A Time-Series Examination of the Quality of Industry-Level U.S. Productivity Data

Sergio H. Lence and Alejandro Plastina
Iowa State University

INTRODUCTION

- Large number of productivity analyses have based on Total Factor Productivity (TFP)
- Industry-level TFP data used to investigate important economic issues
  - Concentration of productivity gains across industries and whether such gains are linked to information technology (Stiroh 2002)
  - Whether automation is labor-displacing (Autor and Salomons 2018)
  - Whether recent rise in capital share can be attributed to increasing automation (Aghion, Jones, and Jones 2019)
  - Impact of sectoral trends in TFP and labor growth on GDP growth (Foerster et al. 2022)

INTRODUCTION

- Industry-level TFP data used to investigate important economic issues (cont.)
  - Contributions of individual industries to U.S. aggregate TFP growth (Jorgenson, Ho, and Samuels 2019)
  - Productivity gap between Europe and the United States in the late 1990s and early 2000s (van Ark, O'Mahony and Timmer 2008)
  - Recent interest in "environmentally-adjusted" TFP indicators, which take into account production of undesirable by-products and externalities, and how intensely natural resources are used (OECD 2020)

INTRODUCTION

- Significant efforts to develop proper measures of individual components of TFP and to evaluate relative merits of alternative aggregation methods
  - OECD (2001); Fuglie, Wang, and Ball (2012); Fuglie (2015); Shumway et al. (2017); USDA-ERS (2021); Szulc (1964); Eltetö and Köves (1964); Jorgenson and Griliches (1967); Caves, Christensen, and Diewert (1982a, 1982b); Bjurek (1996); Balk and Althin (1996); O'Donnell (2012, 2016); Färe and Zelenyuk (2021)
- Few studies exploring quality of real-world TFP data series
  - Alston (2018); Andersen, Alston, and Pardey (2011)

OBJECTIVES

- Examine pairwise consistency of industry-level U.S. TFP series from three alternative highly regarded sources

METHODS

Summary of Proposed Approach

1. Find order of integration of each log(TFP) series for a sector
2. Determine whether there are issues
   - Series have different order of integration
   - Regressions based on properly differenced series indicate issues
3. If issues are found, test whether differences in series is stationary or not
   A. If difference in series is nonstationary, at least one of the log(TFP) series must be of poor quality
      - log(TFP) series for that sector have no long-term relationship because they are not cointegrated
   B. If difference in series is stationary, test whether one series leads the other
      - Leading series is of better quality

METHODS

1. Find order of integration of each log(TFP) series for a sector
   - Start by assuming series is I(j $\leq$ 2) (i.e., integrated of order j $\leq$ 2)
     - Test H_0: I(2) vs. H_1: I(1)
       - Series is I(2) if H_0 not rejected
       - Series is I(1) if H_0 rejected
     - If series found I(1), test H_0: I(1) vs. H_1: I(0)
       - Series is I(1) if H_0 not rejected
       - Series is I(0) if H_0 rejected

METHODS

2. Determine whether regressions based on properly differenced series indicate issues
   - Let x_t and y_t be a sector's two alternative log(TFP) series
     - Stationary series:

$$X_t = \Delta^{jx} x_t \text{ if } x_t \text{ is } I(j_x)$$
$$Y_t = \Delta^{jy} y_t \text{ if } y_t \text{ is } I(j_y)$$

   - Ideally, a = 0, b = 1, and Var(ε) = 0 if Q_t is stationary series based on true (but unobserved) log(TFP) and E(ε) = 0

$$X_t = a_x + b_x Q_t + \varepsilon_{x,t}$$
$$Y_t = a_y + b_y Q_t + \varepsilon_{y,t}$$

   - Let $Z_t = X_t - Y_t$, and run OLS regressions to infer magnitudes of {a, b, Var(ε)}

$$Z_t = \alpha_x + \beta_x X_t + e_{x,t}$$
$$Z_t = \alpha_y + \beta_y Y_t + e_{y,t}$$

METHODS

2. Determine whether regressions based on properly differenced series indicate issues
   - Let $Z_t = X_t - Y_t$, and run OLS regressions to infer magnitudes of {a, b, Var(ε)}

$$Z_t = \alpha_x + \beta_x X_t + e_{x,t}$$
$$Z_t = \alpha_y + \beta_y Y_t + e_{y,t}$$

     - Issues if $\max\{|\alpha_x|, |\beta_x|, |\alpha_y|, |\beta_y|\}$ is large

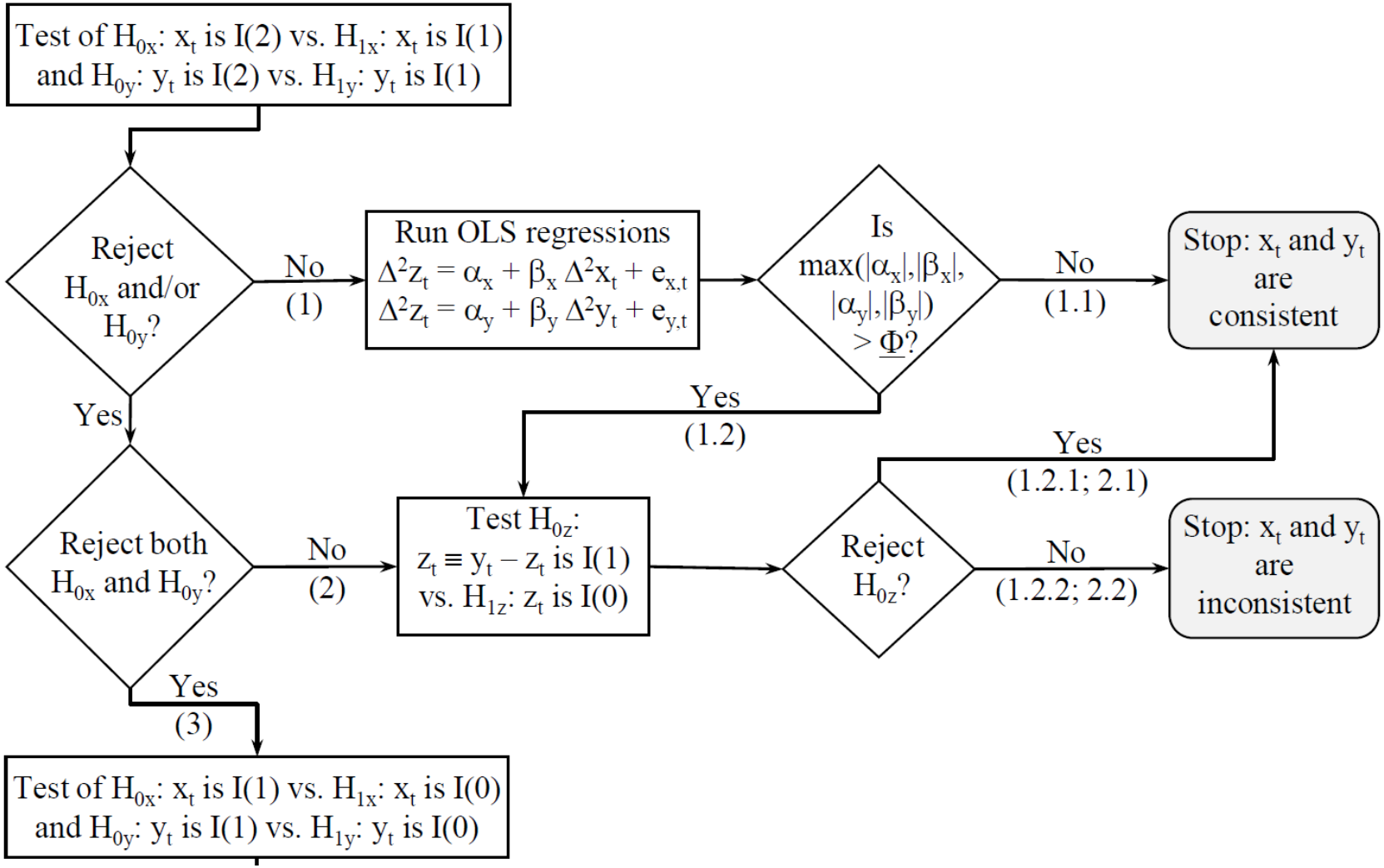
METHODS

3. If issues are found, test whether differences in series is stationary or not
   - Test H_0: $z_t = x_t - y_t$ is I(1) vs. z_t is I(0)
     - z_t series is I(0) if H_0 not rejected
     - z_t series is I(1) if H_0 rejected
       - x_t and y_t are not cointegrated
         - No long-term relationship between x_t and y_t

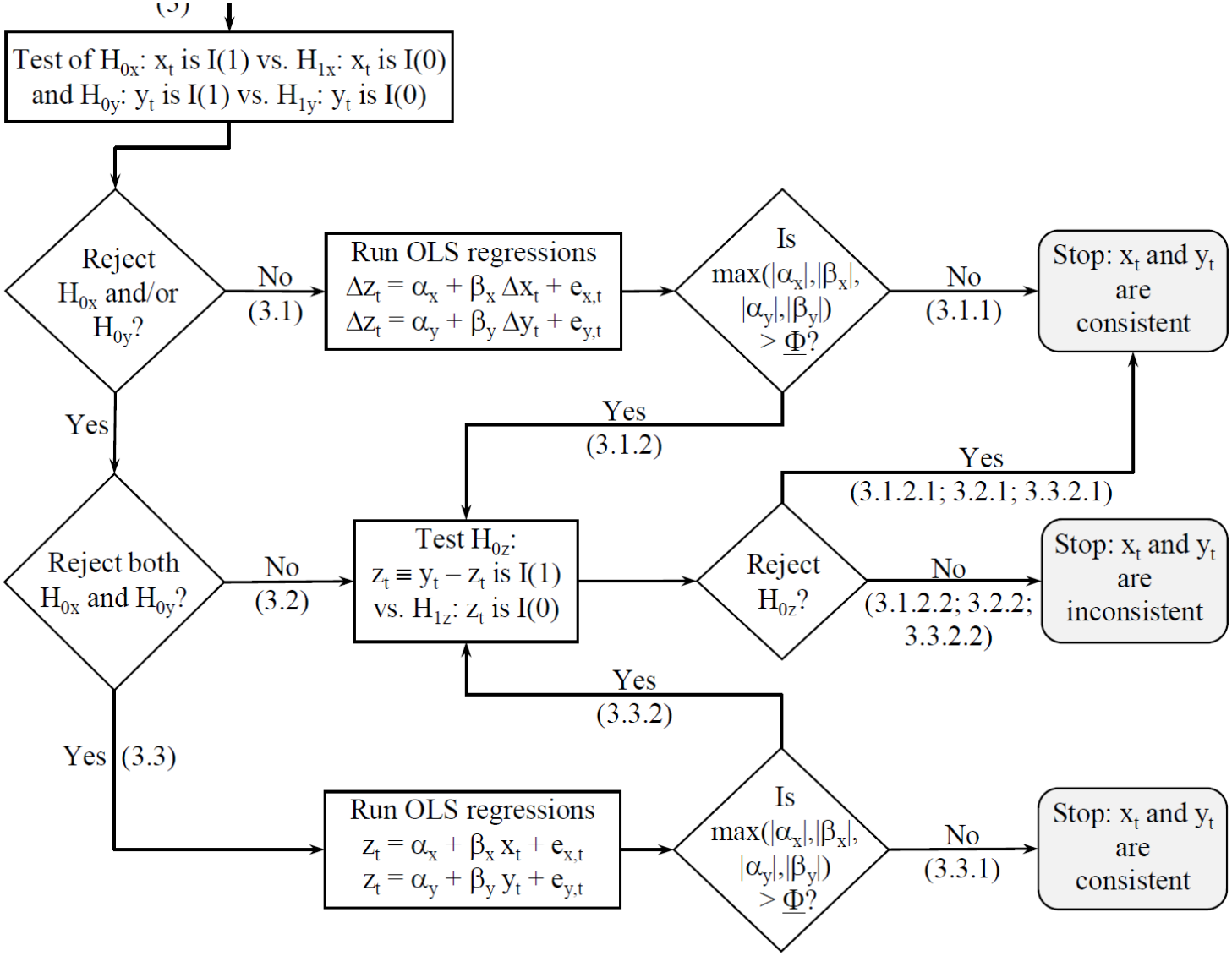
At least one of xt and yt must be a poor measure of the true log(TFP)

METHODS

Figure 1. Econometric strategy to classify pairs of time series observations.



METHODS


Test of H_{0x}: x_t is I(1) vs. H_{1x}: x_t is I(0) and H_{0y}: y_t is I(1) vs. H_{1y}: y_t is I(0)
Reject H_{0x} and/or H_{0y}?
No (3.1)
Run OLS regressions $\Delta z_t = \alpha_x + \beta_x \Delta x_t + e_{x,t}$ $\Delta z_t = \alpha_y + \beta_y \Delta y_t + e_{y,t}$
Is $\max(|\alpha_x|,|\beta_x|, |\alpha_y|,|\beta_y|) > \underline{\Phi}$?
No (3.1.1)
Stop: x_t and y_t are consistent
Yes
Yes (3.1.2)
Yes (3.1.2.1; 3.2.1; 3.3.2.1)
Reject both H_{0x} and H_{0y}?
No (3.2)
Test H_{0z}: $z_t \equiv y_t - z_t$ is I(1) vs. H_{1z}: z_t is I(0)
Reject H_{0z}?
No (3.1.2.2; 3.2.2; 3.3.2.2)
Stop: x_t and y_t are inconsistent
Yes (3.3)
Yes (3.3.2)
Run OLS regressions $z_t = \alpha_x + \beta_x x_t + e_{x,t}$ $z_t = \alpha_y + \beta_y y_t + e_{y,t}$
Is $\max(|\alpha_x|,|\beta_x|, |\alpha_y|,|\beta_y|) > \underline{\Phi}$?
No (3.3.1)
Stop: x_t and y_t are consistent

DATA

- Annual TFP series for 61 U.S. industries
- 3 alternative sources
  - U.S. Bureau of Labor Statistics ("BLS")
    - 1987-2020
    - "Official" TFP series for U.S.
  - U.S. Bureau of Economic Analysis ("BEA")
    - 1987-2020
  - Jorgenson, Ho, and Samuels ("JHS")
    - 1987-2014
    - Output, capital, labor, and materials in real terms
      - Analytical KLEMS-type data (series computed using harmonized definitions and aggregation procedures across industries, to make individual series comparable across industries)
    - TFP index = (real output)/(real input), where real input = Törnqvist input index obtained from capital, labor, and intermediate input series in JHS database

DATA

Table 1. List of industries analyzed, in alphabetical order

Accommodation	Motor vehicles bodies and trailers and parts
Administrative and support services	Nonmetallic mineral products
Air transportation	Oil and gas extraction
Ambulatory health care services	Other services except government
Amusements gambling and recreation industries	Other transportation and support activities
Apparel and leather and allied products	Other transportation equipment
Broadcasting and telecommunications	Paper products
Chemical products	Performing arts spectator sports museums and related activities
Computer and electronic products	Petroleum and coal products
Computer systems design and related services	Pipeline transportation
Construction	Plastics and rubber products
Educational services	Primary metals
Electrical equipment appliances and components	Printing and related support activities
Fabricated metal products	Publishing industries (includes software)
Farms	Rail transportation
Federal Reserve banks credit intermediation and related activities	Real estate

DATA

Food and beverage and tobacco products

Food services and drinking places

Forestry fishing and related activities

Funds trusts and other financial vehicles

Furniture and related products

Hospitals Nursing and residential care facilities

Information and data processing services

Insurance carriers and related activities

Legal services

Machinery

Management of companies and enterprises

Mining except oil and gas

Miscellaneous manufacturing

Miscellaneous professional scientific and technical services

Motion picture and sound recording industries

Rental and leasing services and lessors of intangible assets

Retail Trade

Securities commodity contracts and investments

Social assistance

Support activities for mining

Textile mills and textile product mills

Transit and ground passenger transportation

Truck transportation

Utilities

Warehousing and storage

Waste management and remediation services

Water transportation

Wholesale Trade

Wood products

DATA

- Each series is converted to logs
- Original TFP series are indexes with different base years
  - To err in favor of finding series consistent with each other, each log series is demeaned to conduct each pairwise analysis

DATA

- Each series is converted to logs
- Original TFP series are indexes with different base years
  - To err in favor of finding series consistent with each other, each log series is demeaned to conduct each pairwise analysis
    - Average value = 0 for each series in pairwise analyses

RESULTS

Table 1. Number of FTP series pairs corresponding to each decision and outcome

Decisions and Outcomes	**BLS vs. BEA**	**BLS vs. JHS**	**BEA vs. JHS**
Test H_{0x}: x_t is I(2) vs. H_{1x}: x_t is I(1) and			
H_{0y}: y_t is I(2) vs. H_{1y}: y_t is I(1)			
1. Neither H_{0x} nor H_{0y} is rejected	3	5	6
Run OLS regressions $\Delta^2 z_t = \alpha_x + \beta_x \Delta^2 x_t + e_{x,t}$			
and $\Delta^2 z_t = \alpha_y + \beta_y \Delta^2 y_t + e_{y,t}$			
1.1. $\max(\lvert\alpha_x\rvert,\lvert\beta_x\rvert,\lvert\alpha_y\rvert,\lvert\beta_y\rvert) \le \Phi$ ($\Rightarrow$ Consistent)	2	2	4
1.2. $\max(\lvert\alpha_x\rvert,\lvert\beta_x\rvert,\lvert\alpha_y\rvert,\lvert\beta_y\rvert) > \Phi$	1	3	2
Test H_{0z}: z_t is I(1) vs. H_{1z}: z_t is I(0)			
1.2.1. H_{0z} is rejected ($\Rightarrow$ Consistent)	0	1	0
1.2.2. H_{0z} is not rejected ($\Rightarrow$ Inconsistent)	1	2	2
2. Only one of H_{0x} and H_{0y} is rejected	4	8	9
Test H_{0z}: z_t is I(1) vs. H_{1z}: z_t is I(0)			
2.1. H_{0z} is rejected ($\Rightarrow$ Consistent)	0	1	2
2.2. H_{0z} is not rejected ($\Rightarrow$ Inconsistent)	4	7	7

RESULTS

Table 1. Number of FTP series pairs corresponding to each decision and outcome

Decisions and Outcomes	BLS vs. BEA	BLS vs. JHS	BEA vs. JHS
3. Both H_{0x} and H_{0y} are rejected	54	48	46
Test H_{0x}: x_t is I(1) vs. H_{1x}: x_t is I(0) and			
H_{0y}: y_t is I(1) vs. H_{1y}: y_t is I(0)			
3.1. Neither H_{0x} nor H_{0y} is rejected	29	28	34
Run OLS regressions $\Delta z_t = \alpha_x + \beta_x \Delta x_t + e_{x,t}$			
and $\Delta z_t = \alpha_y + \beta_y \Delta y_t + e_{y,t}$			
3.1.1. $\max(\lvert\alpha_x\rvert,\lvert\beta_x\rvert,\lvert\alpha_y\rvert,\lvert\beta_y\rvert) \le \Phi$ ($\Rightarrow$ Consistent)	16	9	27
3.1.2. $\max(\lvert\alpha_x\rvert,\lvert\beta_x\rvert,\lvert\alpha_y\rvert,\lvert\beta_y\rvert) > \Phi$	13	19	7
Test H_{0z}: z_t is I(1) vs. H_{1z}: z_t is I(0)			
3.1.2.1. H_{0z} is rejected ($\Rightarrow$ Consistent)	2	7	1
3.1.2.2. H_{0z} is not rejected ($\Rightarrow$ Inconsistent)	11	12	6
3.2. Only one of H_{0x} and H_{0y} is rejected	11	9	5
Test H_{0z}: z_t is I(1) vs. H_{1z}: z_t is I(0)			
3.2.1. H_{0z} is rejected ($\Rightarrow$ Consistent)	3	2	0
3.2.2. H_{0z} is not rejected ($\Rightarrow$ Inconsistent)	8	7	5
3.3. Both H_{0x} and H_{0y} are rejected	14	11	7
Run OLS regressions $z_t = \alpha_x + \beta_x x_t + e_{x,t}$ and			
$z_t = \alpha_y + \beta_y y_t + e_{y,t}$			
3.3.1. $\max(\lvert\alpha_x\rvert,\lvert\beta_x\rvert,\lvert\alpha_y\rvert,\lvert\beta_y\rvert) \le \Phi$ ($\Rightarrow$ Consistent)	7	6	5
3.3.2. $\max(\lvert\alpha_x\rvert,\lvert\beta_x\rvert,\lvert\alpha_y\rvert,\lvert\beta_y\rvert) > \Phi$	7	5	2
Test H_{0z}: z_t is I(1) vs. H_{1z}: z_t is I(0)			
3.3.2.1. H_{0z} is rejected ($\Rightarrow$ Consistent)	2	2	1
3.3.2.2. H_{0z} is not rejected ($\Rightarrow$ Inconsistent)	5	3	1

RESULTS

Number of FTP series pairs corresponding to each outcome

Outcomes	BLS vs. BEA	BLS vs. JHS	BEA vs. JHS
Inconsistent	29	31	21
Consistent	32	30	40
B.1. x_t and y_t are I(2)	2	4	6
B.2. x_t and y_t are I(1)	21	18	28
B.3. x_t and y_t are I(0)	9	8	6

RESULTS

Table 1. FTP series pairs corresponding to each outcome

Industry	BLS vs. BEA	BLS vs. JHS	BEA vs. JHS
Accommodation	3.2.2	3.2.2	3.1.1
Administrative and support services	3.1.2.2	3.1.2.2	3.1.2.2
Air transportation	1.2.2	2.2	3.1.1
Ambulatory health care services	3.3.2.2	3.2.2	3.1.2.2
Amusements gambling and recreation industries	3.1.1	3.1.2.1	3.1.2.2
Apparel and leather and allied products	3.2.2	3.1.2.2	3.1.1
Broadcasting and telecommunications	3.1.2.2	2.1	2.2
Chemical products	3.1.2.1	1.2.1	1.1
Computer and electronic products	**1.1**	**1.1**	**1.1**
Computer systems design and related services	3.1.1	2.2	2.2
Construction	2.2	2.2	2.1
Educational services	**3.1.1**	**3.1.1**	**3.1.1**
Electrical equipment appliances and components	3.2.1	3.3.2.2	3.1.1
Fabricated metal products	3.1.2.2	2.2	2.2
Farms	**3.1.1**	**3.1.1**	**3.1.1**
Fed. Reserve banks credit intermediation and rel. activities	**3.3.1**	**3.3.1**	**3.3.1**
Food and beverage and tobacco products	3.3.1	3.3.1	3.1.1
Food services and drinking places	3.1.2.2	3.3.2.1	3.2.2
Forestry, fishing and related activities	3.1.1	3.2.2	3.1.1
Funds, trusts and other financial vehicles	2.2	3.2.2	2.2
Furniture and related products	3.3.1	2.2	1.2.2
Hospitals Nursing and residential care facilities	3.3.2.2	3.3.2.2	3.3.1
Information and data processing services	2.2	3.1.2.2	3.2.2

RESULTS

Table 1. FTP series pairs corresponding to each outcome

Industry	BLS vs. BEA	BLS vs. JHS	BEA vs. JHS
Insurance carriers and related activities	3.2.2	3.1.2.1	3.1.1
Legal services	3.1.1	2.2	2.1
Machinery	3.3.1	3.1.1	3.1.1
Management of companies and enterprises	3.3.2.2	2.2	2.2
Mining except oil and gas	3.1.1	3.1.2.2	3.1.1
Miscellaneous manufacturing	3.1.2.2	3.1.2.1	3.1.1
Miscellaneous professional, scientific and technical services	3.1.1	3.1.2.1	3.1.1
Motion picture and sound recording industries	3.1.1	3.1.2.2	3.1.2.2
Motor vehicles bodies and trailers and parts	3.2.1	3.1.2.1	3.1.1
Nonmetallic mineral products	3.2.2	3.3.1	3.3.1
Oil and gas extraction	**3.1.1**	**3.1.1**	**3.1.1**
Other services except government	3.1.2.2	3.1.2.1	2.2
Other transportation and support activities	3.1.1	3.1.2.2	3.1.2.2
Other transportation equipment	**3.1.1**	**3.1.1**	**3.1.1**
Paper products	3.3.2.2	3.2.2	3.1.1
Perform. arts, spectator sports, museums and related activities	3.1.2.2	3.1.2.2	2.2
Petroleum and coal products	3.3.2.1	3.2.1	3.1.1
Pipeline transportation	**3.1.1**	**3.1.1**	**3.1.1**
Plastics and rubber products	3.2.1	3.2.2	3.1.1
Primary metals	3.1.2.2	3.1.2.2	3.1.1

RESULTS

Table 1. FTP series pairs corresponding to each outcome

Industry	BLS vs. BEA	BLS vs. JHS	BEA vs. JHS
Printing and related support activities	3.1.2.1	3.1.2.2	3.1.1
Publishing industries (includes software)	**3.1.1**	**3.1.1**	**3.1.1**
Rail transportation	3.3.1	3.3.2.2	3.3.2.2
Real estate	3.1.2.2	1.2.2	1.1
Rental and leasing services and lessors of intangible assets	**1.1**	**1.1**	**1.1**
Retail Trade	3.2.2	1.2.2	1.2.2
Securities, commodity contracts and investments	3.1.1	3.3.1	3.1.1
Social assistance	3.2.2	3.3.2.1	3.3.2.1
Support activities for mining	3.1.2.2	3.1.2.2	3.1.1
Textile mills and textile product mills	3.1.2.2	3.1.2.2	3.1.1
Transit and ground passenger transportation	3.3.1	3.1.2.2	3.1.2.2
Truck transportation	**3.3.1**	**3.3.1**	**3.3.1**
Utilities	3.3.2.1	3.2.2	3.2.2
Warehousing and storage	3.2.2	3.1.2.1	3.1.2.1
Waste management and remediation services	3.2.2	3.2.1	3.2.2
Water transportation	**3.1.1**	**3.1.1**	**3.1.1**
Wholesale Trade	3.3.2.2	3.1.1	3.2.2
Wood products	2.2	3.3.1	3.3.1

CONCLUSIONS

Between 1/3 and 1/2 of U.S. industries have TFP series that are unrelated in the long-run

- At least one of the series must be of poor quality
- How confident should we be about conclusions obtained from studies based on them?
- How about TFP series for countries whose data are likely to be even les reliable?
- Why are so few studies investigating this issue (relative to studies using these data to obtain dubious results)