May 6, 2005 - p. 1/31

and the Creation of the Quarterly Workforce The LEHD Infrastructure Files Indicators

John M. Abowd ♣, ♣, Bryce E. Stephens ♣ and Lars Vilhuber ♣

♠ Cornell University

. U.S. Census Bureau, LEHD Program

USCENSUSBUREAU

Introduction

Disclosure-proofing the QWI

Forming Aggregated Estimates: QWI

Infrastructure Files

Publicly available files

Conclusion

The LEHD Infrastructure Files

⇒ What are QWI?

Introduction

Oln this paper

Input Files

⇒ What is it?

What are QWI?

The LEHD Infrastructure Files

Since 2003: publication of Quarterly Workforce Indicators

⇒ What are QWI?

⇒ What is it?

⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files

■ Since 2003: publication of Quarterly Workforce Indicators

The first 21st century statistical system

⊃What is it?

⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

What are QWI?

The LEHD Infrastructure Files

Since 2003: publication of Quarterly Workforce Indicators

The first 21st century statistical system

No additional burden

⇒ What are QWI?

What is it?

⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files

Since 2003: publication of Quarterly Workforce Indicators

The first 21st century statistical system

No additional burden

improve the data

Extensive use of modern statistics to integrate and

⇒ What are QWI?

⇒ In this paper ⇒ What is it?

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

USCENSUSBUREAU

May 6, 2005 - p. 3/31

⇒ In this paper SWhat is it?

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

Since 2003: publication of Quarterly Workforce Indicators

The first 21st century statistical system

No additional burden

Extensive use of modern statistics to integrate and improve the data

State-of-the-art confidentiality protection methods

⇒ What is it?
⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

Since 2003: publication of Quarterly Workforce Indicators

■ The first 21st century statistical system

No additional burden

Extensive use of modern statistics to integrate and improve the data

State-of-the-art confidentiality protection methods

Innovative use of wage records to constitute a frame to integrate data

⊃What are QWI?

What is it?In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

Since 2003: publication of Quarterly Workforce Indicators

The first 21st century statistical system

◆ No additional burden

Extensive use of modern statistics to integrate and improve the data

State-of-the-art confidentiality protection methods

Innovative use of wage records to constitute a frame to integrate data

The first statistical system to use "jobs" as a frame

The LEHD Infrastructure Files

Combines

Introduction What are QWI?

OWhat are QW

⊃In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files

Combines

⊃ What are QWI?

(state) administrative records data on workers (UI Wage

○ What is it?

records)

⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files

Combines

⇒ What are QWI?

◆ (state) administrative records data on workers (UI Wage

(state) administrative records data on firms (QCEW aka

OWhat is it?

records)

⇒ In this paper

Infrastructure Files

Input Files

ES-202)

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files

⇒ What are QWI?

OWhat is it?

⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

Combines

 ◆ (state) administrative records data on workers (UI Wage records)

(state) administrative records data on firms (QCEW aka ES-202)

administrative information on demographics

The LEHD Infrastructure Files

SWhat are QWI?

What is it?

In this paper

Infrastructure Files

Input Files

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

Combines

 ◆ (state) administrative records data on workers (UI Wage records)

(state) administrative records data on firms (QCEW aka ES-202)

administrative information on demographics

surveys on people and firms collected by Census Bureau

3 What are QWI?

Oln this paper ⇒ What is it?

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Combines

- ◆ (state) administrative records data on workers (UI Wage records)
- (state) administrative records data on firms (QCEW aka ES-202)
- administrative information on demographics
- surveys on people and firms collected by Census Bureau
- careful longitudinal edit of person identifiers and economic

Files	
acture	
frastru	
HD In	
he LE	

⇒ What are QWI?

OWhat is it?
Oin this paper

Input Files

Infrastructure Files

Forming Aggregated

Estimates: QWI

.

D

Publicly available files

Conclusion

Combines

- ◆ (state) administrative records data on workers (UI Wage records)
- (state) administrative records data on firms (QCEW aka ES-202)
- administrative information on demographics
- surveys on people and firms collected by Census Bureau
- careful longitudinal edit of person identifiers and economic
- careful longitudinal edit of person and firm characteristics

The LEHD Infrastructure Files Introduction

Describe the construction of the LEHD infrastructure

⇒ What are QWI?

⇒ What is it?

⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

.

Publicly available files

The LEHD Infrastructure Files

Describe the construction of the LEHD infrastructure

... in particular the imputation mechanisms used

⇒ What are QWI?

⊃What is it?

⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files

Describe the construction of the LEHD infrastructure

... in particular the imputation mechanisms used

Describe the computation of the QWI statistics

⇒ What are QWI?

⊃What is it?

⊃ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files

Describe the construction of the LEHD infrastructure

... in particular the imputation mechanisms used

... in particular the imputation mechanisms used

Describe the computation of the QWI statistics

⊃What are QWI?

⇒ What is it?
⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files

Describe the construction of the LEHD infrastructure

... in particular the imputation mechanisms used

... in particular the imputation mechanisms used

Describe the disclosure-proofing mechanism

Describe the computation of the QWI statistics

OWhat are QWI?

⇒ What is it?⇒ In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files

O What are QWI?

OWhat is it? O In this paper

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

Describe the construction of the LEHD infrastructure

... in particular the imputation mechanisms used

Describe the computation of the QWI statistics

... in particular the imputation mechanisms used

Describe the disclosure-proofing mechanism

Describe researcher access to infrastructure files and confidential QWI files

USCENSUSBUREAU

Input Files

The LEHD Infrastructure Files Introduction

Input Files > Wage records: UI

⇒ Employer reports: ES202

⇒ Employer reports: Employer seports: Employer reports: Employ

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Wage records: Ul

The LEHD Infrastructure Files Introduction

report of an individual's UI-covered earnings by an

⇒ Wage records: UI Input Files

employing entity

⇒ Employer reports: ES202

⇒ Demographics

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Wage records: UI

The LEHD Infrastructure Files Introduction

Input Files ⊃Wage records: UI

⇒ Employer reports: ES202

⇒ Demographics

during the quarter

employing entity

appears if at least one dollar was earned by that individual

■ report of an individual's UI-covered earnings by an

Infrastructure Files

Ofcoord A Science

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Wage records: UI

The LEHD Infrastructure Files Introduction

Input Files ⇒ Wage records: UI

⊃ Employer reports: ES202

appears if at least one dollar was earned by that individual

■ report of an individual's UI-covered earnings by an

identifies EARNINGS, EMPLOYER, TIME PERIOD

⇒ Demographics

during the quarter

employing entity

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

Wage records: Ul

The LEHD Infrastructure Files Introduction

employing entity

⇒ Employer reports: ES202 ⇒ Wage records: UI

appears if at least one dollar was earned by that individual

■ report of an individual's UI-covered earnings by an

some limited other state-dependent information available

■ identifies EARNINGS, EMPLOYER, TIME PERIOD

⇒ Demographics

during the quarter

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

Wage records: Ul

The LEHD Infrastructure Files

⇒ Employer reports: ES202 ⇒ Wage records: UI

Demographics

Infrastructure Files

Forming Aggregated Estimates: QWI

Publicly available files

Disclosure-proofing the QWI

Conclusion

 report of an individual's UI-covered earnings by an employing entity appears if at least one dollar was earned by that individual during the quarter

identifies EARNINGS, EMPLOYER, TIME PERIOD

some limited other state-dependent information available

■ in particular, for Minnesota, the ESTABLISHMENT is reported

USCENSUSBUREAU

The LEHD Infrastructure Files

... or QCEW

Employer reports: ES202

Introduction

⇒ Wage records: UI Input Files

D Employer reports: ES202

⇒ Demographics

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files

collected as part of the Covered Employment and Wages

(CEW) (administered by the BLS)

Introduction Input Files

⇒ Wage records: UI

⇒ Employer reports: ES202

⇒ Demographics

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

collected as part of the Covered Employment and Wages

(CEW) (administered by the BLS)

Also used as the inputs to the Business Employment

Input Files

⇒ Employer reports: ES202 ⇒ Wage records: UI

⇒ Demographics

Dynamics (BED)

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

Input Files

⇒ Wage records: UI

⇒ Employer reports: ES202

⇒ Demographics

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

collected as part of the Covered Employment and Wages (CEW) (administered by the BLS)

Also used as the inputs to the Business Employment Dynamics (BED) collects from employers covered by state unemployment insurance programs:

employment

payroll

geographic information

The LEHD Infrastructure Files Introduction

Input Files

Wage records: UI

Demployer reports: ES202

Infrastructure Files

⇒ Demographics

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

collected as part of the Covered Employment and Wages (CEW) (administered by the BLS)

Also used as the inputs to the Business Employment Dynamics (BED) collects from employers covered by state unemployment insurance programs:

employment

payroll

◆ geographic information

■ fundamental unit: 'reporting unit' (≈ establishment)

The LEHD Infrastructure Files

Input Files

⇒ Employer reports: ES202 ⇒ Wage records: UI

⇒ Demographics

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

collected as part of the Covered Employment and Wages (CEW) (administered by the BLS)

Also used as the inputs to the Business Employment Dynamics (BED) collects from employers covered by state unemployment insurance programs:

employment

payroll

geographic information

fundamental unit: 'reporting unit' (≈ establishment)

One report per establishment per quarter is filed

Demographics

The LEHD Infrastructure Files Introduction

Demographics are taken from a number of Census-internal

files derived from administrative data:

Person Characteristics File (PCF)

Census Numident

Input Files ⊃Wage records: UI

D Employer reports: ES202

⇒ Demographics

Infrastructure Files

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

Demographics

The LEHD Infrastructure Files Introduction

⇒ Wage records: UI Input Files

⇒ Employer reports: ES202

⇒ Demographics

Infrastructure Files

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

Demographics are taken from a number of Census-internal files derived from administrative data:

- Person Characteristics File (PCF)
- ◆ Census Numident
- Where available, more detailed data on individuals is also extracted from surveys and censuses:
- CPS
- SIPP
- ACS
- 1990 Census
- 2000 Census

USCENSUSBUREAU

Infrastructure Files

The LEHD Infrastructure Files

Introduction

Input Files

DEHF: Employment History

Infrastructure Files

⇒ GAL: Geocoded Address

Characteristics File DECF: Employer Characteristics File

CF: Individual

Disclosure-proofing the QWI

Forming Aggregated

⇒ Flow so far

Estimates: QWI

Publicly available files

EHF: Employment History Files

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

DEHF: Employment History

Characteristics File ⇒ICF: Individual

⇒ ECF: Employer

⇒ GAL: Geocoded Address Characteristics File

⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

■ Job-level EHF

- complete in-state work history for each individual on Ulwage records.
- one record for each employee-employer combination a
- earnings and employment patterns

EHF: Employment History Files

The LEHD Infrastructure Files

Input Files

Infrastructure Files

DEHF: Employment History

⇒ICF: Individual

Characteristics File ⇒ ECF: Employer

OGAL: Geocoded Address Characteristics File

⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

■ Job-level EHF

- ◆ complete in-state work history for each individual on Ulwage records.
- one record for each employee-employer combination a
- earnings and employment patterns
- Employer and establishment-level employment history
- QCEW-based employment-activity history for every SEIN (employer) and SEINUNIT (establishment)

EHF: Employment History Files

The LEHD Infrastructure Files

Input Files

Infrastructure Files

⇒EHF: Employment History

Characteristics File CE: Individual

DECF: Employer

OGAL: Geocoded Address Characteristics File

> Flow so far

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

■ Job-level EHF

- ◆ complete in-state work history for each individual on Ulwage records.
- one record for each employee-employer combination a
- earnings and employment patterns
- Employer and establishment-level employment history
- QCEW-based employment-activity history for every SEIN (employer) and SEINUNIT (establishment)
- Comparison of employment and activity of SEINs between UI and QCEW files is done for QA purposes, and in preparation of weighting.

The LEHD Infrastructure Files

Demographic information from the PCF is merged with

universe of PIKs from wage records

Introduction

Input Files

Infrastructure Files

⇒EHF: Employment History

Characteristics File ⇒ICF: Individual

Characteristics File **DECF:** Employer

⇒ GAL: Geocoded Address

⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

Demographic information from the PCF is merged with

Input Files

Infrastructure Files

records without a valid match flagged

universe of PIKs from wage records

⇒EHF: Employment History

Characteristics File ⇒ICF: Individual

DECF: Employer

Characteristics File

⇒ GAL: Geocoded Address

⇒ Flow so far

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

Demographic information from the PCF is merged with

Input Files

Infrastructure Files

DEHF: Employment History

CPS and SIPP identifiers are merged on.

records without a valid match flagged

universe of PIKs from wage records

Characteristics File ⇒ICF: Individual

⇒ ECF: Employer

Characteristics File

⇒ GAL: Geocoded Address

⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Demographic information from the PCF is merged with

Input Files

Infrastructure Files

DEHF: Employment History

Characteristics File ⇒ICF: Individual

⇒ ECF: Employer

... gender, education, and age information from the CPS

CPS and SIPP identifiers are merged on.

records without a valid match flagged

universe of PIKs from wage records

Characteristics File

⇒ GAL: Geocoded Address

⇒ Flow so far

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

Demographic information from the PCF is merged with

Input Files

Infrastructure Files

DEHF: Employment History

Characteristics File ⇒ICF: Individual

... gender, education, and age information from the CPS

CPS and SIPP identifiers are merged on.

records without a valid match flagged

universe of PIKs from wage records

⇒ GAL: Geocoded Address Characteristics File ⇒ ECF: Employer

⇒ Flow so far

Data completion

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Demographic information from the PCF is merged with

Input Files

Infrastructure Files

DEHF: Employment History

Characteristics File ⇒ICF: Individual

... gender, education, and age information from the CPS

CPS and SIPP identifiers are merged on.

records without a valid match flagged

universe of PIKs from wage records

Characteristics File ⇒ ECF: Employer

⇒ GAL: Geocoded Address

Data completion

Age

⇒ Flow so far

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files

Input Files

DEHF: Employment History Infrastructure Files

⇒ICF: Individual

Characteristics File

Characteristics File ⇒ ECF: Employer

⇒ GAL: Geocoded Address ⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

Demographic information from the PCF is merged with universe of PIKs from wage records

records without a valid match flagged

CPS and SIPP identifiers are merged on.

... gender, education, and age information from the CPS

Data completion

◆ Age

Gender

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files DEHF: Employment History

Files

CFI - Limpoyment

Characteristics File

D ECF: EmployerCharacteristics File

Characteristics File ⇒ GAL: Geocoded Address

Forming Aggregated

⇒ Flow so far

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

Demographic information from the PCF is merged with universe of PIKs from wage records

records without a valid match flagged

CPS and SIPP identifiers are merged on.

... gender, education, and age information from the CPS

■ Data completion

Age

Gender

Education

The LEHD Infrastructure Files

Input Files

Infrastructure Files

⇒EHF: Employment History

Characteristics File ⇒ICF: Individual

DECF: Employer

⇒ GAL: Geocoded Address Characteristics File

⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Demographic information from the PCF is merged with universe of PIKs from wage records

records without a valid match flagged

CPS and SIPP identifiers are merged on.

... gender, education, and age information from the CPS

Data completion

◆ Age

Gender

Education

County of residence

The LEHD Infrastructure Files

Input Files

Infrastructure Files

⇒EHF: Employment History

⇒ICF: Individual

Characteristics File DECF: Employer

3GAL: Geocoded Address Characteristics File

⇒ Flow so far

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Demographic information from the PCF is merged with universe of PIKs from wage records

records without a valid match flagged

CPS and SIPP identifiers are merged on.

... gender, education, and age information from the CPS

Data completion

◆ Age

Gender

 County of residence Education

are each imputed ten times

The LEHD Infrastructure Files

 Two files: firm and establishment level, quarterly records Introduction

Input Files

Infrastructure Files

⇒EHF: Employment History

Characteristics File ⇒ICF: Individual

DECF: Employer

Characteristics File

⇒ GAL: Geocoded Address

⊃ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

■ Two files: firm and establishment level, quarterly records

Input Files

Inputs:

Infrastructure Files

⇒EHF: Employment History

Characteristics File ⇒ICF: Individual

DECF: Employer

Characteristics File

⇒ GAL: Geocoded Address

⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

■ Two files: firm and establishment level, quarterly records

Input Files

Infrastructure Files

1. ES202

■ Inputs:

⇒EHF: Employment History

FIFF. EIIIDIOYIIIEIIL Files CF: IndividualCharacteristics File

⇒ ECF: Employer

Characteristics File

OGAL: Geocoded Address

O Flow so far

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

■ Two files: firm and establishment level, quarterly records

Input Files

Infrastructure Files

1. ES202

■ Inputs:

⇒ EHF: Employment History

Files CF: Individual

2. UI: supplement information on the ES202, extend

published BLS county-level employment data

Characteristics File CECF: Employer Characteristics File CAL: Geocoded Address

List D Flow so far Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

■ Two files: firm and establishment level, quarterly records

Input Files

Infrastructure Files

1. ES202

■ Inputs:

DEHF: Employment History

Files CF: Individual

2. UI: supplement information on the ES202, extend

published BLS county-level employment data

3. GAL: establishment geocodes

Characteristics File

DECF: Employer
Characteristics File

Characteristics File CAL: Geocoded Address

List ⊃Flow so far Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Two files: firm and establishment level, quarterly records

Input Files

Infrastructure Files

1. ES202

■ Inputs:

DEHF: Employment History

2. UI: supplement information on the ES202, extend

oublished BLS county-level employment data

4. LDB (BLS) for backfilling NAICS information

3. GAL: establishment geocodes

Characteristics File ⇒ICF: Individual

Characteristics File ⇒ECF: Employer

3GAL: Geocoded Address

⇒ Flow so far

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files

Input Files

Infrastructure Files

DEHF: Employment History

Files

Characteristics File

⊃ ECF: Employer Characteristics File

OGAL: Geocoded Address

⊃ Flow so far

Forming Aggregated Estimates: QWI

stimates: Qvvi

Disclosure-proofing the QWI

Publicly available files

Conclusion

Two files: firm and establishment level, quarterly records

Inputs:

1. ES202

2. UI: supplement information on the ES202, extend oublished BLS county-level employment data

3. GAL: establishment geocodes

4. LDB (BLS) for backfilling NAICS information

Longitudinal edits for consistency and data completion

The LEHD Infrastructure Files

Input Files

Infrastructure Files

DEHF: Employment History

O ICF: Individual

Characteristics File CECF: Employer

Characteristics File CAL: Geocoded Address

List

S Flow so far

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

onclusion

Two files: firm and establishment level, quarterly records

Inputs:

1. ES202

2. UI: supplement information on the ES202, extend oublished BLS county-level employment data

3. GAL: establishment geocodes

4. LDB (BLS) for backfilling NAICS information

Longitudinal edits for consistency and data completion

Imputation:

impute SIC if NAICS non-missing and vice-versa

The LEHD Infrastructure Files

Input Files

Infrastructure Files

DEHF: Employment History

⇒ICF: Individual

Characteristics File ⇒ ECF: Employer

Characteristics File

GAL: Geocoded Address

⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Two files: firm and establishment level, quarterly records

Inputs:

1. ES202

2. UI: supplement information on the ES202, extend oublished BLS county-level employment data

3. GAL: establishment geocodes

4. LDB (BLS) for backfilling NAICS information

Longitudinal edits for consistency and data completion

Imputation:

◆ impute SIC if NAICS non-missing and vice-versa

unconditional impute of missing SIC and NAICS codes

The LEHD Infrastructure Files

Input Files

Infrastructure Files

DEHF: Employment History

⇒ICF: Individual

Characteristics File ⇒ ECF: Employer

Characteristics File

OGAL: Geocoded Address

> Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Two files: firm and establishment level, quarterly records

Inputs:

1. ES202

2. UI: supplement information on the ES202, extend oublished BLS county-level employment data

3. GAL: establishment geocodes

4. LDB (BLS) for backfilling NAICS information

Longitudinal edits for consistency and data completion

Imputation:

◆ impute SIC if NAICS non-missing and vice-versa

unconditional impute of missing SIC and NAICS codes

geography conditional on industry

The LEHD Infrastructure Files Introduction

... is a data set containing unique commercial and residential

Input Files

addresses

Infrastructure Files

DEHF: Employment History

Characteristics File ⇒ICF: Individual

DECF: Employer

Characteristics File

⇒ GAL: Geocoded Address

⇒ Flow so far

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

... is a data set containing unique commercial and residential

geocoded to the Census Block and latitude/longitude

Input Files

addresses

Infrastructure Files

⇒EHF: Employment History

coordinates

Files ⊃ICF: Individual

Characteristics File

⇒ ECF: Employer
Characteristics File

Characteristics File CAAL: Geocoded Address

List DFlow so far Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

... is a data set containing unique commercial and residential

geocoded to the Census Block and latitude/longitude

Input Files

addresses

Infrastructure Files

⇒EHF: Employment History

coordinates

Characteristics File ⇒ICF: Individual

Inputs:

DECF: Employer

Characteristics File

⇒ GAL: Geocoded Address

⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

... is a data set containing unique commercial and residential

geocoded to the Census Block and latitude/longitude

Input Files

addresses

Infrastructure Files

⇒EHF: Employment History

coordinates

⇒ICF: Individual

Characteristics File

■ Inputs:

Characteristics File **DECF:** Employer

⇒ GAL: Geocoded Address

1. ES202 data

⇒ Flow so far

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

DEHF: Employment History

Characteristics File ⇒ICF: Individual

⇒ ECF: Employer

⇒ GAL: Geocoded Address Characteristics File

⇒ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

... is a data set containing unique commercial and residential addresses

geocoded to the Census Block and latitude/longitude coordinates

Inputs:

1. ES202 data

2. Census Bureau's Business Register (BR)

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

DEHF: Employment History

⇒ICF: Individual

Characteristics File ⇒ ECF: Employer

Characteristics File

⇒ GAL: Geocoded Address

⊃ Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

Conclusion

... is a data set containing unique commercial and residential addresses

geocoded to the Census Block and latitude/longitude coordinates

Inputs:

1. ES202 data

2. Census Bureau's Business Register (BR)

3. Census Bureau's Master Address File (MAF)

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

⇒ EHF: Employment History Files

Characteristics File

DECF: EmployerCharacteristics File

⇒ GAL: Geocoded Address List

List D Flow so far Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

... is a data set containing unique commercial and residential addresses

geocoded to the Census Block and latitude/longitude coordinates

Inputs:

1. ES202 data

2. Census Bureau's Business Register (BR)

Census Bureau's Master Address File (MAF)

4. American Community Survey Place of Work file (ACS-POW)

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

DEHF: Employment History

OICF: Individual

Characteristics File

D ECF: EmployerCharacteristics File

⇒ GAL: Geocoded Address List

List D Flow so far Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

... is a data set containing unique commercial and residential addresses

geocoded to the Census Block and latitude/longitude coordinates

Inputs:

1. ES202 data

2. Census Bureau's Business Register (BR)

3. Census Bureau's Master Address File (MAF)

4. American Community Survey Place of Work file (ACS-POW)

Addresses are

The LEHD Infrastructure Files Introduction

... is a data set containing unique commercial and residential

geocoded to the Census Block and latitude/longitude

Input Files

addresses

Infrastructure Files

⇒ EHF: Employment History

⇒ICF: Individual

⇒ ECF: Employer

Characteristics File

⊃ Flow so far

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

1. ES202 data coordinates Inputs: ⇒ GAL: Geocoded Address Characteristics File

2. Census Bureau's Business Register (BR)

3. Census Bureau's Master Address File (MAF)

4. American Community Survey Place of Work file (ACS-POW)

Addresses are

1. geocoded

The LEHD Infrastructure Files Introduction

Input Files

⇒ EHF: Employment History Infrastructure Files

Characteristics File ⇒ICF: Individual

Characteristics File ⇒ ECF: Employer

⇒ GAL: Geocoded Address

⊃ Flow so far

Forming Aggregated Estimates: QWI

Publicly available files

Conclusion

Disclosure-proofing the QWI

... is a data set containing unique commercial and residential addresses

geocoded to the Census Block and latitude/longitude coordinates

Inputs:

1. ES202 data

2. Census Bureau's Business Register (BR)

3. Census Bureau's Master Address File (MAF)

4. American Community Survey Place of Work file (ACS-POW)

Addresses are

1. geocoded

2. standardized

The LEHD Infrastructure Files Introduction

... is a data set containing unique commercial and residential

geocoded to the Census Block and latitude/longitude

Input Files

addresses

Infrastructure Files

⇒ EHF: Employment History

coordinates

⇒ICF: Individual

Characteristics File DECF: Employer

Characteristics File

1. ES202 data

Inputs:

⇒ GAL: Geocoded Address

⊃Flow so far

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

4. American Community Survey Place of Work file

3. Census Bureau's Master Address File (MAF)

2. Census Bureau's Business Register (BR)

Publicly available files

Conclusion

Addresses are

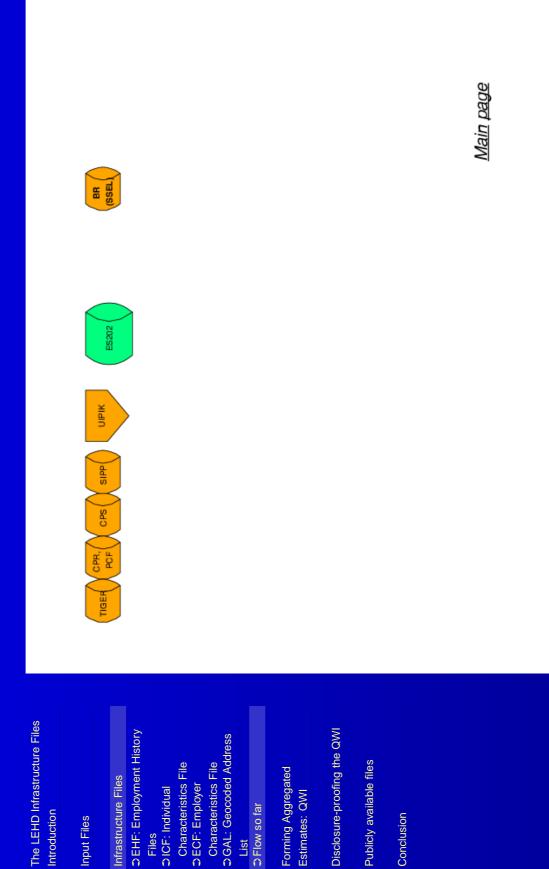
(ACS-POW)

1. geocoded

2. standardized

3. unduplicated (by firm name)

Flow so far



⇒ GAL: Geocoded Address

Forming Aggregated

⊃ Flow so far

Estimates: QWI

Characteristics File

Characteristics File

CE: Individual

DECF: Employer

Infrastructure Files

Introduction

Input Files

Publicly available files

Conclusion

Flow so far

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

DEHF: Employment History

CE: Individual

Characteristics File **DECF:** Employer

⇒ GAL: Geocoded Address Characteristics File

⊃ Flow so far

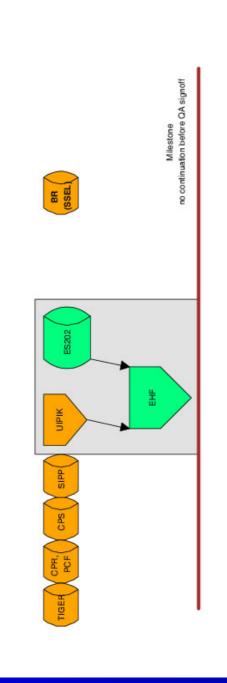
Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion



Main page

Flow so far

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

⇒EHF: Employment History

Files O ICF: Individual Characteristics File CECF: Employer Characteristics File

List O Flow so far

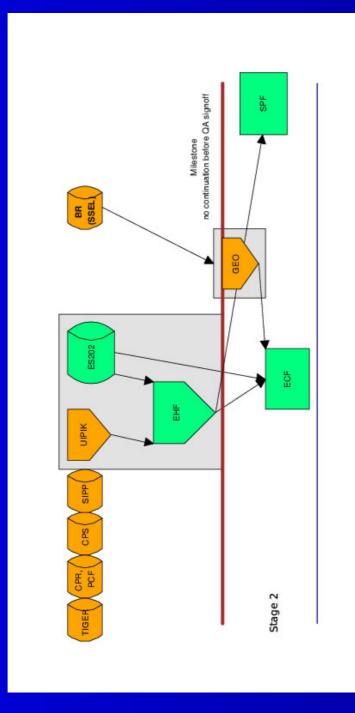
⇒ GAL: Geocoded Address

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion



Main page

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows ⇒ Solution:

Successor-Predecessor

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Simplementation

⇒ Computing the statistics O Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files

Firm identifier:

Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor ⇒ Solution:

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

S Probability Model

⇒ Computing the statistics O Implementation

O Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files

■ Firm identifier: state-specific account number

Introduction

Infrastructure Files

Input Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor ⇒ Solution:

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Implementation

⇒ Computing the statistics O Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Firm identifier:

Input Files

Account numbers can and do change:

Infrastructure Files

Forming Aggregated

Estimates: QWI

⇒ Correction of spurious worker flows

Successor-Predecessor ⇒ Solution:

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Implementation

⇒ Computing the statistics O Implementation

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

Firm identifier:

Input Files

Account numbers can and do change:

change in legal form

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious worker flows

Successor-Predecessor ⇒ Solution:

Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Implementation

⇒ Computing the statistics O Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Firm identifier:

Input Files

Account numbers can and do change:

◆ change in legal form

a merger

Infrastructure Files

⇒ Correction of spurious Forming Aggregated Estimates: QWI

worker flows ⇒ Solution:

Successor-Predecessor

⇒ Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Implementation

⇒ Computing the statistics O Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Firm identifier:

Input Files

Account numbers can and do change:

◆ change in legal form

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious worker flows

Change in firm identifier

a merger

Successor-Predecessor ⇒ Solution:

⇒ Attaching establishment

characteristics to jobs

DU2W: Unit to Worker

⇒ Probability Model Implementation ⇒ Computing the statistics

Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Firm identifier:

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious worker flows Successor-Predecessor

⇒ Attaching establishment characteristics to jobs

⇒ U2W: Unit to Worker

Seriod Probability Model

Simplementation

Computing the statistics Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

Change in firm identifier is the component determining when Account numbers can and do change: a worker changes employers ◆ change in legal form ◆ a merger

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious worker flows

Solution:

Successor-Predecessor

Attaching establishment characteristics to jobs

O U2W: Unit to Worker

UZW: Unit to V

Probability ModelImplementation

O Implementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

■ Firm identifier:

Account numbers can and do change:

◆ change in legal form

◆ a merger

■ Change in <u>firm identifier</u>

■ → non-economic change in identifier creates spurious flow

Solution: Successor-Predecessor File

The LEHD Infrastructure Files

track large worker movements between SEINs

Introduction

Infrastructure Files

Input Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Implementation

⇒ Computing the statistics O Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

Solution: Successor-Predecessor File

The LEHD Infrastructure Files Introduction

Input Files

■ → link entities that have different account numbes, but

constitute the same economic entitiy

track large worker movements between SEINs

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor Attaching establishment

characteristics to jobs

DU2W: Unit to Worker

Seriod Probability Model

Simplementation Simplementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

Solution: Successor-Predecessor File

The LEHD Infrastructure Files Introduction

Input Files

→ link entities that have different account numbes, but

track large worker movements between SEINs

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious worker flows Successor-Predecessor

characteristics to jobs

Attaching establishment

⇒ U2W: Unit to Worker

> Probability Model

Umplementation

Computing the statistics ⇒ Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

relative terms, and the number of workers entering an SEIN, SPF provides a variety of link characteristics, based on the number of workers leaving an SEIN, in both absolute and constitute the same economic entitly again in absolute and relative terms.

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious

Successor-Predecessor worker flows

Attaching establishment characteristics to jobs

⇒ U2W: Unit to Worker

⇒ Probability Model

Computing the statistics Implementation Implementation

Disclosure-proofing the QWI

Conclusion

Publicly available files

track large worker movements between SEINs

■ → link entities that have different account numbes, but constitute the same economic entitly

relative terms, and the number of workers entering an SEIN, SPF provides a variety of link characteristics, based on the number of workers leaving an SEIN, in both absolute and again in absolute and relative terms.

observed to move to a single successor, and that successor absorbs 80% of its employees from a single predecessor, QWI: if 80% of an SEIN's workers (the predecessor) are then all flows between those two account numbers are filtered out, and treated as if they had never existed.

The LEHD Infrastructure Files

Goal: achieve a high level of accuracy and detail

Introduction

Infrastructure Files

Input Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor

⇒ Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Implementation

⇒ Computing the statistics O Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Problem: no establishment identification on wage record

Goal: achieve a high level of accuracy and detail

Infrastructure Files

Forming Aggregated

Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

⇒ U2W: Unit to Worker

⇒ Probability Model Simplementation

Computing the statistics Simplementation

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

Goal: achieve a high level of accuracy and detail

Input Files

Problem:

Infrastructure Files

Forming Aggregated

⇒ Correction of spurious Estimates: QWI

worker flows

Successor-Predecessor ⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Implementation

⇒ Computing the statistics O Implementation

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

Input Files

Problem:

Infrastructure Files

30-40% of state-wide employment in multi-establishment

Goal: achieve a high level of accuracy and detail

Forming Aggregated Estimates: QWI

firms

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment

characteristics to jobs

⇒ U2W: Unit to Worker

⇒ Probability Model

Implementation

Computing the statistics O Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

USCENSUSBUREAU

May 6, 2005 - p. 19/31

The LEHD Infrastructure Files Introduction

Input Files

Problem:

Infrastructure Files

■ 30-40% of state-wide employment in multi-establishment

Goal: achieve a high level of accuracy and detail

Solution: probability model for employment location and

Forming Aggregated Estimates: QWI

firms

⇒ Correction of spurious

worker flows

imputation

⇒ Attaching establishment Successor-Predecessor

characteristics to jobs

⇒ U2W: Unit to Worker

⇒ Probability Model Simplementation Computing the statistics

Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

May 6, 2005 - p. 19/31

The LEHD Infrastructure Files Introduction

Goal: achieve a high level of accuracy and detail

Input Files

Problem:

Infrastructure Files

30-40% of state-wide employment in multi-establishment

Solution: probability model for employment location and

Forming Aggregated Estimates: QWI

firms

⇒ Correction of spurious worker flows

Successor-Predecessor

imputation

Attaching establishment characteristics to jobs

Key elements are:

⇒ U2W: Unit to Worker

⇒ Probability Model

Simplementation

Computing the statistics ⇒ Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated
Estimates: QWI

Correction of spurious worker flows

Successor-Predecessor

Successor Tredecessor File

Staching establishment

characteristics to jobs

OU2W: Unit to Worker

⇒ Probability Model

Simplementation

ImplementationComputing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

Goal: achieve a high level of accuracy and detail

Problem:

30-40% of state-wide employment in multi-establishment firms

Solution: probability model for employment location and imputation

Key elements are:

1. distance between place-of-work and place-of-residence

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

⇒ U2W: Unit to Worker

Seriod Probability Model

Computing the statistics ⇒ Implementation

Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

Goal: achieve a high level of accuracy and detail

Problem:

30-40% of state-wide employment in multi-establishment firms

Solution: probability model for employment location and imputation

Key elements are:

1. distance between place-of-work and place-of-residence

2. distribution of employment across establishments of multi-establishment firms.

The LEHD Infrastructure Files ntroduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious worker flows Successor-Predecessor

Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

> Probability Model

Computing the statistics Simplementation Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

Goal: achieve a high level of accuracy and detail

Problem:

30-40% of state-wide employment in multi-establishment firms

Solution: probability model for employment location and imputation

Key elements are:

1. distance between place-of-work and place-of-residence

distribution of employment across establishments of multi-establishment firms.

Important practical aspects:

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated

Correction of spurious Estimates: QWI worker flows Successor-Predecessor

Attaching establishment

characteristics to jobs

⇒ U2W: Unit to Worker

> Probability Model

Computing the statistics Simplementation

Umplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

Goal: achieve a high level of accuracy and detail

Problem:

30-40% of state-wide employment in multi-establishment firms

Solution: probability model for employment location and imputation

Key elements are:

1. distance between place-of-work and place-of-residence

distribution of employment across establishments of multi-establishment firms.

Important practical aspects:

Non-ignorable missing data imputation

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Correction of spurious Forming Aggregated Estimates: QWI

worker flows

Successor-Predecessor

Attaching establishment

characteristics to jobs ⇒ U2W: Unit to Worker

⇒ Probability Model

Computing the statistics Implementation

Umplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

Goal: achieve a high level of accuracy and detail

Problem:

30-40% of state-wide employment in multi-establishment firms

Solution: probability model for employment location and imputation

Key elements are:

1. distance between place-of-work and place-of-residence

distribution of employment across establishments of multi-establishment firms.

Important practical aspects:

Non-ignorable missing data imputation

Several million imputations every quarter

The LEHD Infrastructure Files Introduction

lacksquare workers i=1,...,.

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor ⇒ Solution:

Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

O Implementation Simplementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

lacksquare workers i=1,...,l

lacksquare firms j=1,...,J

Input Files

Infrastructure Files

Forming Aggregated

Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor ⇒ Solution:

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

O Implementation

O Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

lacksquare workers i=1,...,I

lacksquare firms j=1,...,J

Input Files

Infrastructure Files

lacksquare active establishments at firm j R_{jt}

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

⇒ Solution:

Successor-Predecessor

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

O Implementation Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

lacksquare workers i=1,...,I

firms j=1,...,J

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

lacktriangle quarter t employment of establishment r in firm j N_{jrt}

lacktriangle active establishments at firm j R_{jt}

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

O Implementation Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

lacksquare workers i=1,...,I

firms j = 1, ..., J

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

quarter t employment of establishment r in firm $j \ N_{jrt}$

active establishments at firm $j R_{jt}$

 \blacksquare y_{ijt} establishment at which i was employed

⇒ Correction of spurious worker flows Successor-Predecessor

Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

Seropability Model

Implementation

Computing the statistics

Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

lacksquare workers i=1,...,I

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Disclosure-proofing the QWI

Publicly available files

Conclusion

Computing the statistics

Simplementation Implementation

lacksquare firms active

active establishments at firm $j R_{jt}$ firms j=1,...,J

quarter t employment of establishment r in firm $j \ N_{jrt}$ y_{ijt} establishment at which i was employed

The LEHD Infrastructure Files Introduction

lacksquare workers i=1,...,I

firms j = 1, ..., J

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

quarter t employment of establishment r in firm $j \ N_{jrt}$

active establishments at firm $j R_{jt}$

 y_{ijt} establishment at which i was employed

⇒ Correction of spurious worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

lacksquare individuals employed at firm j

 \mathcal{J}_t firms active

⇒ U2W: Unit to Worker

⇒ Probability Model

Simplementation

Computing the statistics Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

lacksquare workers i=1,...,I

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Simplementation ⇒ Implementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

quarter t employment of establishment r in firm $j N_{jrt}$ y_{ijt} establishment at which i was employed lacktriangle \mathcal{R}_{jt} set of active $(N_{jrt}>0)$ establishments active establishments at firm $j \ R_{jt}$ **\blacksquare** \mathcal{I}_{jt} individuals employed at firm jfirms j = 1, ..., J \mathcal{J}_t firms active

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment

characteristics to jobs **⊃U2W: Unit to Worker**

Description of the property of the property

Computing the statistics ⇒ Implementation Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

lacksquare workers i=1,...,I

firms j=1,...,J

active establishments at firm $j R_{jt}$

quarter t employment of establishment r in firm $j N_{jrt}$

 y_{ijt} establishment at which i was employed

 \mathcal{J}_t firms active

 \mathcal{I}_{jt} individuals employed at firm j

 \mathcal{R}_{jt} set of active $(N_{jrt}>0)$ establishments

 $lack r_{it} \subset \mathcal{R}_{jt}$ set of active establishments that are feasible for

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious worker flows

Attaching establishment Successor-Predecessor

characteristics to jobs **⊃U2W: Unit to Worker**

Description of the property of the property

Computing the statistics Implementation Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

lacksquare workers i=1,...,I

firms j=1,...,J

active establishments at firm $j R_{jt}$

quarter t employment of establishment r in firm $j N_{jrt}$

 y_{ijt} establishment at which i was employed

 \mathcal{J}_t firms active

 \mathcal{I}_{jt} individuals employed at firm j

 \mathcal{R}_{jt} set of active $(N_{jrt}>0)$ establishments

 $\mathcal{R}^{\imath}_{it} \subset \mathcal{R}_{jt}$ set of active establishments that are feasible for

lacksquare Feasibility: an establishment $r\in \mathcal{R}_{jt}^i$ if $N_{jrs}>0$ for every quarter s that i was employed at j.

U2W: Unit to Worker Impute

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious worker flows

Successor-Predecessor

File

Attaching establishment characteristics to jobs

OU2W: Unit to Worker

Impute

O Probability Model

O Implementation O Implementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

lacksquare workers i=1,...,I

firms j=1,...,J

i active establishments at firm $j \ R_{jt}$

quarter t employment of establishment r in firm $j N_{jrt}$

 y_{ijt} establishment at which i was employed

 \blacksquare \mathcal{J}_t firms active

 \blacksquare \mathcal{I}_{jt} individuals employed at firm j

lacktriangle \mathcal{R}_{jt} set of active $(N_{jrt}>0)$ establishments

 $\mathcal{R}^{\imath}_{it} \subset \mathcal{R}_{jt}$ set of active establishments that are feasible for

Feasibility: an establishment $r \in \mathcal{R}^i_{jt}$ if $N_{jrs} > 0$ for every quarter s that i was employed at j.

The LEHD Infrastructure Files

 $p_{ijrt} = \Pr\left(y_{ijt} = r\right)$

Introduction Input Files Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows ⇒ Solution:

Successor-Predecessor

Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

D Implementation

⇒ Computing the statistics ⇒ Implementation

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

 $p_{ijrt} = \Pr\left(y_{ijt} = r\right)$

Input Files

Infrastructure Files

(L)

Forming Aggregated Estimates: QWI

 $p_{ijrt} = \sum_{s \in \mathcal{R}_{it}^i} e^{\alpha_{jst} + x'_{ijst}\beta}$

 $e^{\alpha_{jrt}+x'_{ijrt}\beta}$

⇒ Correction of spurious

worker flows

Successor-Predecessor ⇒ Solution:

Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Simplementation

O Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows ⇒ Solution:

Successor-Predecessor

Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Simplementation Simplementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

USCENSUSBUREAU

 $\sum_{s \in \mathcal{R}_{jt}^i} e^{\alpha_{jst} + x'_{ijst}\beta}$ $\overline{p_{ijrt}} = \Pr\left(y_{ijt} = r\right)$

 $lpha_{jrt}$ establishment- and quarter-specific effect

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

- Correction of spurious
 - worker flows
- Successor-Predecessor
- Attaching establishment characteristics to jobs
 - characteristics to jobs CU2W: Unit to Worker
 - UZW: Unit

⇒ Probability Model

- ⇒ Implementation
- ImplementationComputing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

USCENSUSBUREAU

 $p_{ijrt} = \Pr\left(y_{ijt} = r\right)$

 $p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}_{jt}^i} e^{\alpha_{jst} + x'_{ijst}\beta}}$

 x_{ijrt} time-varying vector, worker and establishment α_{jrt} establishment- and quarter-specific effect

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

O Correction of spurious

Correction of spu worker flows Successor-Predecessor

Attaching establishment characteristics to jobs

○ U2W: Unit to Worker Impure

○ Probability Model

ImplementationImplementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

$$p_{ijrt} = \Pr\left(y_{ijt} = r\right)$$

$$p_{ijrt} = rac{e^{lpha_{jrt} + x'_{ijrt}eta}}{\sum_{s \in \mathcal{R}^i_{jt}} e^{lpha_{jst} + x'_{ijst}eta}}$$

 β effect on probability of being employed at a particular x_{ijrt} time-varying vector, worker and establishment α_{jrt} establishment- and quarter-specific effect establishment

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious

worker flows Solution:

Successor-Predecessor

Tile

Attaching establishment characteristics to jobs

characteristics to jobs C U2W: Unit to Worker

⇒ Probability Model

ImplementationImplementationComputing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

$$p_{ijrt} = \Pr\left(y_{ijt} = r\right)$$

$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}^{i}_{jt}} e^{\alpha_{jst} + x'_{ijst}\beta}}$$

 β effect on probability of being employed at a particular x_{ijrt} time-varying vector, worker and establishment α_{jrt} establishment- and quarter-specific effect establishment

Currently:

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

⇒ U2W: Unit to Worker

⇒ Probability Model

Simplementation ⇒ Implementation Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

$$p_{ijrt} = \Pr\left(y_{ijt} = r\right)$$

$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}^i_{jt}} e^{\alpha_{jst} + x'_{ijst}\beta}}$$

 β effect on probability of being employed at a particular x_{ijrt} time-varying vector, worker and establishment α_{jrt} establishment- and quarter-specific effect establishment

Currently:

 $\star x_{ijrt}$ is linear spline in distance between residence and establishment

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious

- worker flows Solution: Successor-Predecessor
- File C Attaching establishment
 - characteristics to jobs CU2W: Unit to Worker
 - Impute
- ⇒ Probability Model

Simplementation

- ImplementationComputing the statistics
- Disclosure-proofing the QWI

Publicly available files

Conclusion

$$p_{ijrt} = \Pr\left(y_{ijt} = r\right)$$

$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}_{jt}^i} e^{\alpha_{jst} + x'_{ijst}\beta}}$$

 β effect on probability of being employed at a particular x_{ijrt} time-varying vector, worker and establishment α_{jrt} establishment- and quarter-specific effect establishment

Currently:

- $lacktriangle x_{ijrt}$ is linear spline in distance between residence and establishment
- \bullet α_{jrt} is a hierarchical Bayesian model based on N_{jrt} is

The LEHD Infrastructure Files Introduction

Using Minnesota data,

Input Files

Infrastructure Files

Forming Aggregated

Estimates: QWI

⇒ Correction of spurious worker flows

⇒ Solution:

Successor-Predecessor

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

⊃ Implementation

⇒ Computing the statistics

⇒ Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Using Minnesota data,

Input Files

compute posterior modal value of α_{jrt}

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor ⇒ Solution:

Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

⇒ Implementation

⇒ Computing the statistics ⇒ Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Using Minnesota data,

Input Files

Infrastructure Files

evaluate the posterior mode of $p(\beta|\alpha,x,y)$

compute posterior modal value of $lpha_{jrt}$

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor ⇒ Solution:

Attaching establishment

characteristics to jobs

DU2W: Unit to Worker

⇒ Probability Model

⇒ Implementation

⇒ Computing the statistics Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

- Correction of spurious
 - worker flows
- Successor-Predecessor
- Attaching establishment
 - characteristics to jobs
 - ⇒ U2W: Unit to Worker
- Seriod Probability Model
 - ⊃ Implementation
- Simplementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

evaluate the posterior mode of $p(\beta|\alpha,x,y)$ compute posterior modal value of α_{jrt} Using Minnesota data, maximize

$$\log p(\beta | \alpha, x, y) \propto \sum_{t=1}^{T} \sum_{j \in \mathcal{J}_t} \sum_{i \in \mathcal{I}_{jt}} \sum_{r \in \mathcal{R}_{jt}^i} d_{ijrt}$$

$$(\alpha_{jrt} + x'_{ijrt}\beta)$$

$$-\log \left(\sum_{j \in \mathcal{I}_t} e^{\alpha_{jst} + x'_{ijst}\beta} \right)$$

The LEHD Infrastructure Files

lacksquare use mean and variance of eta from Minnesota data

Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor ⇒ Solution:

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

⊃ Implementation ⇒ Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

 \blacksquare take 10 draws of β from the normal approximation (at the

■ use mean and variance of β from Minnesota data

Infrastructure Files

mode) to $p(\beta|\alpha, x, y)$.

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment

characteristics to jobs

DU2W: Unit to Worker

Seriod Probability Model Implementation

⇒ Computing the statistics ⇒ Implementation

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

mode) to $p(\beta|\alpha, x, y)$.

Forming Aggregated Estimates: QWI

lacktriangle use QCEW employment counts, compute 10 values of $lpha_{jt}$

take 10 draws of β from the normal approximation (at the

use mean and variance of β from Minnesota data

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment

characteristics to jobs

⇒ U2W: Unit to Worker

Seriod Probability Model Implementation

Computing the statistics ⇒ Implementation

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious

Successor-Predecessor worker flows

Attaching establishment

characteristics to jobs **DU2W: Unit to Worker**

⇒ Probability Model Simplementation ⇒ Implementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

- take 10 draws of β from the normal approximation (at the mode) to $p(\beta|\alpha, x, y)$.
- use QCEW employment counts, compute 10 values of $lpha_{it}$
- The drawn values of lpha and eta are used to draw 10 imputed values of place of work from to the posterior predictive distribution

(2)
$$p(\tilde{y}|x,y) = \int \int p(\tilde{y}|\alpha,\beta,x,y) p(\alpha|N) p(\beta|\alpha,x,y) d\alpha d\beta$$

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious worker flows

Successor-Predecessor

Attaching establishment

characteristics to jobs ⇒ U2W: Unit to Worker

Description of the property of the property

Umplementation Simplementation Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

use mean and variance of β from Minnesota data

take 10 draws of β from the normal approximation (at the mode) to $p\left(\beta|\alpha,x,y
ight)$. use QCEW employment counts, compute 10 values of $lpha_{it}$

The drawn values of lpha and eta are used to draw 10 imputed values of place of work from to the posterior predictive distribution (2) $p(\tilde{y}|x,y) = \int p(\tilde{y}|\alpha,\beta,x,y) p(\alpha|N) p(\beta|\alpha,x,y) d\alpha d\beta$

■ → 10 establishment identifiers associated with a job spell

The LEHD Infrastructure Files

We now have:

Introduction

Infrastructure Files

Input Files

Forming Aggregated

Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor ⇒ Solution:

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

S Probability Model

⇒ Implementation ⇒ Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

The LEHD Infrastructure Files Introduction

We now have:Jobs identified

Input Files

Infrastructure Files

Forming Aggregated

⇒ Correction of spurious Estimates: QWI

worker flows

⇒ Solution:

Successor-Predecessor

⇒ Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

Simplementation

⇒ Computing the statistics

⇒ Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Jobholder's demographics (age, gender)

Jobs identified

■ We now have:

Forming Aggregated

⇒ Correction of spurious Estimates: QWI

worker flows ⇒ Solution:

Successor-Predecessor

Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

O Implementation Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

→ Jobholder's demographics

→ Jobs identified

■ We now have:

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

⇒ Solution:

Successor-Predecessor

Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model Implementation

⇒ Computing the statistics ⇒ Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Establishment's characteristics (geography and industry)

Jobholder's demographics

→ Jobs identified

■ We now have:

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

O Implementation

Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Establishment's characteristics

Jobholder's demographics

→ Jobs identified

■ We now have:

Forming Aggregated Estimates: QWI

⇒ Correction of spurious worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

O Implementation Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

◆ Establishment's characteristics

Jobholder's demographics

→ Jobs identified

■ We now have:

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Now compute

Successor-Predecessor

Attaching establishment

characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model

O Implementation Implementation

⇒ Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment

characteristics to jobs

⇒ U2W: Unit to Worker

Seriod Probability Model

O Implementation Simplementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

■ We now have:

→ Jobs identified

Jobholder's demographics

Establishment's characteristics

■ Now compute

1. For each job, the relevant variables, defined at the person-level (indicators)

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious worker flows Successor-Predecessor

Attaching establishment characteristics to jobs

⇒ U2W: Unit to Worker

Seriod Probability Model Simplementation Simplementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

We now have:

Jobs identified

Jobholder's demographics

Establishment's characteristics

■ Now compute

1. For each job, the relevant variables, defined at the person-level (indicators) 2. Aggregate (typically sum) to the establishment level

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

⇒ Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

⇒ U2W: Unit to Worker

> Probability Model Simplementation Computing the statistics

O Implementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

We now have:

Jobs identified

Jobholder's demographics

Establishment's characteristics

■ Now compute

1. For each job, the relevant variables, defined at the person-level (indicators) 2. Aggregate (typically sum) to the establishment level

3. → establishment-level statistics, available in RDC

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

⊃U2W: Unit to Worker

⇒ Probability Model Simplementation Simplementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

We now have:

Jobs identified

Jobholder's demographics

◆ Establishment's characteristics

■ Now compute

1. For each job, the relevant variables, defined at the person-level (indicators) 2. Aggregate (typically sum) to the establishment level

⇒ establishment-level statistics, available in RDC

4. Attach weights to each establishment

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment characteristics to jobs

DU2W: Unit to Worker

⇒ Probability Model

Simplementation Umplementation

Computing the statistics

Disclosure-proofing the QWI

Publicly available files

Conclusion

We now have:

Jobs identified

◆ Jobholder's demographics

Establishment's characteristics

Now compute

1. For each job, the relevant variables, defined at the person-level (indicators) 2. Aggregate (typically sum) to the establishment level

⇒ establishment-level statistics, available in RDC

Attach weights to each establishment

5. Attach 'fuzz' factors to each establishment

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious

worker flows

Attaching establishment

Successor-Predecessor

characteristics to jobs ⇒ U2W: Unit to Worker

⇒ Probability Model Umplementation

⇒ Computing the statistics Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

■ We now have:

Jobs identified

Jobholder's demographics

Establishment's characteristics

Now compute

1. For each job, the relevant variables, defined at the person-level (indicators) 2. Aggregate (typically sum) to the establishment level

⇒ establishment-level statistics, available in RDC

Attach weights to each establishment

5. Attach 'fuzz' factors to each establishment

geography-industry-demographic detail Final aggregation to desired

The LEHD Infrastructure Files

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Correction of spurious

worker flows

Successor-Predecessor

Attaching establishment

characteristics to jobs ⇒ U2W: Unit to Worker

⇒ Probability Model

Umplementation

Computing the statistics Simplementation

Disclosure-proofing the QWI

Publicly available files

Conclusion

■ We now have:

Jobs identified

Jobholder's demographics

Establishment's characteristics

Now compute

1. For each job, the relevant variables, defined at the person-level (indicators) 2. Aggregate (typically sum) to the establishment level

⇒ establishment-level statistics, available in RDC

Attach weights to each establishment

5. Attach 'fuzz' factors to each establishment

geography-industry-demographic detail Final aggregation to desired

Disclosure-proof

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Noise-infusion
⇒ Item suppression

Publicly available files

Conclusion

Disclosure-proofing the QWI

Noise-infusion

The LEHD Infrastructure Files Introduction

First layer: workplace-level aggregation

Input Files

Infrastructure Files

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI Shoise-infusion

⇒ Item suppression

Publicly available files

Noise-infusion

The LEHD Infrastructure Files Introduction

First layer: workplace-level aggregation
 infusion of specially constructed noise:

Input Files

Infrastructure Files

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

⇒ Noise-infusion
⇒ Item suppression

Publicly available files

Conclusion

Noise-infusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Disclosure-proofing

⇒ Item suppression

Publicly available files

Conclusion

■ First layer: workplace-level aggregation

infusion of specially constructed noise:

(3)
$$p(\delta_j) = \begin{cases} (b-\delta)/(b-a)^2, & \delta \in [a,b] \\ (b+\delta-2)/(b-a)^2, & \delta \in [2-b,2-a] \end{cases}$$

Noise-infusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Noise-infusion
⇒ Item suppression

Publicly available files

Conclusion

First layer: workplace-level aggregation

◆ infusion of specially constructed noise:

 $(b+\delta-2)/(b-a)^2, \ \delta \in [2-b, 2-a]$ $(b-\delta)/(b-a)^2, \ \delta \in [a,b]$ (3) $p(\delta_j) =$

Result: random noise factor centered around 1 with distortion of at least a-1 and at most b-1.

Noise-infusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Noise-infusion
⇒ Item suppression

Publicly available files

Conclusion

■ First layer: workplace-level aggregation

infusion of specially constructed noise:

 $(b+\delta-2)/(b-a)^2, \ \delta \in [2-b, 2-a]$ $(b-\delta)/(b-a)^2, \ \delta \in \ [a,b]$ (3) $p(\delta_j) =$

 ◆ Result: random noise factor centered around 1 with distortion of at least a-1 and at most b-1.

Important properties:

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Ultern suppression

Publicly available files

Conclusion

■ First layer: workplace-level aggregation

♦ infusion of specially constructed noise:

(3)
$$p(\delta_j) = \begin{cases} (b-\delta)/(b-a)^2, & \delta \in [a,b] \\ (b+\delta-2)/(b-a)^2, & \delta \in [2-b,2-a] \end{cases}$$

 ◆ Result: random noise factor centered around 1 with distortion of at least a-1 and at most b-1.

Important properties:

1. for a given workplace, distortion is always distorted in the same direction (increased or decreased) by the same percentage amount in every period.

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

O Noise-infusion
Oltem suppression

Publicly available files

Conclusion

■ First layer: workplace-level aggregation

infusion of specially constructed noise:

 $(b+\delta-2)/(b-a)^2, \ \delta \in [2-b, 2-a]$ $(b-\delta)/(b-a)^2, \delta \in [a,b]$ (3) $p(\delta_j) =$

 ◆ Result: random noise factor centered around 1 with distortion of at least a-1 and at most b-1.

Important properties:

1. for a given workplace, distortion is always distorted in the same direction (increased or decreased) by the same percentage amount in every period.

distortion cancel out for the vast majority of the estimates. 2. when estimates are aggregated, the effects of the

The LEHD Infrastructure Files Introduction

Second layer: after aggregations

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Item suppression ⇒ Noise-infusion

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Some estimates are based on fewer than three persons or

Second layer: after aggregations

Infrastructure Files

firms.

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Item suppression ⇒ Noise-infusion

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Second layer: after aggregations

Input Files

Some estimates are based on fewer than three persons or

Infrastructure Files

firms.

Forming Aggregated Estimates: QWI

→ suppression of these estimates

Disclosure-proofing the QWI

⇒ Noise-infusion

⇒ Item suppression

Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Noise-infusion⇒ Item suppression

Publicly available files

Conclusion

Second layer: after aggregations

 Some estimates are based on fewer than three persons or firms.

→ suppression of these estimates

Some of the estimates are based on noisy data

The LEHD Infrastructure Files Introduction

Second layer: after aggregations

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Disclosure-proofing to Noise-infusion

Publicly available files

Conclusion

→ suppression of these estimates

Some estimates are based on fewer than three persons or

Some of the estimates are based on noisy data

→ flagged as "substantially distorted"

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Noise-infusion ⇒ Item suppression Publicly available files

Conclusion

Second layer: after aggregations

Some estimates are based on fewer than three persons or

→ suppression of these estimates

Some of the estimates are based on noisy data

→ flagged as "substantially distorted"

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Noise-infusion ⇒ Item suppression Publicly available files

Conclusion

Second layer: after aggregations

Some estimates are based on fewer than three persons or

→ suppression of these estimates

Some of the estimates are based on noisy data

→ flagged as "substantially distorted"

USCENSUSBUREAU

Publicly available files

Disclosure-proofing the QWI

Forming Aggregated

Estimates: QWI

Infrastructure Files

Publicly available files

Publicly available files

Conclusion

The LEHD Infrastructure Files

Introduction

Input Files

The LEHD Infrastructure Files Introduction

■ Published QWI

Input Files

Infrastructure Files

Forming Aggregated

Estimates: QWI

Disclosure-proofing the QWI

Publicly available files Dublicly available files

Conclusion

The LEHD Infrastructure Files

■ Published QWI

Introduction

Input Files

→ http://lehd.dsd.census.gov

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Publicly available files Publicly available files

Conclusion

The LEHD Infrastructure Files Introduction

■ Published QWI

Input Files

→ http://lehd.dsd.census.gov

Infrastructure Files

■ RDC

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files Despectively available files

Conclusion

The LEHD Infrastructure Files Introduction

■ Published QWI

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

◆ Employer characteristics files ECF → LEHD-ECF

→ http://lehd.dsd.census.gov

Disclosure-proofing the QWI

⇒ Publicly available files Publicly available files

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

⇒ Publicly available files Publicly available files

Published QWI

→ http://lehd.dsd.census.gov

◆ Employer characteristics files ECF → LEHD-ECF

Establishment level flow files - Firm-level QWI -> LEHD-QWI

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Description of the properties of the properti Publicly available files

Published QWI

→ http://lehd.dsd.census.gov

◆ Employer characteristics files ECF → LEHD-ECF

Establishment level flow files - Firm-level QWI -> LEHD-QWI

LEHD Business Register Bridge (LEHD-BRB)

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Description of the properties of the properti

Published QWI

→ http://lehd.dsd.census.gov

◆ Employer characteristics files ECF → LEHD-ECF

Establishment level flow files - Firm-level QWI -> LEHD-QWI

LEHD Business Register Bridge (LEHD-BRB)

Human Capital files LEHD-HCF

The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

sciosare-proding rie

Publicly available files > Publicly available files

onclusion

Published QWI

→ http://lehd.dsd.census.gov

◆ Employer characteristics files ECF → LEHD-ECF

Establishment level flow files - Firm-level QWI -> LEHD-QWI

LEHD Business Register Bridge (LEHD-BRB)

◆ Human Capital files LEHD-HCF

→ http://www.ces.census.gov

USCENSUSBUREAU

Conclusion

Disclosure-proofing the QWI

Forming Aggregated

Estimates: QWI

Infrastructure Files

Publicly available files

⊃ Flow so far

Conclusion

The LEHD Infrastructure Files

Introduction

Input Files

The LEHD Infrastructure Files Introduction

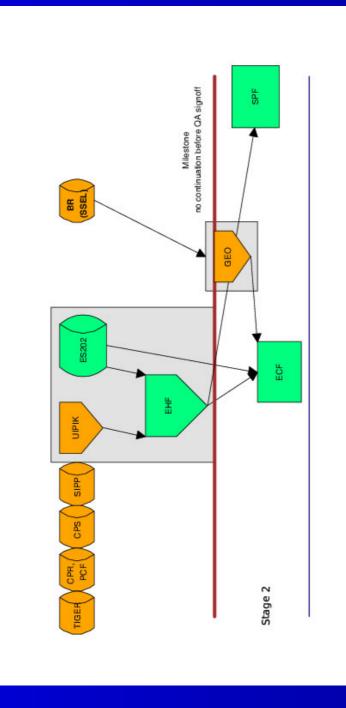
Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI Disclosure-proofing the QWI

Publicly available files

⇒ Flow so far Conclusion



Main page

The LEHD Infrastructure Files Introduction

Input Files

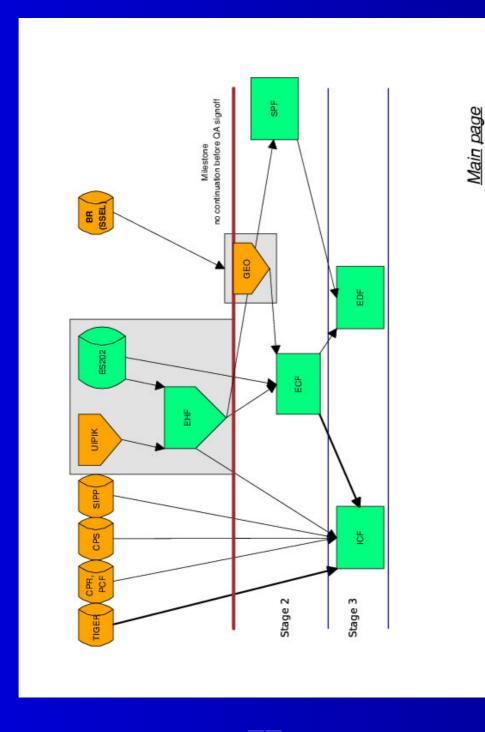
Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion D Flow so far



The LEHD Infrastructure Files Introduction

Input Files

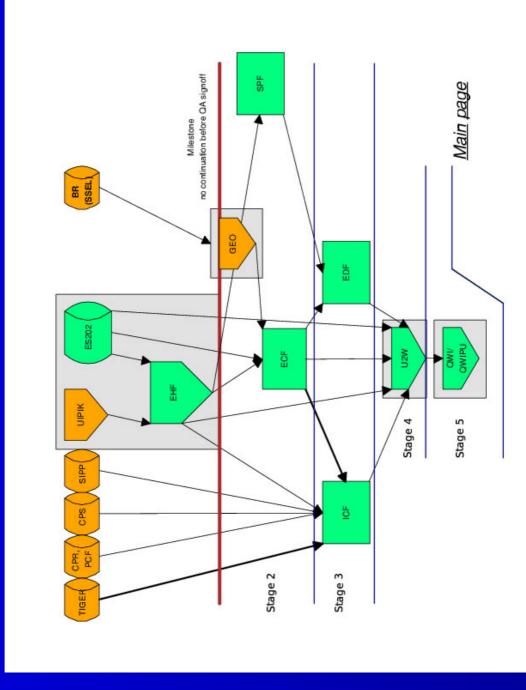
Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion D Flow so far



The LEHD Infrastructure Files Introduction

Input Files

Infrastructure Files

Forming Aggregated Estimates: QWI

Disclosure-proofing the QWI

Publicly available files

Conclusion

⇒ Flow so far

