, Census Final vs. EI "lax" mean vs. EI "strict" mean



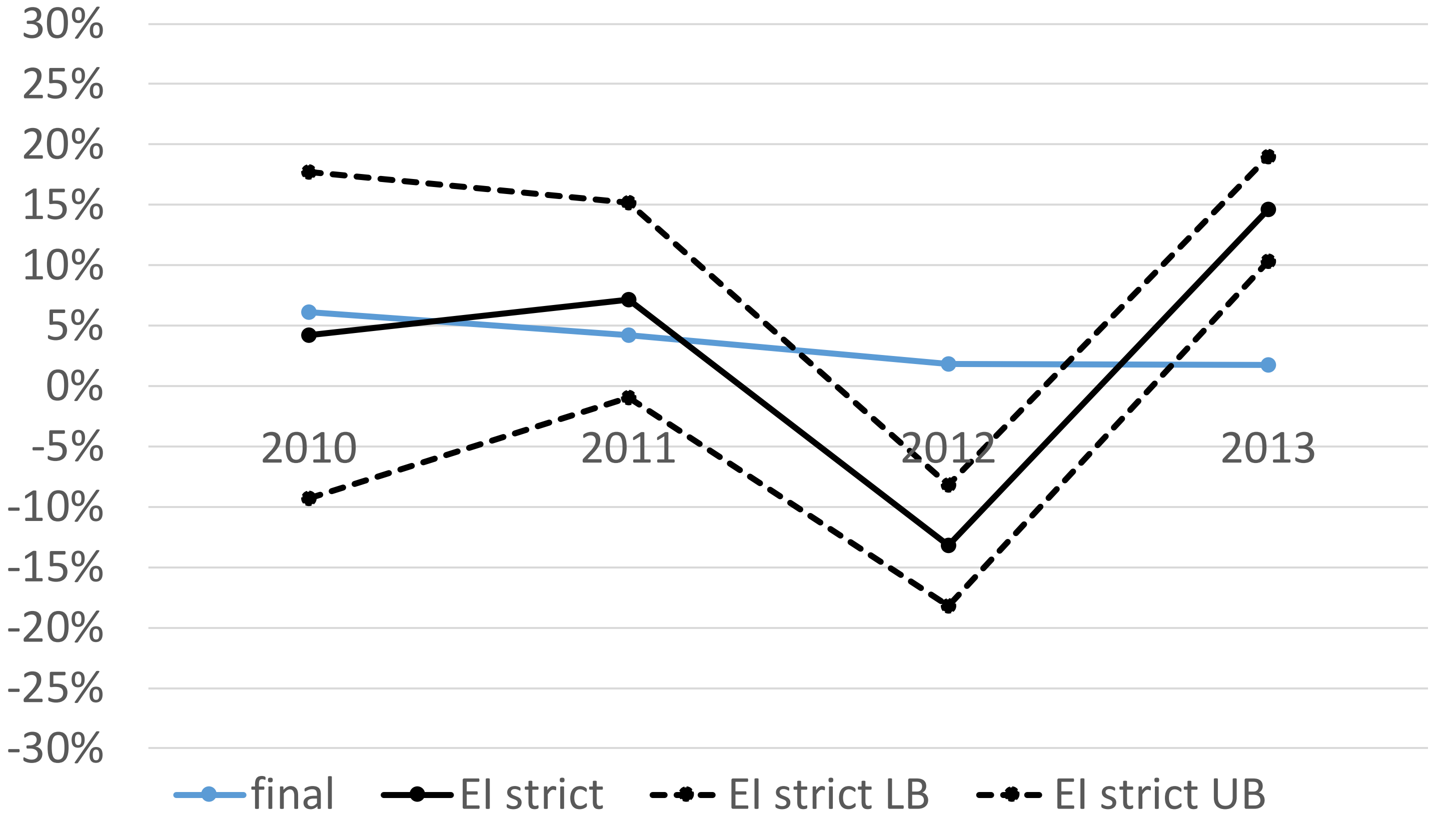
Aggregate Productivity Growth Rates (5-year), 2009-2013 ASM, Census Final vs. EI “lax” mean vs. EI “strict” mean



Aggregate Productivity Growth Rates (annual), Census final vs. EI "lax" with 95% Confidence Intervals



Aggregate Productivity Growth Rates (annual)
Census final vs. EI “strict” with 95% Confidence Intervals



Preliminary Conclusions

- The choice of editing and imputation method has a huge effect on measured aggregate productivity growth (APG) in U.S. manufacturing

- Uncertainty of APG estimates due to imputation is also large

United States® Census Bureau

Next Steps

- Model validation exercises (training & truth sets)
- Why such big differences between Census final vs. our EI APG measures?
    - Are certain industries driving the results?
    - Small plants with big survey weights and big growth rates?
- Look at the effect of editing and imputation on :
    - Persistence of plant-level productivity (e.g., autocorrelation of TFPR)
    - Variability of plant-level TFP “shocks”
    - Labor productivity growth
    - Association between TFPR and exit
    - All of the above by, e.g., firm size

References

- Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2017. "Macro and Micro Dynamics of Productivity: From Devilish Details to Insights" *NBER Working Paper 23666*.
- Kim, Hang J., Lawrence H. Cox, Alan F. Karr, Jerome P. Reiter, and Quanli Wang. 2015. "Simultaneous Edit-Imputation for Continuous Microdata", *Journal of the American Statistical Association*, Vol. 110, No. 511.
- Rotemberg, M. and T. Kirk White. 2020. "Measuring Cross-Country Differences in Misallocation" *CES Working Paper CES-WP-16-50R*.
- White, T. Kirk, Jerome Reiter, and Amil Petrin. 2018. "Imputation in U.S. Manufacturing Data and Its Implications for Productivity Dispersion." *The Review of Economics and Statistics.*

United States® Census Bureau