The LEHD Infrastructure Files

and the Creation of the Quarterly Workforce Indicators

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

♠ Cornell University

♣ U.S. Census Bureau, LEHD Program

Introduction

What are QWI?

- Since 2003: publication of Quarterly Workforce Indicators

What are QWI?

- Since 2003: publication of Quarterly Workforce Indicators
- The first 21st century statistical system

What are QWI?

- Since 2003: publication of Quarterly Workforce Indicators
- The first 21st century statistical system
  - No additional burden

What are QWI?

- 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

What are QWI?

- 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 are QWI?

- 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?

- 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

What is it?

- Combines

What is it?

- Combines
  - (state) administrative records data on workers (UI Wage records)

What is it?

- Combines
  - (state) administrative records data on workers (UI Wage records)
  - (state) administrative records data on firms (QCEW aka ES-202)

What is it?

- Combines
  - (state) administrative records data on workers (UI Wage records)
  - (state) administrative records data on firms (QCEW aka ES-202)
  - administrative information on demographics

What is it?

- 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

What is it?

- 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 firm units

What is it?

- 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 firm units
- careful longitudinal edit of person and firm characteristics

In this paper

- Describe the construction of the LEHD infrastructure

In this paper

- Describe the construction of the LEHD infrastructure
  - ... in particular the imputation mechanisms used

In this paper

- Describe the construction of the LEHD infrastructure
  - ... in particular the imputation mechanisms used
- Describe the computation of the QWI statistics

In this paper

- 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

In this paper

- 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

In this paper

- 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

Input Files

Wage records: UI

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

Wage records: UI

- 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

Wage records: UI

- 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

Wage records: UI

- 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

Wage records: UI

- 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

Employer reports: ES202

... or QCEW

Employer reports: ES202

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

Employer reports: ES202

- 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)

Employer reports: ES202

- 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

Employer reports: ES202

- 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' ($\approx$ establishment)

Employer reports: ES202

- 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' ($\approx$ establishment)
- One report per establishment per quarter is filed

Demographics

- Demographics are taken from a number of Census-internal files derived from administrative data:
  - Person Characteristics File (PCF)
  - Census Numident

Demographics

- 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

Infrastructure Files

EHF: Employment History Files

- Job-level EHF
  - complete in-state work history for each individual on UIwage records.
  - one record for each employee-employer combination – a job
  - earnings and employment patterns

EHF: Employment History Files

- Job-level EHF
  - complete in-state work history for each individual on UIwage records.
  - one record for each employee-employer combination – a job
  - 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

- Job-level EHF
  - complete in-state work history for each individual on UIwage records.
  - one record for each employee-employer combination – a job
  - 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.

ICF: Individual Characteristics File

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

ICF: Individual Characteristics File

- Demographic information from the PCF is merged with universe of PIKs from wage records
- records without a valid match flagged

ICF: Individual Characteristics File

- 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.

ICF: Individual Characteristics File

- 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

ICF: Individual Characteristics File

- 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

ICF: Individual Characteristics File

- 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

ICF: Individual Characteristics File

- 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

ICF: Individual Characteristics File

- 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

ICF: Individual Characteristics File

- 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

ICF: Individual Characteristics File

- 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

  are each imputed ten times

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records
- Inputs:

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records
- Inputs:
  1. ES202

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records
- Inputs:
  1. ES202
  2. UI: supplement information on the ES202, extend published BLS county-level employment data

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records
- Inputs:
  1. ES202
  2. UI: supplement information on the ES202, extend published BLS county-level employment data
  3. GAL: establishment geocodes

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records
- Inputs:
  1. ES202
  2. UI: supplement information on the ES202, extend published BLS county-level employment data
  3. GAL: establishment geocodes
  4. LDB (BLS) for backfilling NAICS information

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records
- Inputs:
  1. ES202
  2. UI: supplement information on the ES202, extend published BLS county-level employment data
  3. GAL: establishment geocodes
  4. LDB (BLS) for backfilling NAICS information
- Longitudinal edits for consistency and data completion

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records
- Inputs:
  1. ES202
  2. UI: supplement information on the ES202, extend published 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

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records
- Inputs:
  1. ES202
  2. UI: supplement information on the ES202, extend published 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

ECF: Employer Characteristics File

- Two files: firm and establishment level, quarterly records
- Inputs:
  1. ES202
  2. UI: supplement information on the ES202, extend published 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

GAL: Geocoded Address List

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

GAL: Geocoded Address List

- ... is a data set containing unique commercial and residential addresses
- geocoded to the Census Block and latitude/longitude coordinates

GAL: Geocoded Address List

- ... is a data set containing unique commercial and residential addresses
- geocoded to the Census Block and latitude/longitude coordinates
- Inputs:

GAL: Geocoded Address List

- ... is a data set containing unique commercial and residential addresses
- geocoded to the Census Block and latitude/longitude coordinates
- Inputs:
  1. ES202 data

GAL: Geocoded Address List

- ... 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)

GAL: Geocoded Address List

- ... 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)

GAL: Geocoded Address List

- ... 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)

GAL: Geocoded Address List

- ... 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

GAL: Geocoded Address List

- ... 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

GAL: Geocoded Address List

- ... 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

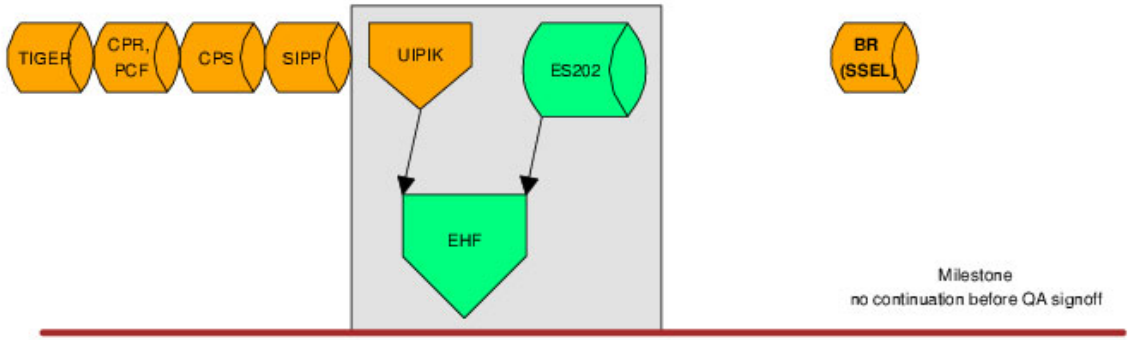
GAL: Geocoded Address List

- ... 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
  3. unduplicated (by firm name)

Flow so far

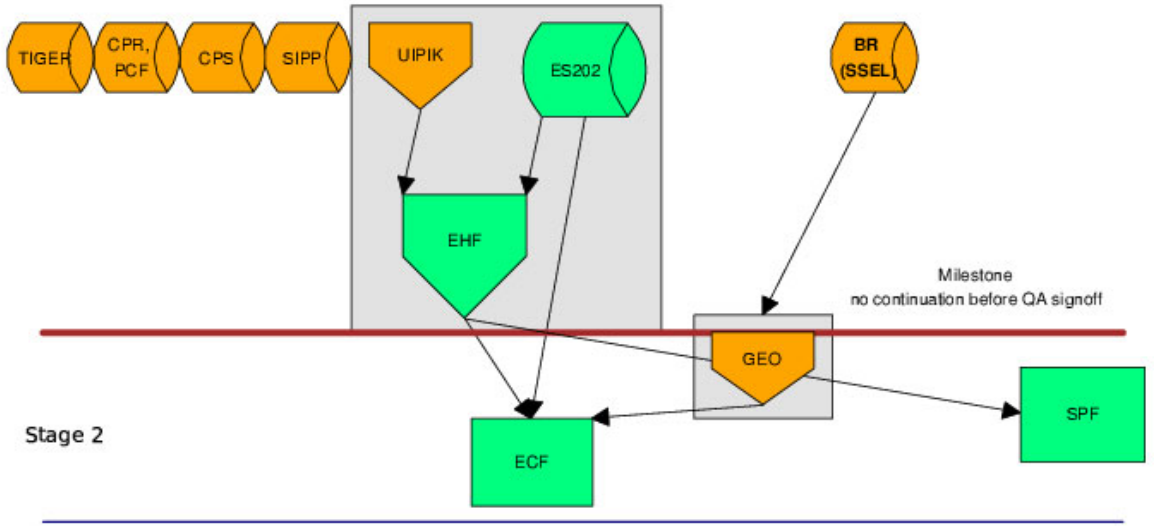
Main page

Flow so far



Main page

Flow so far



Main page

Forming Aggregated Estimates: QWI

Correction of spurious worker flows

- Firm identifier:

Correction of spurious worker flows

- Firm identifier: state-specific account number

Correction of spurious worker flows

- Firm identifier:
- Account numbers can and do change:

Correction of spurious worker flows

- Firm identifier:
- Account numbers can and do change:
  - change in legal form

Correction of spurious worker flows

- Firm identifier:
- Account numbers can and do change:
  - change in legal form
  - a merger

Correction of spurious worker flows

- Firm identifier:
- Account numbers can and do change:
  - change in legal form
  - a merger
- Change in firm identifier

Correction of spurious worker flows

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

Correction of spurious worker flows

- Firm identifier:
- Account numbers can and do change:
  - ✦ change in legal form
  - ✦ a merger
- Change in firm identifier

- $\longrightarrow$ non-economic change in identifier creates spurious flow

Solution: Successor-Predecessor File

- track large worker movements between SEINs

Solution: Successor-Predecessor File

- track large worker movements between SEINs
- $\longrightarrow$ link entities that have different account numbes, but constitute the same economic entitiy

Solution: Successor-Predecessor File

- track large worker movements between SEINs
- $\longrightarrow$ link entities that have different account numbes, but constitute the same economic entitiy
- SPF provides a variety of link characteristics, based on the number of workers leaving an SEIN, in both absolute and relative terms, and the number of workers entering an SEIN, again in absolute and relative terms.

Solution: Successor-Predecessor File

- track large worker movements between SEINs
- $\longrightarrow$ link entities that have different account numbes, but constitute the same economic entitiy
- SPF provides a variety of link characteristics, based on the number of workers leaving an SEIN, in both absolute and relative terms, and the number of workers entering an SEIN, again in absolute and relative terms.
- QWI: if 80% of an SEIN's workers (the predecessor) are observed to move to a single successor, and that successor absorbs 80% of its employees from a single predecessor, then all flows between those two account numbers are filtered out, and treated as if they had never existed.

Attaching establishment characteristics to jobs

- Goal: achieve a high level of accuracy and detail

Attaching establishment characteristics to jobs

- Goal: achieve a high level of accuracy and detail
- Problem: no establishment identification on wage record

Attaching establishment characteristics to jobs

- Goal: achieve a high level of accuracy and detail
- Problem:

Attaching establishment characteristics to jobs

- Goal: achieve a high level of accuracy and detail
- Problem:
- 30-40% of state-wide employment in multi-establishment firms

Attaching establishment characteristics to jobs

- 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

Attaching establishment characteristics to jobs

- 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:

Attaching establishment characteristics to jobs

- 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

Attaching establishment characteristics to jobs

- 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.

Attaching establishment characteristics to jobs

- 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.
- Important practical aspects:

Attaching establishment characteristics to jobs

- 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.
- Important practical aspects:
  - Non-ignorable missing data imputation

Attaching establishment characteristics to jobs

- 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.
- Important practical aspects:
  - Non-ignorable missing data imputation
  - Several million imputations every quarter

U2W: Unit to Worker Impute

- workers $i = 1, ..., I$

U2W: Unit to Worker Impute

- workers $i = 1, ..., I$
- firms $j = 1, ..., J$

U2W: Unit to Worker Impute

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

U2W: Unit to Worker Impute

- 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}

U2W: Unit to Worker Impute

- 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

U2W: Unit to Worker Impute

- 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

U2W: Unit to Worker Impute

- 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

U2W: Unit to Worker Impute

- 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

U2W: Unit to Worker Impute

- 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
- $\mathcal{R}^i_{jt} \subset \mathcal{R}_{jt}$ set of active establishments that are feasible for worker i.

U2W: Unit to Worker Impute

- 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
- $\mathcal{R}^i_{jt} \subset \mathcal{R}_{jt}$ set of active establishments that are feasible for worker i.
- Feasibility: an establishment $r \in \mathcal{R}^i_{jt}$ if $N_{jrs} > 0$ for every quarter s that i was employed at j.

U2W: Unit to Worker Impute

- 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
- $\mathcal{R}^i_{jt} \subset \mathcal{R}_{jt}$ set of active establishments that are feasible for worker i.
- Feasibility: an establishment $r \in \mathcal{R}^i_{jt}$ if $N_{jrs} > 0$ for every quarter s that i was employed at j.

Probability Model

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

Probability Model

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

(1)
$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}^i_{jt}} e^{\alpha_{jst} + x'_{ijst}\beta}}$$

Probability Model

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

(1)
$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}_{jt}^{i}} e^{\alpha_{jst} + x'_{ijst}\beta}}$$

α_{jrt} establishment- and quarter-specific effect

Probability Model

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

$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}_{jt}^{i}} e^{\alpha_{jst} + x'_{ijst}\beta}} \tag{1}$$

α_{jrt} establishment- and quarter-specific effect

x_{ijrt} time-varying vector, worker and establishment

Probability Model

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

(1)
$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}_{jt}^{i}} e^{\alpha_{jst} + x'_{ijst}\beta}}$$

α_{jrt} establishment- and quarter-specific effect

x_{ijrt} time-varying vector, worker and establishment

β effect on probability of being employed at a particular establishment

Probability Model

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

(1)
$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}_{jt}^{i}} e^{\alpha_{jst} + x'_{ijst}\beta}}$$

α_{jrt} establishment- and quarter-specific effect

x_{ijrt} time-varying vector, worker and establishment

β effect on probability of being employed at a particular establishment

Currently:

Probability Model

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

(1)
$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}_{jt}^{i}} e^{\alpha_{jst} + x'_{ijst}\beta}}$$

α_{jrt} establishment- and quarter-specific effect

x_{ijrt} time-varying vector, worker and establishment

β effect on probability of being employed at a particular establishment

Currently:

- x_{ijrt} is linear spline in distance between residence and establishment

Probability Model

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

(1)
$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}_{jt}^{i}} e^{\alpha_{jst} + x'_{ijst}\beta}}$$

α_{jrt} establishment- and quarter-specific effect

x_{ijrt} time-varying vector, worker and establishment

β effect on probability of being employed at a particular establishment

Currently:

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

Implementation

Using Minnesota data,

Implementation

Using Minnesota data,

compute posterior modal value of α_{jrt}

Implementation

Using Minnesota data,

compute posterior modal value of α_{jrt}

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

Implementation

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

$$\log p\left(\beta|\alpha, x, y\right) \quad \propto \quad \sum_{t=1}^{T} \sum_{j \in \mathcal{J}_t} \sum_{i \in \mathcal{I}_{jt}} \sum_{r \in \mathcal{R}_{jt}^{i}} d_{ijrt} \left(\alpha_{jrt} + x'_{ijrt}\beta - \log\left(\sum_{s \in \mathcal{R}_{jt}^{i}} e^{\alpha_{jst} + x'_{ijst}\beta}\right)\right)$$

Implementation

- use mean and variance of β from Minnesota data

Implementation

- use mean and variance of β from Minnesota data
- take 10 draws of β from the normal approximation (at the mode) to $p\left(\beta|\alpha, x, y\right)$.

Implementation

- use mean and variance of β from Minnesota data
- take 10 draws of β from the normal approximation (at the mode) to $p\left(\beta|\alpha, x, y\right)$.
- use QCEW employment counts, compute 10 values of α_{jt}

Implementation

- use mean and variance of β from Minnesota data
- take 10 draws of β from the normal approximation (at the mode) to $p\left(\beta|\alpha, x, y\right)$.
- use QCEW employment counts, compute 10 values of α_{jt}
- The drawn values of α and β are used to draw 10 imputed values of place of work from to the posterior predictive distribution

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

Implementation

- use mean and variance of β from Minnesota data
- 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 α_{jt}
- The drawn values of α and β are used to draw 10 imputed values of place of work from to the posterior predictive distribution

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

- $\longrightarrow$ 10 establishment identifiers associated with a job spell

Computing the statistics

- We now have:

Computing the statistics

- We now have:
  - Jobs identified

Computing the statistics

- We now have:
  - Jobs identified
  - Jobholder's demographics (age, gender)

Computing the statistics

- We now have:
  - Jobs identified
  - Jobholder's demographics

Computing the statistics

- We now have:
  - Jobs identified
  - Jobholder's demographics
  - Establishment's characteristics (geography and industry)

Computing the statistics

- We now have:
  - Jobs identified
  - Jobholder's demographics
  - Establishment's characteristics

Computing the statistics

- We now have:
  - Jobs identified
  - Jobholder's demographics
  - Establishment's characteristics
- Now compute

Computing the statistics

- 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)

Computing the statistics

- 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

Computing the statistics

- 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. $\longrightarrow$ establishment-level statistics, available in RDC

Computing the statistics

- 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. $\longrightarrow$ establishment-level statistics, available in RDC
  4. Attach weights to each establishment

Computing the statistics

- 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. $\longrightarrow$ establishment-level statistics, available in RDC
  4. Attach weights to each establishment
  5. Attach 'fuzz' factors to each establishment

Computing the statistics

- 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. $\longrightarrow$ establishment-level statistics, available in RDC
  4. Attach weights to each establishment
  5. Attach 'fuzz' factors to each establishment
  6. Final aggregation to desired geography-industry-demographic detail

Computing the statistics

- 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. $\longrightarrow$ establishment-level statistics, available in RDC
  4. Attach weights to each establishment
  5. Attach 'fuzz' factors to each establishment
  6. Final aggregation to desired geography-industry-demographic detail
- Disclosure-proof

Disclosure-proofing the QWI

Noise-infusion

- First layer: workplace-level aggregation

Noise-infusion

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

Noise-infusion

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

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

Noise-infusion

- First layer: workplace-level aggregation
  - infusion of specially constructed noise:
  - $$(3)\ p(\delta_j) = \begin{cases} (b-\delta)\Big/(b-a)^2, & \delta \in [a,b] \\ (b+\delta-2)\Big/(b-a)^2, & \delta \in [2-b, 2-a] \end{cases}$$
  - Result: random noise factor centered around 1 with distortion of at least $a-1$ and at most $b-1$.

Noise-infusion

- First layer: workplace-level aggregation
  - infusion of specially constructed noise:
  - $$(3)\ p(\delta_j) = \begin{cases} (b-\delta)\Big/(b-a)^2,\ \delta \in [a,b] \\ (b+\delta-2)\Big/(b-a)^2,\ \delta \in [2-b,2-a] \end{cases}$$
  - Result: random noise factor centered around 1 with distortion of at least $a-1$ and at most $b-1$.
- Important properties:

Noise-infusion

- First layer: workplace-level aggregation
  - infusion of specially constructed noise:
  - $$(3)\ p(\delta_j) = \begin{cases} (b-\delta)\Big/(b-a)^2,\ \delta \in [a,b] \\ (b+\delta-2)\Big/(b-a)^2,\ \delta \in [2-b,2-a] \end{cases}$$
  - 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.

Noise-infusion

- First layer: workplace-level aggregation
  - infusion of specially constructed noise:
  - $$(3)\ p(\delta_j) = \begin{cases} (b-\delta)\Big/(b-a)^2,\ \delta \in [a,b] \\ (b+\delta-2)\Big/(b-a)^2,\ \delta \in [2-b,2-a] \end{cases}$$
  - 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.
  2. when estimates are aggregated, the effects of the distortion cancel out for the vast majority of the estimates.

Item suppression

- Second layer: after aggregations

Item suppression

- Second layer: after aggregations
  - Some estimates are based on fewer than three persons or firms.

Item suppression

- Second layer: after aggregations
  - Some estimates are based on fewer than three persons or firms.
  - $\longrightarrow$ suppression of these estimates

Item suppression

- Second layer: after aggregations
  - Some estimates are based on fewer than three persons or firms.
  - $\longrightarrow$ suppression of these estimates
  - Some of the estimates are based on noisy data

Item suppression

- Second layer: after aggregations
  - Some estimates are based on fewer than three persons or firms.
  - $\longrightarrow$ suppression of these estimates
  - Some of the estimates are based on noisy data
  - $\longrightarrow$ flagged as “substantially distorted”

Item suppression

- Second layer: after aggregations
  - Some estimates are based on fewer than three persons or firms.
  - $\longrightarrow$ suppression of these estimates
  - Some of the estimates are based on noisy data
  - $\longrightarrow$ flagged as “substantially distorted”

Item suppression

- Second layer: after aggregations
  - Some estimates are based on fewer than three persons or firms.
  - $\longrightarrow$ suppression of these estimates
  - Some of the estimates are based on noisy data
  - $\longrightarrow$ flagged as “substantially distorted”

Publicly available files

Publicly available files

- Published QWI

Publicly available files

- Published QWI

  → http://lehd.dsd.census.gov

Publicly available files

- Published QWI
  → http://lehd.dsd.census.gov
- RDC

Publicly available files

- Published QWI
  → http://lehd.dsd.census.gov
- RDC
  - Employer characteristics files ECF → LEHD-ECF

Publicly available files

- Published QWI
  → http://lehd.dsd.census.gov
- RDC
  - Employer characteristics files ECF → LEHD-ECF
  - Establishment level flow files - Firm-level QWI → LEHD-QWI

Publicly available files

- Published QWI
  → http://lehd.dsd.census.gov
- RDC
  - Employer characteristics files ECF → LEHD-ECF
  - Establishment level flow files - Firm-level QWI → LEHD-QWI
  - LEHD Business Register Bridge (LEHD-BRB)

Publicly available files

- Published QWI
  → http://lehd.dsd.census.gov
- RDC
  - 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

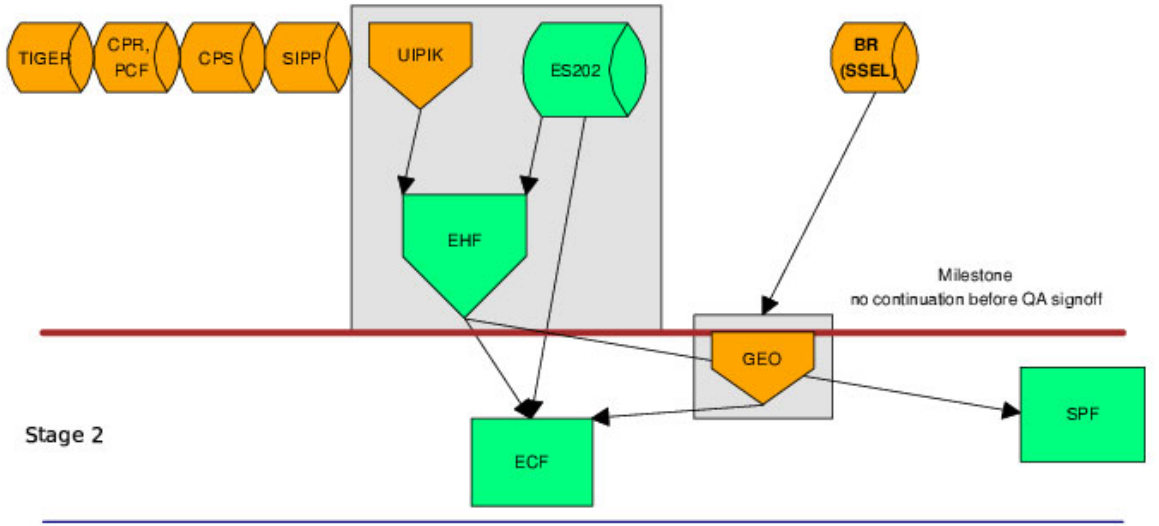
Publicly available files

- Published QWI
  → http://lehd.dsd.census.gov
- RDC
  - ✦ 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

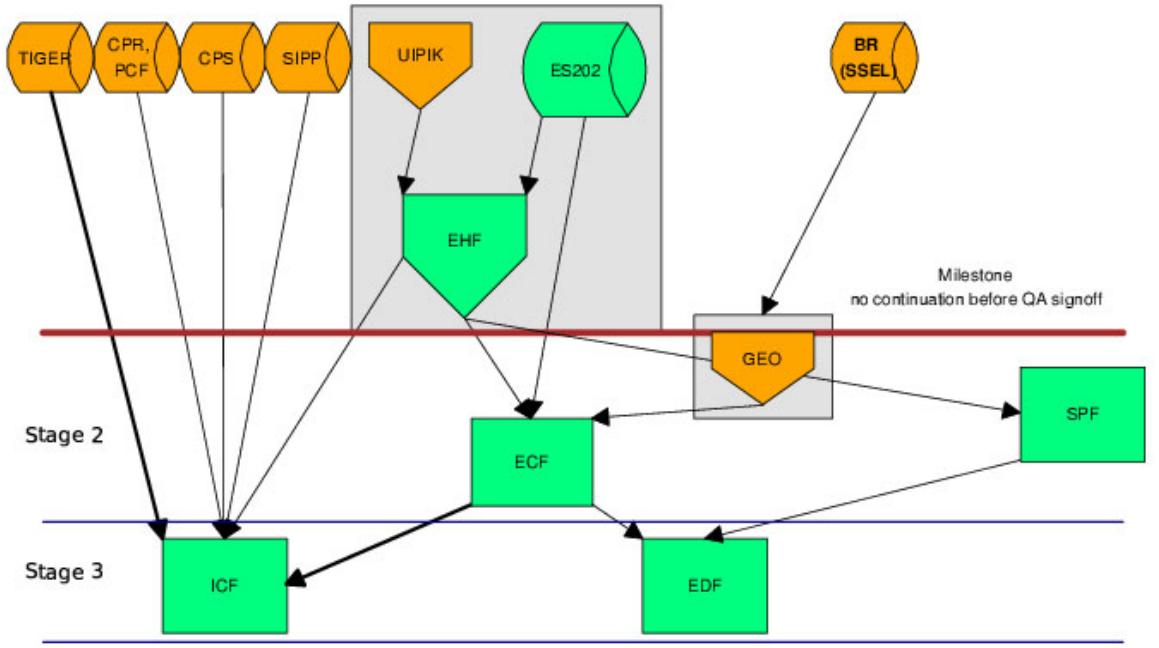
Conclusion

Flow so far



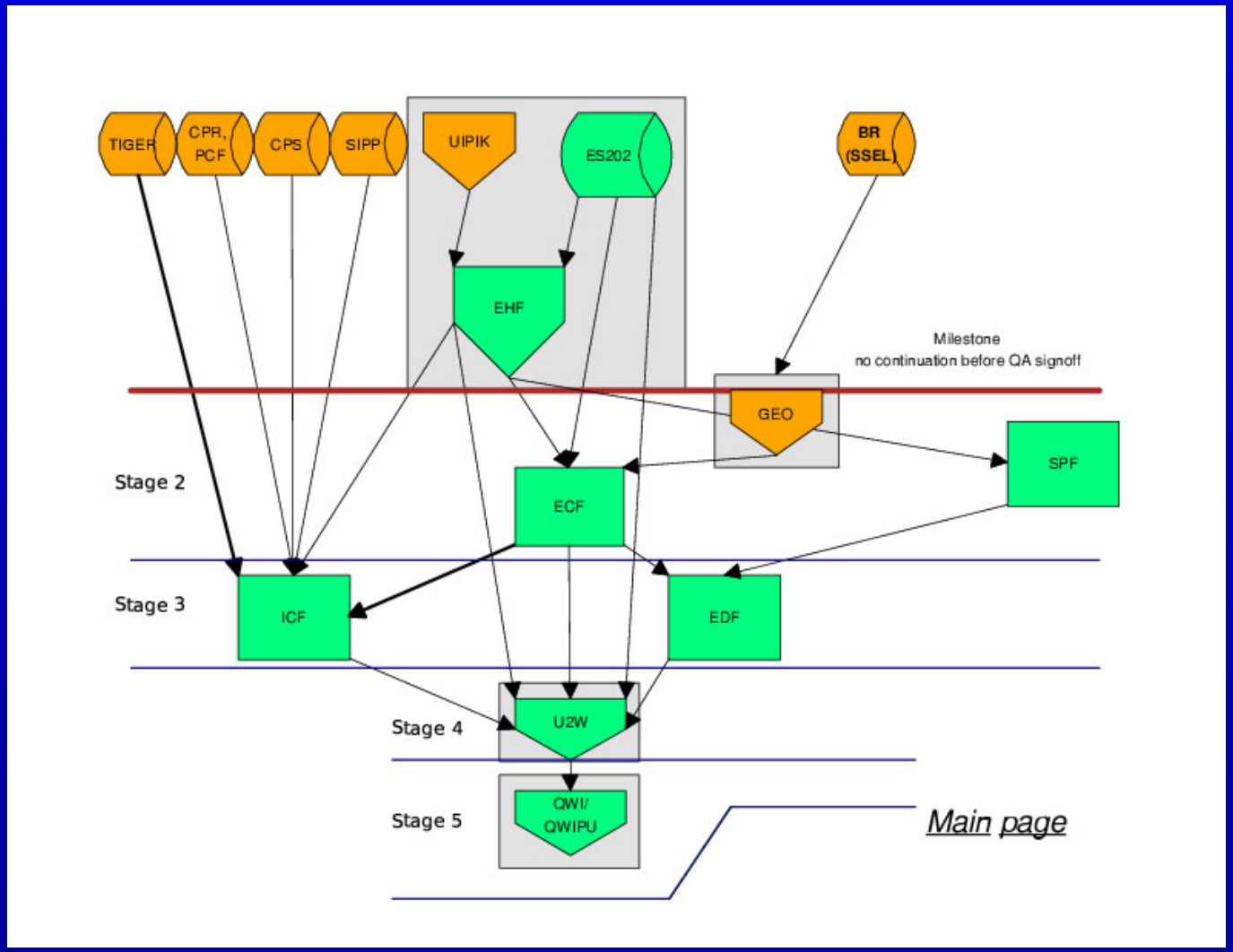
Main page

Flow so far



Main page

Flow so far



Flow so far