CREST-ENSAE Mini-course
Microeconometrics of Modeling Labor Markets Using Linked Employer-Employee Data

John M. Abowd
May 30, 2013
Topics

• May 30: Basics of analyzing complex linked data
• June 3: Basics of graph theory with applications to labor markets
• June 6: Matching and sorting models
• June 10: Endogenous mobility models
Lecture 1

• Population, frames, populations and samples
• Household/individual frames
• Business entity frames
• Integrated frames
• Relational databases
• The structure of longitudinally integrated labor market microdata
• Public-use data from the U.S. Census Bureau LEHD program
Tools for Today’s Lecture

• Much of today’s lecture is based on material from my Cornell course (taught jointly with Lars Vilhuber)

• That course, and all of its exercises, can be accessed at:

  INFO 7470 Understanding Social and Economic Data
Basic Definitions

- Universe (target population)
- Out-of-scope population
- Frame population
- Geography
- Population demographics
- Business entity demographics
- Government entities
Universe = Target Population

- Theoretical construct specifying every entity that satisfies a set of explicit qualifying conditions
- In probability models describing the statistical process of estimation (either finite-population or super-population methods), the universe is the event that occurs with probability 1
Target Population

• The “target population” is any entity satisfying the set of conditions that specify the universe. “Universe” and “Target Population” are synonyms
• Example 1: “Human population” All people, male and female, child and adult, living in a given geographic area at a particular date
• Example 2: “Establishment population” A business, industrial, service or governmental unit at a single location that distributes goods or performs services on a particular date (or during a given period)
Out-of-scope Population

• An entity that is outside either the geographic region under study or fails to meet another specific restriction imposed on the target population

• Example 1: When the in-scope population is “persons age 16 or over living in households,” persons age 15 or younger and all persons living in group quarters are out of scope

• Example 2: When the in-scope population is “employer establishments,” all establishments with zero employees are out of scope
Frame Population

• Set of target population, or universe, entities that can be selected into a sample or census
• Also called a sampling frame
• The frame population or sampling frame is the physical manifestation of the universe—if an entity is not on the frame (or one of the frames for multi-frame sampling), then it cannot be in the census or survey
• Simple cases: (single frame sampling) a list of all addresses to be sampled; list of all people to be sampled; list of all businesses to be sampled
• Complex sample designs use multiple frame populations to get better coverage of the target population or universe
• Complex frame example: Current Population Survey or Survey of Income and Program Participation (also: Enquête d’Emploi)
Geography

• The geography of a sampling frame assigns to every latitude and longitude a fundamental geographic area
• Geographic entities can be assembled uniquely by aggregating geographic areas
• The basic geographic entity for the U.S. Census is the “block”
Standard Hierarchy of Census Geography Entities
Entities

• In sampling frame development every geographic location (latitude and longitude) that contains a structure (natural or man-made) capable of originating economic activity is classified as a domicile, business, or both

• Entities are placed in the frame by declaring the target population to be humans beings (all domiciles including group quarters), economic (all businesses and service organizations), and government

• Notice that both for-profit and not-for-profit business activity are covered in the business entity scope
Population Demographics

• Human populations are usually categorized by current living quarters when designing demographic sampling frames
• Distinguish between household living quarters and group living quarters
• Frames based on landline, mobile telephone or Internet service provider are not domicile base
Business Entity Demographics

• Business entities have “establishments” as their basic unit
• Establishment: a business or industrial unit at a single location that distributes goods or performs services
• Establishments are collected into companies
• Business entity demographics separately track establishments (physical business locations) and companies (economic organizations owning establishments)
Business Entity Demographics

• Company: (or “enterprise”) all the establishments that operate under the ownership or control of a single organization. A company may be a commercial business, service, or membership organization
• A company may consist of one or several establishments
• A company may operate at one or several locations
• A company may operate in one or more economic activities
• A company includes all subsidiary organizations, all establishments that are majority-owned by the company or any subsidiary, and all the establishments that can be directed or managed by the company or any subsidiary.
Single-unit Companies (SU)

• Definition: Companies for which the location and the company are one and the same
• A single-unit business, service agency, or membership organization is one for which all the economic activity of the owner or owners is conducted at a single location
• Example: the “Shop Around The Corner” in the movie of the same name (or “You’ve got mail”)

30 May 2013 © John M. Abowd and others, 2013
Multi-unit Companies (MU)

• Definition: companies that have more than one location

• A multi-unit business, service organization, or governmental agency is one for which the owners conduct economic activity at more than one physical location

• Example 1: all the manufacturing and management locations of the General Motors Corporation constitute a multi-unit company

• Example 2: all the service delivery locations of the Salvation Army constitute a multi-unit company
Government Entities

• Definition: public entity created by the U.S. constitution, state constitutions or the statutes of a state

• In the United States these are divided into:
  – National government (U.S. constitution)
  – State government (state constitutions)
  – Local government (statutory entities created by states)
    • General purpose
    • Public school systems
    • Special districts
Business and Government Entity Activity Classifications

• An entity that is engaged in economic or governmental activity may be classified in several ways
  – Ownership: the legal form of organization (public/private; corporate, partnership, sole proprietorship)
  – Activity: An industry is the most detailed category available in North American Industrial Classification System to describe business activities
Business and Government Entity Activity Classifications

• NAICS provides hundreds of separate industry categories, unique categories that reflect different methods used to produce goods and services.

• Industry categories are used to classify, collect, process, publish, and analyze business statistics.

• [NAICS documentation](#)
Demographic Sampling Frames

• Comprehensive mapping of location of the in-scope population with standardized geography
• Comprehensive list of domiciles in the geographic area covered
• Characteristics of inhabitants of the domiciles (stratifying variables)
Economic Sampling Frames

• Comprehensive mapping of location of the in-scope activity with standardized geography
• Comprehensive list of addresses of business and government establishments in the geographic area covered
• Economic activity and size measures for the entity (stratifying variables)
• Updating of organizational structure (by survey)
Basic Relations Connecting Frames

• Geographic relations
• Business relations
Geographic Relations

• Economic and demographic frames are directly connected through the use of common geographic identifiers
• A single location can be associated with household (demographic) activity, economic activity, both, or neither (undeveloped)
• No U.S. statistical agency maintains a single geographically integrated frame for households and businesses
Business Relations

• Demographic and economic sampling frames can be connected by economic relations between the households and businesses
• Supplier-customer relations
• Employer-employee relations
Housing Frames

• Swedish registers
• Census MAF/TIGER system (MTdb)
• Housing address list updates
• How do you combine the information?
Register-based Population Censuses

• All personal addresses in Sweden are maintained in national registers
• The registers are updated by the individual when he/she moves using the national PIN
• All businesses and governmental activity use the registers to find the person
• The registers can be used as the sampling frame to conduct a census or survey
  – Swedish register-based census 2005
  – German register-based census
Employer Business Frame Populations

• Census Employer Business Register
• BLS Establishment Register
• Establishment births and deaths
• The problem of false births and deaths
• The problem of multiple activity codes
Census Employer Business Register

• Frame population consists of businesses that file income tax returns
• Employers are identified from the tax forms
• Multiunit and single unit businesses are determined in the Economic Census
• Updates to the MU/SU classification are determined by the annual Report of Organization Survey
• Overview from census.gov
Employment/Job Frame Populations

• A job is a relation between an employer and an employee
• Target population of jobs depends on definitions of employer and employee
• Frame population must be constructed from legal employment definitions
• No U.S. agency maintains a sampling frame for jobs
• Closest is the LEHD Infrastructure File System Employment History File at the Census Bureau
Job Frame Populations

- Dynamic frames
- Defining an employer
- Defining an employee
Job Frame Populations

• The problem of integration on the employer side
  – The job frame does not define a complete employer population frame, even if it is universal, unless the activity definitions for the employer and the employee report match

• The problem of integration on the employee side
  – The job frame does not define a complete individual population frame, even if it is universal, unless the individual activity definition only includes employment (no unemployment or non-participation)
Geographic Link Frames

- Transportation network frames
- Defining a workplace location
- Defining a residence location
- Frame maintenance for the workplace-residence pair
JOB FRAMES AND RELATED STATISTICS
A Different View

People

Firms and establishments
New Entity: Jobs
The View from the People Side
CPS, ACS, ...

People

Jobs

Firms and establishments
The View from the Firm Side Economic Census, QCEW
The View from the Firm Side: Occupational Employment Statistics

People
- Plumber, $x
- Manager, $y
- Designer, $z

Jobs

Firms and establishments
- 3
- 2
- 1
Dynamics of Jobs

People

Jobs

Firms and establishments
Establishment Censuses and Surveys

• Quarterly Census of Employment and Wages (QCEW)
  – Quarterly count of workers and wage earnings from employers (about 98% of all workers), a “near census”
  – Collected through state-federal partnership from administrative data (state UI systems, Unemployment Compensation for Federal Employees)
  – The base frame for BLS-managed surveys of establishments and firms
QCEW Detail

• Key record is a **employer**-filed mandatory record on covered workers and wages

• Establishment detail provided by **mandatory** Multiple Worksite Report, breaking out employment by establishment
  – However, compliance is not universal
  – Exception: Minnesota

• Information on location, industry, employment, wages
QCEW-based Surveys

• Job Openings and Labor Turnover Survey (JOLTS)
  – Data on vacancies, hires, separations
  – Survey of ~16,000 establishments

• Current Employment Statistics (CES)
  – Monthly survey of ~145,000 firms on hours, employment, earnings
  – Basis of monthly “Employment Situation”, together with CPS-derived household statistics
QCEW-based Statistics

• Occupational Employment Statistics (OES)
  – Sample of firms, and occupations within firms
  – Bi-annual, provides estimates of employment and wages for ~800 occupations
  – Published at national, state, metropolitan levels as well as for certain industries

• http://www.bls.gov/oes/
What’s Missing?

• Employers and establishment report on characteristics of their workers that they know
  – Occupation
  – Fact of employment
  – Wages
  – Tasks
  – Hiring/firing in general
• Workers report on characteristics of their employers and jobs that they know
  – Industry
  – Occupation
  – Wages
  – Hiring and firing as it affects them
  – Their family structure (marriage, children, health)
Linking Workers to Employers

• Validation studies
  – Mellow & Sider (1983), Bound & Krueger (1991), etc.

• Unemployment insurance wage records for specific states
  – Anderson & Meyer (1992, 1994, etc.)
National Studies Linking Universe Worker to Universe Firms/Establishments

• Late 1990s
  – Including Abowd, Kramarz, Margolis (1999) for France
  – Bingley & Westergård-Nielsen (1996), Belzil (2000) for Denmark
  – Others

• See Haltiwanger, Lane, Spletzer, Theeuwes and Troske (1999), *The Creation and Analysis of Employer-Employee Matched Data*
United States

• 1999: start of the LEHD program
• Resulting infrastructure links
  – Worker-level wage records from UI systems
  – Establishment-level QCEW reports
  – Demographic information available at the U.S. Census (including extracts of IRS 1040 and SSA Numident, augmented by Census Bureau staff)
  – Firm-level information from the Business Register and economic censuses and surveys
LEHD Program

• Support from NSF and other agencies

• Public-use products:
  – Quarterly Workforce Indicators
    • New amount of detail on workforce and the dynamics of the workforce at county/MSA/WIB level (by gender, by age, by detailed industry)
  – OnTheMap/LODES
    • Detail on links between workplaces and residences, at Census block level, with workplace and residence block detail on workforce (LODES is public-use data from OnTheMap graphical application)

• Restricted-access data:
  – LEHD Snapshot (S2004, S2008, S2011) in RDCs
Part 1:
Overview of Methodology

BASIC CONCEPTS OF THE LEHD INFRASTRUCTURE
Underlying Concepts: Construction of the LEHD Infrastructure

• Goals:
  • Understand concepts, basics of construction of LEHD infrastructure and processing
  • Understand QWI measures, how data is used in LEHD products
  • Understand differences between QWI measures and other measures
    • Related to definitions
    • Related to data sources/construction
Reference Materials

• QWI Comprehensive Index
  – Provides crosswalk between various naming conventions
  – Indicates how measures are used in online products

• Comparison of employment definitions
  – CPS vs. QCEW vs. LEHD

• QWI Cheatsheet
  – Reference for the structure of the public released QWI files
Additional References

- *The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators* – Primary reference for LEHD methodology
- Technical papers also available in the (historical) LEHD technical paper and CES working paper series on various topics
- Resources on the VirtualRDC for the Quarterly Workforce Indicators, OnTheMap, and the LEHD Infrastructure File System
Basic Concepts and Definitions

• Dates
  – Year and quarter
  – Boundary between quarters is the employment reference date

• Employer (SEIN)
  – single Unemployment Insurance (UI) account in a given state’s UI wage reporting system

• Location (SEINUNIT) - work place location

• Employee (PIK)
  – at least one employer reports earnings of at least one dollar of UI-covered earnings for an individual
Basic Concepts: Job

- Job (PIK-SEIN-SEINUNIT)
  - coupling of specific individual with specific employer and location in a given year/quarter

- The job is the basic unit of analysis within the LEHD Infrastructure

- Jobs are linked across years and quarters to develop longitudinal measures
Basic Concepts: Employment

• Level of measurement
  – Job (PIK-SEIN-SEINUNIT)

• Level of estimation
  – Worker’s job (PIK-SEIN-SEINUNIT)
  – Primary job (OnTheMap PIK-SEIN-SEINUNIT most earnings)
  – All jobs in a firm (SEIN) and location (SEINUNIT)
  – All jobs for a worker (PIK)

• Employment status
  – Point-in-time
  – Full quarter (continuous during quarter- stable)
Basic Concepts: Beginning of Period Employment

- Will reference as “b” or “B”
  - lowercase b – job level
  - uppercase B – jobs aggregated to establishment or higher level
- Primary measure of employment for QWI and OnTheMap
- Developed from job history
  - Defined when job is present in previous and current quarter
- Conceptually and empirically similar to QCEW Month 1 employment (Mon1)
  - Definitions, data sources, and methodology result in differences
### Employment History

- Jobs are linked across years and quarters to develop an individual’s employment history with a firm
  - PIK-SEIN-SEINUNIT level
- The reference quarter is noted at $t$
  - Earlier quarters are negative, later positive
- For calculation of measures,
  - RED indicates positive earnings
  - BLACK indicates zero earnings
  - BLUE (background) indicates time period not referenced
Details: Jobs

• See $m$ and $M$ on comprehensive index
• This variable is turned on ($m=1$) for every wage record in a state’s UI system that reports earnings of at least $\$1$ in $t$
• This is a job for $m$ (PIK-SEIN-SEINUNIT)
• This is a count of all persons ever paid by an employer at a location
  – By itself, it is not comparable to any other job-based statistic in the US system
• Released in QWI public use files as “EmpTotal” and labeled as “Employment reference quarter: Counts”
• Not reported in QWI Online, Industry Focus
Details: Employment – Beginning of Period

- See $b$ and $B$ on comprehensive index
- This variable is turned on whenever an individual has positive earnings in both the previous and current quarters
  - $m=1$ for last quarter $(t-1)$ and this quarter $(t)$
- $b=1$ means an individual was employed at a particular employer and location (PIK-SEIN-SEINUNIT) on the first calendar day of the quarter
- $B$ is the count of beginning of quarter employment for an employer location (SEIN-SEINUNIT)
- This is the main employment measure used in QWI and OnTheMap
Details: Employment – End of Period

• See $e$ and $E$ on comprehensive index

• This variable is turned on whenever an individual has positive earnings in both the current and next quarters
  – $m=1$ for this quarter ($t$) and next quarter ($t+1$)

• $e=1$ means an individual was employed at a particular employer and location (PIK-SEIN-SEINUNIT) on the last calendar day of the quarter

• $E$ is the count of end of quarter employment for an employer location (SEIN-SEINUNIT)

• This variable is only reported in the QWI public use files, as EmpEnd.
Details: Employment – Full Period

- See $f$ and $F$ on comprehensive index
- This variable is turned on whenever an individual has positive earnings in the last, current and next quarters
  
  $m=1$ for last quarter (t-1), this quarter (t) and next quarter (t +1)
- $f=1$ means an individual was employed at a particular employer and location (PIK-SEIN-SEINUNIT) throughout the current quarter
- $F$ is the count of full quarter employment for an employer location (SEIN-SEINUNIT)
- This variable is reported in the QWI public use files as EmpS and on OnTheMap as “Employment, Stable Jobs”
QWI Estimates:
Employment Measures
48 States, Private Sector Only

- Flow employment
- Beginning-of-period employment
- Full-quarter employment
QWI Estimates: Employment Percent Change
48 States, Private Sector Only
Basic Concepts: Earnings

• Point in time earnings
  – Defined for a reference group meeting a particular employment definition at a point in time (End-of-quarter employment)

• Full-quarter earnings
  – Defined for a reference group of full-quarter employment (Full-quarter employment, Full-quarter hires, Full-quarter new hires, Full-quarter separations)

• Average earnings based on wage record earnings for the indicated quarter divided by 3 (monthly estimate)
  • In graphics, dollar sign ($) indicates reference quarter for earnings
**Details: Earnings – End-of-Quarter**

- See Z_W2 on comprehensive index
- Earnings in quarter $t$ are accumulated into W2 whenever an individual has positive earnings in the current and next quarter
  - $m=1$ for this quarter ($t$) and next quarter ($t+1$)
- Z_W2 is the average monthly earnings of end-of-quarter employees for an employer location (SEIN-SEINUNIT)
- This variable is reported in the QWI public use files as “EarnEnd”
Details: Earnings – Full Quarter

• See $Z_{W3}$ on comprehensive index

• Earnings in quarter $t$ are accumulated into $W3$ whenever an individual has positive earnings in the last, current and next quarters

  \[ m=1 \text{ for last quarter } (t-1), \text{ this quarter } (t) \text{ and next quarter } (t+1) \]

• $Z_{W3}$ is the average monthly earnings of full quarter employees for an employer location (SEIN-SEINUNIT)

• This variable is reported in the QWI public use files as “EarnS”, on QWI Online as “Avg Monthly Earnings”, in Industry Focus as “Average monthly earnings for all workers” (and for growth), and on OnTheMap as “Average Monthly Earnings, Stable Jobs”
Details: Earnings Total Payroll

- See $w_1$ and $W_1$ on comprehensive index
- Earnings in quarter $t$ are accumulated into $W_1$ whenever an individual has positive earnings current quarter
  
  \[ m=1 \text{ for this quarter (t)} \]
- $W_1$ is the total quarterly earnings of employees for an employer location (SEIN-SEINUNIT)
- This variable is reported in the QWI public use files as “Payroll”
- This variable covers the same concept as “Total Quarterly Wages” on public-use QCEW.
QWI Estimates:
Average Monthly Earnings
48 States, Private Sector Only

Average Monthly Earnings of full-quarter employees
Average earnings of end-of-period employees
Job-Based Measures: Employment Flows

• Measures use longitudinal job history to identify employment patterns
  – Previous and following quarters are referenced to check for positive earnings
Details: Employment Flows-Accessions

- See $a1$ and $A$ on comprehensive index
- An accession ($a1$) is turned on whenever an individual has positive earnings in the current and not in the previous quarter
  
  $m=0$ for last quarter (t-1) and $m=1$ this quarter (t)
- $a1=1$ means an individual was newly employed at a particular employer and location (PIK-SEIN-SEINUNIT) during the current quarter
- $A$ is the count of all accessions for an employer location (SEIN-SEINUNIT)
- This variable is reported in the QWI public use files as HirA
Details: Employment Flows-

New Hires

• See \( h1 \) and \( H \) on comprehensive index
• A new hire (\( h1 \)) is turned on whenever an individual has an accession, with no earnings from the employer during the previous four quarters
  \[ m=0 \text{ for last four quarters (t-4 to t-1) and } m=1 \text{ this quarter (t)} \]
• \( h1=1 \) means an individual was newly hired at a particular employer and location (PIK-SEIN-SEINUNIT) during the current quarter
• \( H \) is the count of all new hires for an employer location (SEIN-SEINUNIT)
• This variable is reported in the QWI public use files as “HirN” and on QWI Online as “New Hires”
Details: Employment Flows-Recalls

- See \( r_1 \) and \( R \) on comprehensive index
- A recall (\( r_1 \)) is turned on whenever an individual has an accession, and also received earnings from the employer during the earlier three quarters
  
  \[
  m = 1 \text{ for at least one of quarters } (t-4, t-3, t-2), \ m = 0 \text{ for last quarter } (t-1) \text{ and } m = 1 \text{ this quarter } (t)
  \]

- \( r_1 = 1 \) means an individual was recalled at a particular employer and location (PIK-SEIN-SEINUNIT) during the current quarter
- \( R \) is the count of all recalls for an employer location (SEIN-SEINUNIT)
- This variable is reported in the QWI public use files as “HirR”
Details: Employment Flows-Separations

- See $s1$ and $S$ on comprehensive index
- A separation ($s1$) is turned on whenever an individual has no wage record in the next quarter but is present this quarter
  
  \[ m=1 \text{ for current quarter (t) and } m=0 \text{ next quarter (t+1)} \]
- $s1=1$ means an individual was separated at a particular employer and location (PIK-SEIN-SEINUNIT) during the current quarter
- $S$ is the count of all separations for an employer location (SEIN-SEINUNIT)
- This variable is reported in the QWI public use files as “Sep,” and on QWI Online as “Separations”
QWI Estimates:
Accessions, Separations, and New Hires
48 States, Private Sector Only

(in Millions)

Accessions
Separations
New hires

© John M. Abowd and others, 2013
QWI Estimates:
Accessions, Separations, New Hires
Percent Change
48 States, Private Sector Only

Delta Accessions
Delta Separations
Delta New hires
Details: Employment Flows-Full Quarter Accessions

- See $a_3$ and $FA$ on comprehensive index
- A full-quarter accession ($a_3$) is turned on whenever an individual is full-quarter employed this quarter, but not last quarter
  - $m=0$ for quarter $(t-2)$ and $m=1$ last quarter $(t-1)$, this quarter $(t)$ and next quarter $(t+1)$
- $a_3=1$ means an individual was full-quarter employed at a particular employer and location (PIK-SEIN-SEINUNIT) during the current quarter for the first time
- $FA$ is the count of all full-quarter accessions for an employer location (SEIN-SEINUNIT)
- This variable is reported in the QWI public use files as “HirAS”
Details: Employment Flows-
Full-Quarter New Hires

• See $h_3$ and $H_3$ on comprehensive index
• A full-quarter new hire ($h_3$) is turned on whenever an individual has a full-quarter accession, with no association with the employer during the four quarters preceding hire
  
  $m=0$ for four quarters ($t-5$ to $t-2$) and $m=1$ last quarter ($t-1$), this quarter ($t$) and next quarter ($t+1$)
• $h_3=1$ means an individual was newly hired at a particular employer and location (PIK-SEIN-SEINUNIT) last quarter and became a full-quarter employee during the current quarter
• $H_3$ is the count of all new full-quarter hires at an employer location (SEIN-SEINUNIT)
• This variable is reported in the QWI public use files as “HirNS”, on Industry Focus as “Number of New Hires” and on OnTheMap as “New Hires, Stable Jobs”
Details: Employment Flows-
Full Quarter Separations

• See \( s3 \) and \( FS \) on comprehensive index
• A full-quarter separation \((s3)\) is turned on whenever an individual has no wage record in the next quarter but was full-quarter employed in the previous quarter
  – \( m=1 \) for quarter \((t-2)\), last quarter \((t-1)\), current quarter \((t)\) and \( m=0 \) next quarter \((t+1)\)
• \( s3=1 \) means an individual was separated at a particular employer and location from a job that was full-quarter during the previous quarter (can’t be full-quarter this quarter)
• \( FS \) is the count of all full-quarter separations for an employer location \((SEIN-SEINUNIT)\)
• This variable is reported in the QWI public use files as “SepS” and on OnTheMap as “Separations, Stable Jobs”
QWI Estimates – Full Quarter: Accessions, Separations, and New Hires
48 States, Private Sector Only

Flow into full-quarter employment
Full-quarter new hires
Flow out of full-quarter employment

© John M. Abowd and others, 2013
QWI Estimates – Full Quarter: Accessions, Separations, New Hires
Percent Change
48 States, Private Sector Only

Delta Flow into full-quarter employment
Delta Full-quarter new hires
Delta Flow out of full-quarter employment

© John M. Abowd and others, 2013 82
Firm-Based Measures: Flows, Creations, DestrucNons

• Calculated at the establishment (SEIN-SEINUNIT) level
  – Also calculated at establishment level for each demographic segment

• Use Emp (B), EmpEnd (E), and EmpS (F) employment measures as inputs

• Measures are aggregated across establishments
  – Separately aggregated by demographic segment (caution to follow)
Firm Job Flows: What is being measured?

• These measures display dynamics at the firm (or establishment) level, which is between the individual level and the top level aggregates

• Measures may be related as follows:
  – Even if an industry is declining in total employment, individual firms may be growing
    • Net Job Flows < 0, Job Creation > 0
  – Even if a firm is shrinking, it still may be hiring individuals
    • Job Destruction > 0, New Hires > 0
Details: Firm Job Flows – Job Creations

- See $JC$ on comprehensive index
- A job creation is only defined at the employer-location level. A job is created when an employer has greater end of period employment than beginning of period employment
  
  $$JC = \max(E - B, 0)$$
- $JC$ is the count of all job creations for an employer location (SEIN-SEINUNIT)
- This variable is reported in the QWI public use files as “FrmJbGn,” on QWI Online as “Job Creation,” and on OnTheMap in the QWI report as “Firm Job Gain”
Details: Firm Job Flows – Job Destinations

- See $JD$ on comprehensive index
- A job destruction is only defined at the employer-location level. A job is destroyed when an employer has greater beginning of period employment than end of period employment
  \[
  JD = \max(B - E, 0)
  \]
- $JD$ is the count of all job destructions for an employer location (SEIN-SEINUNIT)
- This variable is reported in the QWI public use files as “FrmJbLs” and on OnTheMap in the QWI report as “Firm Job Loss”

• See JF on comprehensive index

• Net flows are the difference between creations and destructions

\[ JF = JC - JD = E - B \]

• JF is the count of all net job flows for an employer location (SEIN-SEINUNIT)

• This variable is reported in the QWI public use files as “FrmJbC” and on QWI Online as “Net Job Flows”
QWI Estimates:
Firm Job Flows
48 States, Private Sector Only
LEHD PUBLIC-USE STATISTICS: QWI
LEHD Public-use Data
QWI Measures

• Jobs are aggregated to generate estimates of QWI measures at desired levels
• Measures are all built from linearly aggregable components
  – Means are calculated, rather than medians
• Some measures on the public use files may be further aggregated, others not
  – Some firm-based flow measures are not aggregable
  – Components of average earnings measures not available
  – Turnover may be recalculated using component pieces
LEHD Public-use Data
QWI Aggregation Levels - Establishment

• Establishment level characteristics for aggregation:
  – Geography
    • State totals
    • County, Metro, Workforce Investment Board areas
  – Industry
    • All industries
    • NAICS Sectors, Sub-sectors (3-digit), Industry groups (4-digit)
  – Ownership
    • All (1-5)
    • Private-only (5)
  – Firm size (5 detailed categories)
  – Firm age (5 detailed categories)
LEHD Public-use Data: QWI Aggregation Levels - Employee

• Employee level characteristics for aggregation
  – Age
    • All ages
    • 14-18, 19-24, 25-34, 35-44, 45-54, 55-64, 65-99
    • (Workforce Investment Act age categories)
  – Sex
    • Both sexes
    • Male, Female
  – Education (since 2011Q1)
    • Only for individuals age 25+
      – Less than a High School Diploma
      – High School Diploma, No College
      – Some college or Associate’s Degree
      – Bachelor’s Degree or Above
LEHD Public-use Data:
QWI Aggregation Levels - Employee

• Employee level characteristics for aggregation
  – Race (since 2011Q1)
    • OMB categories
      – White alone
      – African-American or Black alone
      – Asian or Pacific alone
      – Native Hawaiian or Other Pacific Islander alone
      – American Indian or Alaska Native alone
      – Two or More Races
  – Ethnicity (since 2011Q1)
    • Hispanic or Latino
    • Not Hispanic or Latino
LEHD Public Use: QWI Measures

• The QWI public use files contain a series of 30 measures
  – Employment: stock/flow measures
  – Earnings

• These measures are reported for all of the aggregation listed in previous slides

• The entire time series is re-estimated and re-released in every data release (“R2012Q3”)

• A broader range of theoretical measures are defined in the technical papers, though many are not estimated in regular production
QWI Public Use Files: File Naming

-qwi_ak_wia_county_naicssec_all.csv.gz

- Components:
  - State
  - Demographic breakdown
    - Currently all files are age by sex, noted *wia*
    - Expect this notation to change in the future
  - Region
  - Industry aggregation level
  - Ownership: *all* or *private-only*

- Reference: [http://www.vrdc.cornell.edu/qwipu/QWI-cheatsheet.txt](http://www.vrdc.cornell.edu/qwipu/QWI-cheatsheet.txt)
QWI Public Use Files: Status Flags

• Data quality flags
  – start with s, followed by name of variable
    • -2 no data available in this category for this quarter
    • -1 data not available to compute this estimate
    • 0 zero employment estimated or zero estimated denominator in a ratio, zero released
    • 1 OK
    • 5 Value suppressed because it does not meet US Census Bureau publication standards.
    • 9 data significantly distorted, distorted value released
  – Suppression does not mean zero
QWI Public Use Files: Other Notes

• Text files, comma-separated values, with labels for classifications
  – Compressed using gzip
    • PC-based software can be used to decompress
    • If using on UNIX, file may be processed without decompressing (see SAS filename statements)

• Data can be processed using statistical software (e.g., SAS, STATA, SPSS) or database software (e.g., Access)
  – Excel 2007 may also be able to manipulate complete QWI files

• Most aggregations contain both totals and components
  – Be sure to set all appropriate filters to extract only the records you want
UNDERSTANDING LEHD SOURCES AND STRUCTURE
Data Flow View of the LEHD Infrastructure
# QWI Production Process: Key Stages

<table>
<thead>
<tr>
<th>Process</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHF</td>
<td>Employment History File</td>
<td>Longitudinal employment history for individuals</td>
</tr>
<tr>
<td>GAL</td>
<td>Geocoded Address List</td>
<td>Assigns coordinates to establishments</td>
</tr>
<tr>
<td>ECF</td>
<td>Employer Characteristics File</td>
<td>Longitudinal establishment-level data: employment, NAICS, geography, ownership; Imputes when necessary; Maintains “fuzz factors”; calculates weights</td>
</tr>
<tr>
<td>SPF</td>
<td>Successor-Predecessor File</td>
<td>Identifies successor-predecessor relationships based on employment flows</td>
</tr>
<tr>
<td>ICF</td>
<td>Individual Characteristics File</td>
<td>Individual level demographic data; Imputes when necessary</td>
</tr>
<tr>
<td>U2W</td>
<td>Unit-to-Worker</td>
<td>Multiply-imputes establishment to jobs at multi-establishment firms</td>
</tr>
<tr>
<td>QWI</td>
<td>Quarterly Workforce Indicators</td>
<td>Generate final estimates, apply weighting, confidentiality protections ((QWIPU = QWI Public Use file))</td>
</tr>
</tbody>
</table>
LEHD Processing: Merging QCEW and UI Data

Quarterly Census of Employment and Wages
- Firm and Establishment (Single/Multi-unit)
- Geography
- Industry
- Ownership

Unemployment Insurance Wage Records
- Firm-Worker (most states)
  OR
- Establishment-Worker (Minnesota only)
- Wages
- Job history
- Link to demography

UI Account Number
- Firm Level (SEIN)
- Establishment Level (SEIN-SEINUNIT) Minnesota only
LEHD Processing: Successor-Predecessor

• Adjustments to account for administrative changes to firms
  – mergers, divestitures, etc.

• Transitions may be identified through:
  – report on the QCEW
    • firm level and establishment level
  – finding large employment flow from the individual wage records
    • firm level only

• Individual job history at predecessor is concatenated with job history for same PIK at successor for purposes of calculating QWI measures
  – Important CAVEAT for RDC work: this does NOT generate a new firm identifier – researchers need to apply similar logic to their research extracts
LEHD Processing: Unit-to-Worker Impute

• Necessary to impute establishment to a job when not available
  – Currently only Minnesota reports establishments on wage data

• Individual job histories are assembled

• Establishment (multiply) imputed to longitudinal job, with the following predictors:
  – Proximity of residence to establishment
  – Size of establishment

• Establishment history (allowing for predecessors) must be consistent with individual job history
LEHD Processing: Weighting

• QWI B is benchmarked against QCEW Mon1 employment

• Firm-level weights (within bounds) are applied to adjust employment towards Mon1 employment

• Secondary weights are applied to match statewide private-only employment

• Weights are calculated at ECF stage, applied at QWI
LEHD Processing: QCEW-QWI Differences

- Sub-state adjustments are not currently applied to QWI data
- While state employment totals should be quite close, sub-state estimates will display deviations from benchmark
  - County, industry employment totals, or smaller cells
- These differences can come from any of a number of QWI-specific processing steps
  - Specific differences observed in the data may also result from an interaction of several sources of deviation
Causes of Differences:
Measure Definition

• B and Month 1 do not capture exactly the same universe
  – An individual may count towards either one of the measures, but not towards the other

• Differences generally minor, but may be noticeable in some industries with particular seasonal patterns
  – e.g., education, agriculture
Causes of Differences:
BLS Data Editing

• LEHD data receipts
  – Before 2004 LEHD received BLS edited data from state partners
  – Since 2004 LEHD does not receive BLS edited data from states (CIPSEA)
• BLS QCEW file may be edited/different from that which LEHD receives
  – Completeness
  – Imputed employment
  – Industry/geography changes
• Statewide totals are close (<1% off)
• LEHD QA will periodically note BLS QCEW data inconsistent with internal LEHD QCEW micro-data
Causes of Differences: Noise Infusion ("Fuzzing")

• Why infuse noise into data?
  – Reduce the amount of cell suppression while preserving confidentiality and analytic validity

• Properties of noise
  – Every data item is distorted by a minimum amount
  – For a given workplace, data are always distorted in the same direction, by the same percentage in every period and release of QWI’s
  – When aggregated, the effects of the distortion cancel out for the vast majority of the estimates
Noise infusion in QWI

Causes of Differences: Noise Infusion ("Fuzzing")

• QWI statistics are flagged when the value is significantly distorted (Status flag 9)
• See infrastructure document, section 6, for more details
• More about this later in the lecture on Confidentiality Protection
Causes of Differences: UI Wage Data Reporting

• Firm may fail to report wage records
  – QCEW still reported or imputed

• Firm may report wage records and QCEW records on different account numbers
  – Successor/predecessor mistiming
  – Public sector issues

• PIK (SSN) miscoding prevents linking wage records to same longitudinal job (Abowd & Vilhuber, 2005)
Causes of Differences: Industry Assignment

- Most establishments are assigned based on the reported NAICS_AUX
- For earlier years in the data series, the reported SIC code is probabilistically mapped to the current NAICS codes
  - Imputes may also be used for transitions between 1997, 2002, and 2007 NAICS
- LDB data are used for NAICS back-coding purposes when the file has been provided by state
- Variations in algorithms between LEHD and BLS may result in differences
  - NAICS sector 55 (management of companies) displays particular issues during SIC-NAICS transition
Causes of Differences: Geographic Coding

• LEHD performs own geo-coding of addresses
  – Generates lat-long for distance measures, allows custom geography

• Address data are processed along with address data from other sources

• Results may differ from BLS assignments
  – Marginal shift over county line
  – Significant relocation

• Effort currently underway to reengineer LEHD geographic assignment to improve results
Causes of Differences: Multiple Worksites (U2W)

• QCEW can report Mon1 by building directly from establishment (with geo/industry info)

• LEHD “No transfer” assumption a single job spell to be reported to the same establishment
  – Job spell – PIK-SEIN relationship that does not contain four consecutive quarters with zero earnings.

• A change in firm structure can make it impossible to replicate counts given constraint
  – Long-term differences may result from new, large establishments appearing without predecessor
Causes of Differences: Successor-Predecessor

• QCEW can, again, build up estimates directly from establishment
  – Does not matter for month1 purpose if predecessor existed

• LEHD must have information from previous and following quarters for range of measures
  – If a new firm appears, and that firm does not have a predecessor (with same employees), jobs at the new firm will not count towards primary LEHD B employment in that quarter
Data Irregularity: Missing UI Records

• Impact of large firm that fails to report UI wage data (or reports late) in 2009Q2.

<table>
<thead>
<tr>
<th></th>
<th>M EmpTotal</th>
<th>B Emp</th>
<th>E EmpEnd</th>
<th>F EmpS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009 Q1</td>
<td>-</td>
<td>-</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>2009 Q2</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>2009 Q3</td>
<td>-</td>
<td>↓</td>
<td>-</td>
<td>↓</td>
</tr>
<tr>
<td>2009 Q4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Data Irregularity: Spike in Wage Records

- Impact of large firm that displays unusual spike for only 2009Q2
  e.g., back pay for a court settlement

<table>
<thead>
<tr>
<th></th>
<th>M EmpTotal</th>
<th>B Emp</th>
<th>E EmpEnd</th>
<th>F EmpS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009 Q1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Data Irregularity: Unidentified Succ-Pred

- Firm reported under account X in 2009Q1, account Y in 2009Q2 (same geography, industry, job count);
  - Transition not identified in LEHD processing

<table>
<thead>
<tr>
<th></th>
<th>M EmpTotal</th>
<th>B Emp</th>
<th>E EmpEnd</th>
<th>F EmpS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009 Q1</td>
<td>-</td>
<td>-</td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td>2009 Q2</td>
<td>-</td>
<td>▼</td>
<td>-</td>
<td>▼</td>
</tr>
<tr>
<td>2009 Q3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Data Irregularity: Missing UI Records

• Impact of large firm that fails to report UI wage data (or reports late) in 2009Q2.
  – Assume most workers had stable employment before and after

<table>
<thead>
<tr>
<th></th>
<th>A HirA</th>
<th>H HirN</th>
<th>S Sep</th>
<th>FA HirAS</th>
<th>H3 HirNS</th>
<th>FS SepS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009 Q1</td>
<td>-</td>
<td>-</td>
<td>↑</td>
<td>-</td>
<td>-</td>
<td>↑</td>
</tr>
<tr>
<td>2009 Q2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q3</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>↑</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Data Irregularity: Spike in Wage Records

- Impact of large firm that displays unusual spike for only 2009Q2
  e.g., back pay for a court settlement for fired workers

<table>
<thead>
<tr>
<th></th>
<th>A HirA</th>
<th>H HirN</th>
<th>S Sep</th>
<th>FA HirAS</th>
<th>H3 HirNS</th>
<th>FS SepS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009 Q1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q2</td>
<td>↑</td>
<td>↑/-</td>
<td>↑</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* If workers had not received earnings in the past year, they would be new hires
Data Irregularity: Unidentified Successor-Predecessor

• Firm reported under account X in 2009Q1, account Y in 2009Q2 (same geography, industry, job count)
  – Transition not identified in LEHD processing
  – Assume most workers had stable employment before and after

<table>
<thead>
<tr>
<th></th>
<th>A HirA</th>
<th>H HirN</th>
<th>S Sep</th>
<th>FA HirAS</th>
<th>H3 HirNS</th>
<th>FS SepS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009 Q1</td>
<td>-</td>
<td>-</td>
<td>↑</td>
<td>-</td>
<td>-</td>
<td>↑</td>
</tr>
<tr>
<td>2009 Q2</td>
<td>↑</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>↑</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009 Q4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Firm Job Flows: Be Careful about Aggregation

- Note that for categories like age and sex, the published net job flows for the subcategories will sum to the margin.
- But for gross Job Creation and gross Job Destruction this is not true.
- (Job Creation for men) + (Job Creation for women) does not equal (total Job Creation)
  - For example, a job could be created at a firm and filled by a woman, while another job at the same firm is destroyed, previously filled by a man.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Creation</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Job Destruction</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Net Job Flows</td>
<td>-1</td>
<td>+1</td>
<td>0</td>
</tr>
</tbody>
</table>
Summary

• Households/Individuals, Business/Establishments and Jobs all contribute to longitudinally integrated labor market microdata

• Public use products can shed much light on the data and should be mastered before attempting the microdata