Longitudinal Employer - Household Dynamics

Technical paper No. TP-2002-13

The Creation of the Employment Dynamics Estimates

Date	:	July 26, 2002
Prepared by	:	John M. Abowd, Paul A. Lengermann, and Lars Vilhuber
Contact	:	Ronald Prevost (Ronald.C.Prevost@census.gov)
		U.S. Census Bureau, LEHD Program
		FB 2138-3
		4700 Silver Hill Rd.
		Suitland, MD 20233 USA

This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications, and is released to inform interested parties of ongoing research and to encourage discussion of work in progress. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grant SES-9978093 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging, and the Alfred P. Sloan Foundation. The views expressed herein are attributable only to the author(s) and do not represent the views of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau is preparing to support external researchers' use of these data; please contact Ronald Prevost (Ronald.C.Prevost@census.gov), U.S. Census Bureau, LEHD Project, FB 2138-3, 4700 Silver Hill Rd., Suitland, MD 20233, USA.

The Creation of the Employment Dynamics Estimates

John M. Abowd, Paul A. Lengermann, and Lars Vilhuber*

July 26, 2002

1 Introduction

The Longitudinal Employer Household Dynamics (LEHD) Program is a new state/federal partnership between the Census Bureau and ten states. Both sides gain from this partnership. States fulfill their mandate of providing high quality local labor market information to their customers. The Census Bureau uses state UI wage record and ES202 data to fulfill its Title 13 mandate: improving the Census Bureau's economic and demographic censuses, surveys, and intercensal population estimates.

The Memoranda of Understanding (MOU) between the Census Bureau and the state partners specify that this is a voluntary partnership. Research beyond that specified in the MOU must have the express written authorization of the state data custodian. States receive three key products from Census: (1) quarterly employment indicators about the state economy at detailed industry and geography; (2) enhanced UI wage records; and (3) information about successor/predecessor firms.

The Census Bureau receives UI wage records and ES-202 establishment records from the states. As a part of its Title 13 mission, the Bureau then uses these products to integrate information about the individuals (place of residence, sex, birth date, place of birth, race, education) with information about the employer (place of work, industry, employment, sales). These integrated files are used to improve the Census Bureau's demographic surveys, like the Current Population Survey, the Survey of Income and Program Participation, and the American Community Survey. They are also used to improve the Census Bureau's Business Register, which is the sampling frame for all its economic data and the initial contact frame for the Economic Census.

This document describes how the LEHD Program uses these integrated files to produce the Employment Dynamics Estimates, which are the source of the quarterly employment indicators returned to the states. In the following section, we provide definitions of employer, employee, job, and all associated worker flow, job flow, and earnings measures. We then give an overview of the different raw data inputs and how they are treated and adjusted in Section 3. In a system that focusses on the dynamics at the individual and firm level, proper identification of the entities is important, and we briefly highlight the steps undertaken to this end, although the finer detail is available in separate documents. The raw data are then aggregated and standardized into a series of component files, as described in Section 4. Finally, Section 5 illustrates how they are brought together to create the EDE statistics. It will soon become clear to the reader that the level of detail potentially available with these statistics requires special attention to the confidentiality of the the underlying entities. How their identity is protected is described in Section 6. A glimpse at some examples from the final statistics are provided in Section 7.

*The authors acknowledge the substantial contributions of the staff and senior research fellows of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Program. This document, based on other LEHD Program papers and brochures, was prepared for the NBER Summer Institute Conference on Personnel Economics, August 1-2, 2002. This research is a part of the LEHD Program at the U.S. Census Bureau, which is partially supported by the National Science Foundation Grant SES-9978093 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging, and the Alfred P. Sloan Foundation. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the U.S. Census Bureau, Cornell University, or the National Science Foundation. Confidential data from the LEHD Program were used in this paper. The U.S. Census Bureau is preparing to support external researchers' use of these data under a protocol to be released in the near future. For further information, contact Ronald C. Prevost LEHD Program Director (ronald.c.prevost@census.gov).

2 Fundamental Concepts

2.1 Dates

The EDE is a quarterly data system with calendar year timing. We use the notation YYYY:Q to refer to a year and quarter combination. For example, 1999:4 refers to the fourth quarter of 1999, which includes the months October, November, and December.

2.2 Employer

An employer in the EDE system consists of a single Unemployment Insurance (UI) account in a given state's UI wage reporting system. For statistical purposes the EDE system creates an employer identifier called an State Employer Identification Number (SEIN) from the UI-account number and information about the state (FIPS code). Thus, within the EDE system, the SEIN is a unique identifier within and across states but the entity to which it refers is a UI account.

2.3 Employee

Individual employees are identified by their Social Security Numbers (SSN) on the UI wage records that provide the input to the EDE. To protect privacy and confidentialty of the SSN and the individual's name, a different branch of the Census Bureau removes the name and replaces the SSN with an internal Census identifier called a Protected Identity Key (PIK).

2.4 Job

The EDE system definition of a job is the association of an individual (PIK) with an employer (SEIN) in a given year and quarter. The EDE system stores the entire history of every job that an individual holds. Estimates are based on the definitions presented below, which formalize how the EDE system estimates the start of a job (accession), employment status (beginning- and end-of-quarter employment), continuous employment (full-quarter employment), the end of a job (separation), and average earnings for different groups.

2.5 Unemployment Insurance wage records (the EDE system universe)

The Employment Dynamics Estimates are built upon concepts that begin with the report of an individual's UI-covered earnings by an employing entity (SEIN). An individual's UI wage record enters the EDE system if at least one employer reports earnings of at least one dollar for that individual (PIK) during the quarter. Thus, the job must produce at least one dollar of UI-covered earnings during a given quarter to count in the EDE system. The presence of this valid UI wage record in the EDE system triggers the beginning of calculations that estimate whether that individual was employed at the beginning of the quarter, at the end of the quarter, and continuously throughout the quarter. These designations are discussed below. Once these point-in-time employment measures have been estimated for the individual, further analysis of the individual's wage records results in estimates of full-quarter employment, accessions, separations (point-in-time and full-quarter), job creations and destructions, and a variety of full-quarter average earnings measures.

2.6 Employment at a point in time

Employment is estimated at two points in time during the quarter, corresponding to the first and last calendar days. An individual is defined as employed at the beginning of the quarter when that individual has valid UI wage records for the current quarter and the preceding quarter. Both records must apply to the same employer (SEIN). An individul is defined as employed at the end of the quarter when that individual has valid UI wage records for the current quarter and the subsequent quarter. Again, both records must show the same employer. The EDE system uses beginning and end of quarter employment as the basis for constructing worker and job flows. In addition, these measures are used to check the external consistency of the data, since a variety of employment estimates are available as point-in-time measures. Many federal statistics are based upon estimates of employment as of the 12th day of particular months. The Census Bureau uses March 12 as the reference date for employment measures contained in its Business Register

and on the Economic Censuses and Surveys. The BLS "Covered Employment and Wages (CEW)" series, which is based on the ES-202 data, use the 12th of each month as the reference date for employment. The EDE system cannot use exactly the same reference date as these other systems because UI wage reports do not specify additional detail regarding the timing of these payments. EDE research has shown that the point-in-time definitions used to estimate beginning and end of quarter employment track the CEW month one employment estimates well at the level of an employer (SEIN).

2.7 Employment for a full quarter

The concept of full quarter employment estimates individuals who are likely to have been continuously employed throughout the quarter at a given employer. An indivdual is defined as full-quarter-employed if that individual has valid UI-wage records in the current quarter, the preceding quarter, and the subsequent quarter at the same employer (SEIN). That is, in terms of the point-in-time definitions, if the individual is employed at the same employer at both the beginning and end of the quarter, then the individual is considered full-quarter employed in the EDE system.

Consider the following example. Suppose that an individual has valid UI wage records at employer *A* in 1999:2, 1999:3, and 1999:4. This individual does not have a valid UI wage record at employer *A* in 1999:1 or 2000:1. Then, according to the definitions above, the individual is employed at the end of 1999:2, the beginning and end of 1999:3, and the beginning of 1999:4 at employer *A*. The EDE system treats this individual as a full-quarter employee in 1999:3 but not in 1999:2 or 1999:4. Full-quarter status is not defined for either the first or last quarter of available data.

2.8 Point-in-time estimates of accession and separation

An accession occurs in the EDE system when it encounters the first valid UI wage record for a job (an individual (PIK)-employer (SEIN) pair). Accessions are not defined for the first quarter of available data from a given state. The EDE definition of an accession can be interpreted as an estimate of the number of new employees added to the payroll of the employer (SEIN) during the quarter. The individuals who acceded to a particular employer were not employed by that employer during the previous quarter but received at least one dollar of UI-covered earnings during the quarter of accession.

A separation occurs in the current quarter of the EDE system when it encounters no valid UI wage record for an individual-employer pair in the subsequent quarter. This definition of separation can be interpreted as an estimate of the number of employees who left the employer during the current quarter. These individuals received UI-covered earnings during the current quarter but did not receive any UI-covered earnings in the next quarter from this employer. Separations are not defined for the last quarter of available data.

2.9 Accession and separation from full-quarter employment

Full-quarter employment is not a point-in-time concept. Full-quarter accession refers to the quarter in which in-dividual first attains full-quarter employment status at a given employer. Full-quarter separation occurs in the last full-quarter that an individual worked for a given employer.

As noted above, full-quarter employment refers to an estimate of the number of employees who were employed at a given employer during the entire quarter. An accession to full-quarter employment, then, involves two additional conditions that are not relevant for ordinary accessions. First, the individual (PIK) must still be employed at the end of the quarter at the same employer (SEIN) for which the ordinary accession is defined. At this point (the end of the quarter where the accession occured and the beginning of the next quarter) the individual has acceded to continuing-quarter status. An accession to continuing-quarter status means that the individual acceded in the current quarter and is end-of-quarter employed. Next the EDE system must check for the possibility that the individual becomes a full-quarter employee in the subsequent quarter. An accession to full-quarter status occurs if the individual acceded in the previous quarter, and is employed at both the beginning and end of the current quarter. Consider the following example. An individual's first valid UI wage record with employer *A* occurs in 1999:2. The individual, thus acceded in 1999:2. The same individual has a valid wage record with employer *A* in 1999:3. The EDE system treats this individual as end-of-quarter employed in 1999:2 and beginning of quarter employed in 1999:3. The individual, thus, acceded to continuing-quarter status in 1999:2. If the individual also has a valid UI wage record at employer *A* in 1999:4, then the individual is full-quarter employed in 1999:3. Since 1999:3 is the first quarter of full-quarter employment, the EDE system considers this individual an accession to full-quarter employment in 1999:3.

Full-quarter separation works much the same way. One must be careful about the timing, however. If an individual separates in the current quarter, then the EDE system looks at the preceding quarter to determine if the individual was employed at the beginning of the current quarter. An individual who separates in a quarter in which that person was employed at the beginning of the quarter is a separation from continuing-quarter status in the current quarter. Finally, the EDE system checks to see if the individual was a full-quarter employee in the preceding quarter. An indivdidual who was a full quarter employee in the previous quarter is treated as a full-quarter separation in the quarter in which that person actually separates. Note, therefore, that the definition of full-quarter separation preserves the timing of the actual separation (current quarter) but restricts the estimate to those individuals who were full-quarter status in the preceding quarter. For example, suppose that an individual separates from employer *A* in 1999:3. This means that the individual had a valid UI wage record at employer *A* in 1999:3 but did not have a valid UI wage record at employer *A* in 1999:4. The separation is dated 1999:3. Suppose that the individual had a valid UI wage record at employer *A* in 1999:2. Then, a separation from continuing quarter status occured in 1999:3. Finally, suppose that this individual had a valid UI wage record at employer *A* in 1999:1. Then, this individual was a full-quarter employee at employer *A* in 1999:2. The EDE system records a full-quarter separation in 1999:3.

2.10 Point-in-time estimates of new hires and recalls

The EDE system refines the concept of accession into two subcategories: new hires and recalls. In order to do this, the EDE system looks at a full year of wage record history prior to the quarter in which an accession occurs. If there are no valid wage records for this job (PIK-SEIN) during the four quarters preceding an accession, then the accession is called a new hire; otherwise, the accession is called a recall. Thus, new hires and recalls sum to accessions. For example, suppose that an individual accedes to employer *A* in 1999:3. Recall that this means that there is a valid UI wage record for the individual 1 at employer *A* in 1999:3 but not in 1999:2. If there are also no valid UI wage records for individual 1 at employer *A* for 1999:1, 1998:4 and 1998:3, then the EDE system designates this accession as a new hire of individual 1 by employer *A* in 1999:3. Consider a second example in which individual 2 accedes to employer *B* in 2000:2. Once again, the accession implies that there is not a valid wage record for individual 2 at employer *B* in 2000:1. If there is a valid wage record for individual 2 at employer *B* in 1999:4, 1999:3, or 1999:2, then the EDE system designates the accession of individual 2 to employer *B* as a recall in 2000:2. New hire and recall data, because they depend upon having four quarters of historical data, only become available one year after the data required to estimate accessions become available.

2.11 New hires and recalls to and from full-quarter employment

Accessions to full-quarter status can also be decomposed into new hires and recalls. The EDE system accomplishes this decomposition by classifying all accession to full-quarter status who were classified as new hires in the previous quarter as new hires to full-quarter status in the current quarter. Otherwise, the accession to full-quarter status is classified as a recall to full-quarter status. For example, if individual 1 accedes to full-quarter status at employer *A* in 1999:4 then, according to the definitions above, individual 1 acceded to employer *A* in 1999:3 and reached full-quarter status in 1999:4. Suppose that the accession to employer *A* in 1999:3 was classified as a new hire, then the accession to full quarter status in 1999:4 is classified as a full-quarter new hire. For another example, consider individual 2 who accedes to full-quarter status at employer *B* in 2000:3. Suppose that the accession of individual 2 to employer *B* in 2000:2, which is implied by the full-quarter accession in 2000:3, was classified by the EDE system as a recall in 2000:2; then, the accession of individual 2 to full-quarter status at employer *B* in 2000:3 is classified as a recall to full-quarter status.

2.12 Job creations and destructions

Job creations and destructions are defined at the employer (SEIN) level and not at the job (PIK-SEIN) level. To construct an estimate of job creations and destructions, the EDE system totals beginning and ending employment for each quarter for every employer in the UI wage record universe, that is, for an employer who has at least one valid UI wage record during the quarter. The EDE system actually uses the Davis et al. (1996) formulas for job creation and destruction (see definitions in Appendix A on page 19). Here, we use a simplified definition. If end-of-quarter employment is greater than beginning-of-quarter employment, then the employer has created jobs. The EDE system sets job creations in this case equal to end-of-quarter employment less beginning-of-quarter employment. The estimate

of job destructions in this case is zero. On the other hand, if beginning-of-quarter employment exceeds end-of-quarter employment, then this employer has destroyed jobs. The EDE system computes job destructions in this case as beginning-of-period employment less end-of-period employment. The EDE system sets job creations to zero in this case. Notice that either job creations are positive or job destructions are postive, but not both. Job creations and job destructions can simultaneously be zero if beginning-of-quarter employment equals end-of-quarter employment. There is an important suptelty regarding job creations and destructions when they are computed for different sex and age groups within the same employer. There can be creation and destruction of jobs for certain demographic groups within the employer without job creation or job destruction occuring overall. That is, jobs can be created for some demographic groups and destroyed for others even at enterprises that have no change in employment as a whole.

Here is a simple example. Suppose employer *A* has 250 employees at the beginning of 2000:3 and 280 employees at the end of 2000:3. Then, employer *A* has 30 job creations and zero job destructions in 2000:3. Now suppose that of the 250 employees 100 are men and 150 are women at the beginning of 2000:3. At the end of the quarter suppose that there are 135 men and 145 women. Then, job creations for men are 35 and job destructions for men are 0 in 2000:3. For women in 2000:3 job creations are 0 and job destructions are 5. Notice that the sum of job creations for the employer by sex (35 + 0) is not equal to job creations for the employer as a whole (30) and that the sum of job destructions by sex (0 + 5) is not equal to job destructions for the employer as a whole.

2.13 Net job flows

Net job flows are also only defined at the level of an employer (SEIN). They are the difference between job creations and job destructions. Net job flows are, thus, always equal to end-of-quarter employment less beginning of quarter employment.

Returning to the example in the description of job creations and destructions. Employer *A* has 250 employees at the beginning of 2000:3 and 280 employees at the end of 2000:3. Net job flows are 30 (job creations less job destructions or beginning-of-quarter employment less end-of-quarter employment). Suppose, once again that employment of men goes from 100 to 135 from the beginning to the end of 2000:3 and employment of women goes from 150 to 145. Notice, now, that net job flows for men (35) plus net job flows for women (-5) equals net job flows for the employer as a whole (30). Net job flows are additive across demographic groups even though gross job flows (creations and destructions) are not.

Some useful relations among the worker and job flows include:

- Net job flows = job creations - job destructions
- Net job flows = end-of-quarter employment - beginning-of-period employment
- Net job flows = accessions - separations

These relations hold for every demographic group and for the employer as a whole. Additional identities are shown in Appendix A.

2.14 Full-quarter job creations, job destructions and net job flows

The EDE system applies the same job flow concepts to full-quarter employment to generate estimates of full-quarter job creations, full-quarter job destructions, and full-quarter net job flows. Full-quarter employment in the current quarter is compared to full-quarter employment in the preceding quarter. If full-quarter employment has increased between the preceding quarter and the current quarter, then full-quarter job creations are equal to full-quarter employment in the current quarter less full-quarter employment in the preceding quarter. In this case full-quarter job destructions are zero. If full-quarter employment has decreased between the previous and current quarters, then full-quarter job destructions are equal to full-quarter employment in the preceding quarter minus full-quarter employment in the current quarter. In this case, full-quarter job destructions are zero. Full-quarter net job flows equal full-quarter job creations minus full-quarter job destructions. The same identities that hold for the regular job flow concepts hold for the full-quarter concepts.

2.15 Average earnings of end-of-period employees

The average earnings of end-of-period employees is estimated by first totaling the UI wage records for all individuals who are end-of-period employees at a given employer in a given quarter. Then the total is divided by the number of end-of-period employees for that employer and quarter.

2.16 Average earnings of full-quarter employees

Measuring earnings using UI wage records in the EDE system presents some interesting challenges. The earnings of end-of-quarter employees who are not present at the beginning of the quarter are the earnings of accessions during the quarter. The EDE system does not provide any information about how much of the quarter such individuals worked. The range of possibilities goes from 1 day to every day of the quarter. Hence, estimates of the average earnings of such individuals may not be comparable from quarter to quarter unless one assumes that the average accession works the same number of quarters regardless of other conditions in the economy. Similarly, the earnings of beginning-of-quarter who are not present at the end of the quarter represent the earnings of separations. These present the same comparison problems as the average earnings of accessions; namely, it is difficult to model the number of weeks worked during the quarter. If we consider only those individuals employed at the firm in a given quarter who were neither accessions nor separations during that quarter, we are left, exactly, with the full-quarter employees, as discussed above.

The EDE system measures the average earnings of full-quarter employees by summing the earnings on the UI wage records of all individuals at a given employer who have full-quarter status in a given quarter then dividing by the number of full-quarter employees. For example, suppose that in 2000:2 employer *A* has 10 full-quarter employees and that their total earnings are $\$300,000$. Then, the average earnings of the full-quarter employees at *A* in 2000:2 is $\$30,000$. Suppose, further that 6 of these employees are men and that their total earnings are $\$150,000$. So, the average earnings of full-quarter male employees is $\$25,000$ in 2000:2 and the average earnings of female full-quarter employees is $\$37,500$ $(= \$150,000/4)$.

2.17 Average earnings of full-quarter accessions

As discussed above, a full-quarter accession is an individual who acceded in the preceding quarter and acheived full-quarter status in the current quarter. The EDE system measures the average earnings of full-quarter accessions in a given quarter by summing the UI wage record earnings of all full-quarter accessions during the quarter and dividing by the number of full-quarter accessions in that quarter.

2.18 Average earnings of full-quarter new hires

Full-quarter new hires are accessions to full-quarter status who were also new hires in the preceding quarter. The average earnings of full-quarter new hires are measured as the sum of UI wage records for a given employer for all full-quarter new hires in a given quarter divided by the number of full-quarter new hires in that quarter.

2.19 Average earnings of full-quarter separations

Full-quarter separations are individuals who separate during the current quarter who were full-quarter employees in the previous quarter. The EDE system measures the average earnings of full-quarter separations by summing the earnings for all individuals who are full-quarter status in the current quarter and who separate in the subsequent quarter. This total is then divided by full-quarter separations in the subsequent quarter. The average earnings of full-quarter separations is, thus, the average earnings of full-quarter employees in the current quarter who separated in the next quarter. Note the dating of this variable.

2.20 Average periods of non-employment for accessions, new hires, and recalls

As noted above an accession occurs when a job starts; that is, on the first occurance of an SEIN-PIK pair following the first quarter of available data. When the EDE system detects an accession, it measures the number of quarters (up to a maximum of four) that the individual spent non-employed in the state prior to the accession. The EDE system estimates the number of quarters spent non-employed by looking for all other jobs held by the individual at

any employer in the state in the preceding quarters up to a maximum of four. If the EDE system doesn't find any other valid UI-wage records in a quarter preceding the accession it augments the count of non-employed quarters for the individual who acceded, up to a maximum of four. Total quarters of non-employment for all accessions is divided by accessions to estimate average periods of non-employment for accessions.

Here is a detailed example. Suppose individual 1 and individual 2 accede to employer *A* in 2000:1. In 1999:4, individual *A* does not work for any other employers in the state. In 1999:1 through 1999:3 individual 1 worked for employer *B*. Individual 1 had one quarter of non-employment preceding the accession to employer *A* in 2000:1. Individual 2 has no valid UI wage records for 1999:1 through 1999:4. Indivdiual 2 has four quarters of non-employment preceding the accession to employer *A* in 2000:1. The accessions to employer *A* in 2000:1 had an average of 2.5 quarters of non-employment in the state prior to accession.

Average periods of non-employment for new hires and recalls are estimated using exactly analogous formulas except that the measures are estimated separately for accesions who are also new hires as compared with accession who are recalls.

2.21 Average number of periods of non-employment for separations

Analogous to the average number of periods of non-employment for accessions prior to the accession, the EDE system measures the average number of periods of non-employment in the state for individuals who separated in the current quarter, up to a maximum of four. When the EDE system detects a separation, it looks forward for up to four quarters to find valid UI wage records for the individual who separated and other employers in the state. Each quarter that it fails to detect any such jobs is counted as a period of non-employment, up to a maximum of four. The average number of periods of non-employment is estimated by dividing the total number of periods of non-employment for separations in the current quarter by the number of separations in the quarter.

2.22 Average changes in total earnings for accessions and separations

The EDE system measures the change in total earnings for individuals who accede or separate in a given quarter. For an individual accession in a given quarter, the EDE system computes total earnings from all valid wage records for all of the individual's employers in the preceding quarter. The system then computes the total earnings for the same individual for all valid wage records and all employers in the current quarter. The acceding individual's change in earnings is the difference between the current quarter earnings from all employers and the preceding quarter earnings from all employers. The average change in earnings for all accessions is the total change in earnings for all accesions divided by the number of accessions.

The EDE system computes the average change in earnings for separations in an analogous manner. The system computes total earnings from all employers for the separating indivdiual in the current quarter and subtracts total earnings from all employers in the subsequent quarter. The average change in earnings for all separations is the total change in earnings for all separations divided by the number of separations.

Here is an example for the average change in earnings of accessions. Suppose individual 1 accedes to employer *A* in 2000:3. Earnings for individual 1 at employer *A* in 2000:3 are $\$8,000$. Individual 1 also worked for employer *B* in 2000:2 and 2000:3. Individual 1's earnings at employer *B* were $\$7,000$ and $\$3,000$ in in 2000:2 and 2000:3, respectively. Individual 1's change in total earnings between 2000:3 and 2000:2 was $\$4,000\,(=\$8,000+\$3,000-\$7,000)$. Individual 2 also acceded to employer *A* in 2000:3. Individual 2 earned $\$9,000$ from employer *A* in 2000:3. Individual 2 had no other employers during 2000:2 or 2000:3. Individual 2's change in total earnings is $\$9,000$. The average change in earnings for all of employer *A*'s accessions is $\$6,500\,(=(\$4,000+\$9,000)/2)$, the average change in total earnings for individuals 1 and 2.

3 Input files for the EDE

The EDE (and all other LEHD statistics) are constructed from a limited number of standardized files. The underlying data, as mentioned above, are extracted from UI administrative files from each participating state, as well as from the (typically independently created) files from the ES-202 system. Further information comes from Census-internal files, such as the Census Personal Characteristics File (PCF). Other statistical products, though not the EDE, may also use data from the Census's Longitudinal Research Database.

Once the longitudinal identification has been improved, data from the original adminstrative files as well as internal Census files is parsed, and saved in the form of standardized files. For each entity level (individual or firm), two files are constructed, one containing time-aggregated and time-invariant information, the other containing detailed time-varying information.[1]

Table 1 provides an overview of these files.

Table 1: Files used for EDE

Input data	*Dimension*
UI wage data	(quarterly, active individuals)
ES-202 reports	(quarterly, active firms)
Census PCF	(no. of individuals)

Base data	*Dimension*
ICF	(no. of individuals)
EHF	(no. of individuals *x* years present in data)
ECF	(no. of firms)

3.1 Longitudinal consistency of identifiers

Both the UI and the ES-202 data files in the state administrations are built up from a mix of paper and electronic records. Firms in the ES-202 system are identified by a (UI tax) account number attributed by the state. As with all firm identities, an account number can change for a number of reasons over time, not all of which are distinguishable economic entities for the purpose of these statistics. States take great care to follow the legal entities in their system, but account numbers may nevertheless change for reasons which economists may not consider legitimate for the purposes of the EDE. On the other hand, the UI data concern a large number of individuals with a legally unique identifier, the Social Security Number, and individuals are normally not allowed to change their Social Security Number. However, coding errors are less well identified by the UI administration, and are thus more prevalent than for the firm identifiers. In both cases, the longitudinal integrity of the identifier is compromised, creating spurious job creation and job destruction, and biasing all other point-in-time estimates as well.

At LEHD, each problem has been addressed by a specially designed process. SSN coding errors are addressed by a process which reattaches miscoded records to the relevant within-firm work history of a person, relying on name and earnings information (see Vilhuber (2001, 2002a,b) for examples). The identification of a firm's economic successor is achieved by identification of large inter-firm employment transfers, which is only possible through the combination of ES-202 and UI records(see McKinney (2002) for an overview).

4 Component files for the EDE process

4.1 Individual Characteristics File

The *Individual Characteristics File* (ICF)for each state contains one record for every person who is ever employed in that state over the time period spanned by the state's unemployment insurance records. The ICF is created in several steps. The information in the UI wage records is processed first. The resulting data are then merged with data on each

[1] The firm-level time-varying information is currently being developed.

individual from the Census PCF file. The data are then linked to the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP). (A link to the American Community Survey (ACS) will be implemented in a future release.) Finally age and sex are imputed for those people who do not match an identifier in the PCF file. Then, the ICF goes through final processing. Each step is briefly described below.

Initial processing of the UI wage records

The initial input data set for the ICF is the Employment History File. This data file contains PIKs, wages, year, quarter, and state firm identifiers (SEINs). The input records are in the format job/year/quarter, resulting in multiple records per PIK. The initial processing compiles a list of unique PIKs and counts the number of unique employers for each PIK in each year and year/quarter. PIKs which never have positive earnings are dropped.

Merge data from the Census PCF file

The PCF contains information about gender, race, citizenship status, place of birth, date of birth, and date of death for approximately 97% of the PIKs in the UI wage data. This information is merged onto the unique list of PIKs created in the first step.

Create links to the CPS and SIPP

Using PIK/survey identifier crosswalks created by Census, the CPS and SIPP ID variables are merged onto the unique list of PIKs from Step Two. These identification variables enable one to link individuals to the 1983-1997 CPS surveys and the 1984, 1990-1993, and 1996 SIPP panels.

Age and sex imputation and final processing

Since approximately 3% of the PIKs found in the UI wage records do not match to the PCF file, multiple imputation methods are used to assign age and gender to these individuals. The gender imputation is done using a logit model to predict the probability of being male and then assigning gender based on this probability. The age imputation is done using a multinomial logit to predict the probability of being in a given age category and then assigning an age based on this probability and the distribution of ages within the category. The final processing creates the actual ICF data set, which conforms to all LEHD variable and data set specifications.

4.2 Employment History File

The *Employment History File* (EHF) is designed to store the complete in-state work history for each individual that appears in the UI wage records. The EHF for each state contains one record for each employee-employer combination in that state in each year. Every individual who is employed during a given year will then have one observation per employer for that year. The data in the EHF are stored in the form of annual records mainly to save space. Annual earnings have been calculated for future use, but the quarterly earnings variables are also retained in the data. The current job flow analysis focuses entirely on quarterly earnings and quarterly employment flows. For the purposes of the worker and job flow analysis, the only necessary variables are the person and employer identifiers and the quarterly earnings variables. The presence of positive quarterly earnings is used in the job flow analysis not only to compute earnings and payroll statistics but also to determine an individual's employment status each quarter. This information is used for calculating flows of workers into and out of jobs. An individual is considered to be employed by a particular employer during a quarter if (1) a record is present in the raw data for that particular PIK/SEIN combination and (2) the wage field for this record contains a positive (non-zero) value. Together, the quarterly employment status and the quarterly earnings variable provide sufficient information to compute all of the other statistics used in the employment flow analysis.

4.3 Employer Characteristics File

The *Employer Characteristics File* (ECF)is based on information from each state's Department of Employment Security. The data are collected as part of the Covered Employment and Wages (CEW) program, also known as the ES-202

program, which is administered by the U.S. Bureau of Labor Statistics (BLS). This cooperative program between the states and the federal government collects employment, payroll, and location information from employers covered by state unemployment insurance programs.

In order to calculate the job flow statistics by county and Standard Industry Classification (SIC) additional data processing is necessary. The original ES-202 data as received from the states are establishment-based. However the individual wage and employment data from the unemployment insurance program do not contain sufficient information to identify the place of work for employees of multi-unit firms. A multi-unit firm is likely to have establishments in more than one county and/or SIC and a method must be chosen to assign a firm to one county and SIC. At LEHD, the firm is assigned the employment-weighted mode (most prevalent) county and SIC of its establishments.

As a final step missing values of county and SIC are filled in using non-missing data in adjacent year quarters. If a firm is likely to change industries or county relatively infrequently, then this is a reasonable procedure, but it may understate movement for firms that have missing data and relocate often.

The result of the data processing described above is a file with information on county, SIC, and the number of employees for every SEIN that has at least one employee with positive earnings in a given year-quarter.

xxx Talk about files here: inhale the short descriptions from the EDE methods file. xxx

5 Forming Aggregated Estimates

Aggregating the EDE data is a four step process, which can be summarized as follows:

1. The basic variables, as discussed above, are created for each employment history (PIK-SEIN pair) and for every quarter that the pair exits.
2. The EDE system sums for each employer the following variables: beginning-of-period employment, end-of-period employment, accessions, new hires, recalls, separations, full-quarter employment, full-quarter accessions, full-quarter new hires, total earnings of full-quarter employees, total earnings of full-quarter accessions, and total earnings of full-quarter new hires. Job creations, job destructions, and net job flows are estimated for each employer using the beginning and end of quarter employment estimates for that employer. The first-layer of disclosure-proofing is also applied at this step.
3. The employer-level variables in the list above are summed over the relevant aggregating unit (county or SIC division) for each quarter. Average earnings of full-quarter employees, full-quarter accessions, and full-quarter new hires are estimated by taking the ratio of total earnings of the relevant category to the total number of individuals in that category. For example, aveargе earnings of full-quarter men ages 55-64 for a given year, quarter and county is the ratio of total earnings of full-quarter men ages 55-64 to the number of full-quarter men ages 55-64 in that year, quarter, and county.
4. The beginning-of-quarter employment for each county or SIC division is controlled (raked) to the BLS estimate of total county employment in month one of that quarter from the Covered Employment and Wages series. At this point the other estimates and the demographic groups are also raked to preserve the underlying relations among the variables.

The detailed description that follows makes explicit the links to the sets of files described earlier. The statistics calculated in this section are based on definitions summarized in Abowd et al. (1999) and Davis and Haltiwanger (1999). As mentioned before, employment is measured at two points in time (beginning and end of quarter) and according to two concepts (any employment status and full-quarter employment status). Worker flows are captured by accessions and separations with respect to both employment status concepts. Job flows are captured by gross job creation and destruction at the firm level, again according to both employment concepts. Accessions are further separated into new hires and recalls. Earnings and earnings change statistics are calculated for each of the worker flow categories as well as for both employment statuses.

The worker and employment flow statistics reported at the county and SIC division level are calculated through a multi-step process.The EHF (see Subsection 4.2), which contains individual work and earnings histories, is combined with information from the ICF (see Subsection 4.1) to incorporate demographic characteristics of workers such as age and sex. For each worker in each year and quarter, an array of jobs at various SEINs is stored. The statistics listed in Subsection A on page 19 are computed, when appropriate, for each individual/job/quarter combination. The statistics are then aggregated to the SEIN level by age and sex to create a file of totals for each

SEIN/year/quarter/agegroup/sexgroup combination. Both the Workforce Investment Act (WIA) and CPS age groups are used. The totals are stored by age/sex group as well as further aggregated within SEIN over age and sex group to produce the overall total for the SEIN as well as marginal totals for sex and age (for example, the total for females of all ages). All totals are then aggregated twice more: once to the industry level and once to the county level. At this point the statistics are in their final form except for the handling of disclosure issues, as discussed below.

5.1 Examples

The following tables provide an example of how the flow statistics are computed for four hypothetical individuals who work at three hypothetical employers over a two year sample period. All individuals and firms in this example are fictitious. Table 2 summarizes the earnings history of each individual as it would appear in the employment history file. Table 3 on the following page presents the individual level employment flow statistics that can be computed from the individual work histories. Note that individual 1 leaves employer *X* at some point during the second quarter of 1995, and that she begins working for employer *Y* during the same quarter. In Table 3, employment flow statistics as defined in Subsection A have been computed for every quarter of every job worked by Person *1*. Person *1* is considered to be employed at employer *X* from 1994:1 – 1995:2. Hence, *e=1* from 1994:1 through 1995:1 since she is still employed at *X* at the end of each of these quarters. Similarly, *b=1* from 1994:2 through 1995:2 since she is employed at *X* from the very beginning of these quarters. Note that *b* is missing in 1994:1. The first quarter of the analysis is out-of-scope for *b*, since it depends on employment information from the previous quarter. Also note that for in-scope periods, end-of-quarter employment at time *t* is equal to beginning-of quarter employment at time *t + 1*. In Subsection A, this identity (Identity 1) is defined for aggregates, but as shown in the example it holds at the individual level as well.

			Earnings			
PIK	SEIN	Year	Q1	Q2	Q3	Q4
1	X	1994	4500	4500	4800	4800
1	X	1995	5000	3500		
1	Y	1995		2000	6500	6900
2	Y	1994	1800	1800	1800	1800
2	Y	1995	2000			
2	Z	1995			2500	3000
3	Z	1994	5500	5500	5500	5500
3	Z	1995	6000	6000	6000	6000
4	X	1994	3700	3700	3800	3800
4	X	1995	4000		4200	4300

Table 2: Example of individual earnings histories, EHF

Moving on, *f=1* for Person *1*/Employer *X* from 1994:2-1995:1, but *f* is missing during 1994:1, which is out-of-scope, and *f=0* during 1995:2 because she is no longer employed at *X* in 1995:3. In 1995:2 *s=1* and *fs=1* for Person *1*/ Employer *X* because she separates from Employer *X* sometime during this quarter and appears to have been in this job for the entire preceding quarter (1995:1). In 1995:2, *a=1* for Person *1* and Employer *Y* because she enters a relationship with Employer *Y* sometime during this quarter, and *fa=1* in 1995:3 because this is her first full quarter at Employer *Y*. New hires, *h*, is also 1 because she has no previous relationship with Employer *Y* in the last four quarters, and recalls *r=0* for the same reason. A variety of wage measures are also calculated for each individual: *w1* is simply the wage earned at each job each quarter, while measures such as *w2*, *w3*, *wa* are calculated as an individual's wage if he or she meets a certain criteria (*e=1* for *w2*, *f=1* for *w3*, *etc*.).

In Table 4 on page 13, the individual statistics are aggregated to the employer level by summing individual statistics by SEIN. *E* for Employer *X* in 95:1, then, is the sum of *e* over all individuals working at *X* in 1995:1 (in this case individuals 1 and 4). Since *e=1* for Individual *1*, who remains with Employer *X* the next quarter, and *e=0* for Individual *4*,

Table 3: Employment flow statistics at the individual level

PIK	SEIN	YR:QTR	b	e	f	a	h	r	s	fa	fs	$w1$	$w2$	$w3$	wa	dwa	na	nh	nr	ws	dws	ns
1	X	94:1	.	1	.	.	.	.	0	.	.	4500	4500	.	.	.	.	.	.	.	.	.
1	X	94:2	1	1	1	0	.	.	0	.	.	4500	4500	4500	.	.	.	.	.	.	.	.
1	X	94:3	1	1	1	0	.	.	0	0	0	4800	4800	4800	.	.	.	.	.	.	.	.
1	X	94:4	1	1	1	0	.	.	0	0	0	4800	4800	4800	.	.	.	.	.	.	.	.
1	X	95:1	1	1	1	0	0	0	0	0	0	5000	5000	5000	.	.	.	.	.	.	.	.
1	X	95:2	1	0	0	0	0	0	1	0	1	3500	.	.	.	.	.	.	.	3500	−1500	0
1	Y	95:2	0	1	0	1	1	0	0	0	0	2000	2000	.	2000	−1500	0	0	.	.	.	.
1	Y	95:3	1	1	1	0	0	0	0	1	0	6500	6500	6500	.	.	.	.	.	.	.	.
1	Y	95:4	1	.	.	0	0	0	.	.	.	6900	.	.	.	.	.	.	.	.	.	.
2	Y	94:1	.	1	.	.	.	.	0	.	.	1800	1800	.	.	.	.	.	.	.	.	.
2	Y	94:2	1	1	1	0	.	.	0	.	.	1800	1800	1800	.	.	.	.	.	.	.	.
2	Y	94:3	1	1	1	0	.	.	0	0	0	1800	1800	1800	.	.	.	.	.	.	.	.
2	Y	94:4	1	1	1	0	.	.	0	0	0	1800	1800	1800	.	.	.	.	.	.	.	.
2	Y	95:1	1	0	0	0	0	0	1	0	1	2000	0	.	.	.	.	.	.	2000	200	1
2	Z	95:3	0	1	0	1	1	0	0	0	0	2500	2500	.	2500	500	1	1	.	.	.	.
2	Z	95:4	1	.	.	0	0	0	.	.	.	3000	.	.	.	.	.	.	.	.	.	.
3	Z	94:1	.	1	.	.	.	.	0	.	.	5500	5500	.	.	.	.	.	.	.	.	.
3	Z	94:2	1	1	1	0	.	.	0	.	.	5500	5500	5500	.	.	.	.	.	.	.	.
3	Z	94:3	1	1	1	0	.	.	0	0	0	5500	5500	5500	.	.	.	.	.	.	.	.
3	Z	94:4	1	1	1	0	.	.	0	0	0	5500	5500	5500	.	.	.	.	.	.	.	.
3	Z	95:1	1	1	1	0	0	0	0	0	0	6000	6000	6000	.	.	.	.	.	.	.	.
3	Z	95:2	1	1	1	0	0	0	0	0	0	6000	6000	6000	.	.	.	.	.	.	.	.
3	Z	95:3	1	1	1	0	0	0	0	0	0	6000	6000	6000	.	.	.	.	.	.	.	.
3	Z	95:4	1	.	.	0	0	0	.	.	.	6000	.	.	.	.	.	.	.	.	.	.
4	X	94:1	.	1	.	.	.	.	0	.	.	3700	3700	.	.	.	.	.	.	.	.	.
4	X	94:2	1	1	1	0	.	.	0	.	.	3700	3700	3700	.	.	.	.	.	.	.	.
4	X	94:3	1	1	1	0	.	.	0	0	0	3800	3800	3800	.	.	.	.	.	.	.	.
4	X	94:4	1	1	1	0	.	.	0	0	0	3800	3800	3800	.	.	.	.	.	.	.	.
4	X	95:1	1	0	0	0	0	0	1	0	1	4000	0	.	.	.	.	.	.	4000	200	1
4	X	95:3	0	1	0	1	0	1	0	0	0	4200	4200	.	4200	200	1	.	1	.	.	.
4	X	95:4	1	.	.	0	0	0	.	.	.	4300	.	.	.	.	.	.	.	.	.	.

who has no wage record with Employer *X* the next quarter, *E=1*. Similarly, since *a=0* for both individuals this quarter (both worked at *X* last quarter also), *A=0*. The job flow at Employer *X*, defined as the net increase in employment over that quarter, is calculated as the difference between the number of end-of-quarter jobs held and the number of beginning-of quarter jobs held. Thus, $JF = E - -B$, or 1 – 2 = -1 in this case. Because there was a negative net job flow of 1 this quarter, job creation *JC= 0* and job destruction *JD=1*. Total payroll *W1* is also computed for each employer; for Employer *X* in 1995:1 it is simply the sum of the wages paid to individuals 1 and 4: $5000 + $4000 = $9000. Individual 1 also had end-of-quarter wages *w2=5000* because she was end-of-quarter employed at *X* this period. For Individual *4*, *w2=0* because *e=0* at *X* in 1995:1. Total end of quarter wages *W2* for Employer *X* in 1995:1 is then calculated as the sum of wages at all end-of-quarter jobs. In this case, it is simply $5000 since Individual 1 has the only end-of-quarter job at *X* in 1995:1.

Table 4: Employment flow statistics at the SEIN level

SEIN	YR:QTR	B	E	F	A	H	R	S	FA	FS	JF	JC	JD	FJF	FJC	FJD	$W1$	$W2$	$W3$	WA	dWA	NA	NH	NR	WS	dWS	NS
X	94:1	.	2	.	.	.	.	0	.	.	.	.	.	.	.	.	8200	8200	.	.	.	.	.	.	.	.	.
X	94:2	2	2	2	0	.	.	0	.	.	0	0	0	.	.	.	8200	8200	8200	.	.	.	.	.	.	.	.
X	94:3	2	2	2	0	.	.	0	0	0	0	0	0	0	0	0	8600	8600	8600	.	.	.	.	.	.	.	.
X	94:4	2	2	2	0	.	.	0	0	0	0	0	0	0	0	0	8600	8600	8600	.	.	.	.	.	.	.	.
X	95:1	2	1	1	0	0	0	1	0	1	−1	0	1	−1	0	1	9000	5000	5000	.	.	.	.	.	4000	200	1
X	95:2	1	0	0	0	0	0	1	0	1	−1	0	1	−1	0	1	3500	.	.	.	.	.	.	.	3500	−1500	0
X	95:3	0	1	0	1	0	1	0	0	0	1	1	0	0	0	0	4200	4200	.	4200	200	1	.	1	.	.	.
X	95:4	1	.	.	0	0	0	.	.	.	.	.	.	.	.	.	4300	.	.	.	.	0	.	.	.	.	.
Y	94:1	.	1	.	.	.	.	0	.	.	.	.	.	.	.	.	1800	1800	.	.	.	.	.	.	.	.	.
Y	94:2	1	1	1	0	.	.	0	.	.	0	0	0	.	.	.	1800	1800	1800	.	.	.	.	.	.	.	.
Y	94:3	1	1	1	0	.	.	0	0	0	0	0	0	0	0	0	1800	1800	1800	.	.	.	.	.	.	.	.
Y	94:4	1	1	1	0	.	.	0	0	0	0	0	0	0	0	0	1800	1800	1800	.	.	.	.	.	.	.	.
Y	95:1	1	0	0	0	0	0	1	0	1	−1	0	1	−1	0	1	2000	0	.	.	.	.	.	.	2000	200	1
Y	95:2	0	1	0	1	1	0	0	0	0	1	1	0	0	0	0	2000	2000	.	2000	−1500	0	0	.	.	.	.
Y	95:3	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	6500	6500	6500	.	.	.	.	.	.	.	.
Y	95:4	1	.	.	0	0	0	.	.	.	.	.	.	.	.	.	6900	.	.	.	.	.	.	.	.	.	.
Z	94:1	.	1	.	.	.	.	0	.	.	.	.	.	.	.	.	5500	5500	.	.	.	.	.	.	.	.	.
Z	94:2	1	1	1	0	.	.	0	.	.	0	0	0	.	.	.	5500	5500	5500	.	.	.	.	.	.	.	.
Z	94:3	1	1	1	0	.	.	0	0	0	0	0	0	0	0	0	5500	5500	5500	.	.	.	.	.	.	.	.
Z	94:4	1	1	1	0	.	.	0	0	0	0	0	0	0	0	0	5500	5500	5500	.	.	.	.	.	.	.	.
Z	95:1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	6000	6000	6000	.	.	.	.	.	.	.	.
Z	95:2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	6000	6000	6000	.	.	.	.	.	.	.	.
Z	95:3	1	2	1	1	1	0	0	0	0	1	1	0	0	0	0	8500	8500	6000	2500	500	1	1	.	.	.	.
Z	95:4	2	.	.	0	0	0	.	.	.	.	.	.	.	.	.	9000	.	.	.	.	.	.	.	.	.	.

Several identities from Subsection A are illustrated in Table 4. Once again Identity 1 ($B_{jt} = E_{jt-1}$) is noticeable just from glancing at the columns of numbers B and E. Identity 3, $E_{jt} = B_{jt} + A_{jt} - S_{jt}$ also holds whenever all four variables are in-scope. For example, for Employer *X* in 1995:1, *E = 1 = 2 + 0 – 1*. For this employer in 1995:2, *E = 0 = 1 + 0 – 1*. Identity 4, $JF_{jt} = JC_{jt} - JD_{jt}$ is also true: for *X* in 1995: 1 *JF = -1 = 0 – 1*. Identity 5, $E_{jt} = B_{jt} + JC_{jt} - JD_{jt}$, (*X* in 1995:1 : *E = 1 = 2 + 0 – 1*) and Identity 6, $A_{jt} - S_{jt} = JC_{jt} - JD_{jt}$, (*X* in 1995: 1 : *A – S = 0 – 1 = 0 – 1 = JC – JD*). Finally, Identity 15, the total payroll identity ($W_{1jt} = W_{2jt} + WS_{jt}$) is met in all cases. For example, for SEIN *X* in 1995: 2, *W1 = $9000 = 5000 + 4000 = W2 + WS*. When *WS is missing*, as in most cases, *W1* and *W2* are simply equal because every wage is an end-of-quarter wage.

Table 5: Employment flow statistics aggregated to form county totals

YR:QTR	B	E	F	A	H	R	S	FA	FS	JF	JC	JD	FJF	FJC	FJD	$W1$	Z_W2	Z_W3	Z_WA	Z_dWA	Z_NA	Z_NH	Z_NR	Z_dWS	Z_NS
94:1	0	4	0	0	0	0	0	0	0	.	.	.	.	.	.	15500	3875	.	.	.	.	.	.	.	.
94:2	4	4	4	0	0	0	0	0	0	0	0	0	.	.	.	15500	3875	3875	.	.	.	.	.	.	.
94:3	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	15900	3975	3975	.	.	.	.	.	.	.
94:4	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	15900	3975	3975	.	.	.	.	.	.	.
95:1	4	2	2	0	0	0	2	0	2	−2	0	2	−2	0	2	17000	5500	5500	.	.	.	.	.	3000	200
95:2	2	2	1	1	1	0	1	0	1	0	1	1	−1	0	1	11500	4000	6000	2000	−1500	0	0	.	3500	−1500
95:3	2	4	2	2	1	1	0	1	0	2	2	0	1	1	0	19200	4800	6250	3350	350	1	1	1	.	.
95:4	4	0	0	0	0	0	0	0	0	.	.	.	.	.	.	20200	.	.	.	.	.	.	.	.	.

In Table 5, the SEIN-level statistics are aggregated in a similar way to create total flows and average wages. These can be thought of as county totals if the hypothetical universe includes just a single county. The total flows are computed exactly as the employer level flows in Table 4. For 1995:1, total jobs at the end of quarter, total *E*, is just the sum of *E* for *X*, *Y*, and *Z*: 1 + 0 + 1 = 2. Note that this is the same as the sum of all individual *e* in Table 3 for 1995:1. Total accessions are computed similarly (*A = 0 + 0 + 0*) as are total wages (*W1 = 9000 + 2000 + 6000 = 17000*). Average wages (for example, *Z_W2*, *Z_WA*) are computed by summing total wages for *X*, *Y*, and *Z*, and dividing by the total number of individuals used to calculate the particular wage measure. For example, *Z_W2* for 1995: 1 is computed as the sum of *W2* for *X*, *Y*, and *Z* where defined (5000 + 6000) divided by the total number of end-of-quarter positions (*E = 2*) for an average end-of-quarter wage of *Z_W2* = $5500. *Z_WA* is undefined for this quarter because there are no accessions this quarter. In 1995:2 they are computed as the sum of *WA* for *X*, *Y*, and *Z* where defined (2000, since *WA* is only defined for *Y* this quarter) divided by the total number of accessions this quarter (1) so the average wage to accessions in 1995:2 is simply $2000.

6 Disclosure Proofing the EDE

Disclosure proofing is the set of methods used by statistical agencies to protect the confidentiality of the identity of and information about the individuals and businesses that form the underlying data in the system. In the EDE system, disclosure proofing is required to protect the information about individuals and businesses that contribute to the UI wage records, the ES-202 quarterly reports, and the Census Bureau demographic data that have been integrated with these sources. There are three layers of and disclosure proofing in the EDE system.

The first layer occurs at stage two in the production of the estimates, the stage at which employer-level estimates are made. At this stage, the EDE system infuses specially constructed noise into the estimates of all of the employer-level measures. This noise is designed to have two very important properties. First, for a given employer, the data are always distorted in the same direction (increased or decreased) by the same percentage amount in every period. Second, the statistical properties of this distortion are such that when the estimates are aggregated to the county or SIC division level the effects of the distortion cancel out for the vast majority of the estimates.

The second layer of confidentiality protection occurs when the employer-level measures are aggregated to the county or SIC division level. The data from many individuals and businesses are combined into a (relatively) few estimates. This aggregation helps to conceal the exact information about any of the individuals or businesses that underlie the estimate. At this level of confidentiality protection, some of the estimates turn out to be based on fewer than three persons. These estimates are suppressed. In addition, some of the estimates are based on data that are still substantially influenced by the noise that was infused in the first layer. These estimates are flagged as substantially distorted.

The final layer of confidentiality protection occurs when the beginning-of-quarter employment estimate for the county or SIC division as a whole is raked to the published BLS estimates from the Covered Employment and Wages series. This raking has two effects. First, aggregates that the BLS suppresses (because of conditions in its disclosure-

proofing system) are also suppressed by the EDE system. Second, the EDE system does not produce an independent estimate of overall employment for the aggregate. The EDE system, thus supplements the Covered Employment and Wages program by providing worker flows, job flows, full-quarter employment estimates, and demographic detail, none of which can be easily estimated from the ES-202 quarterly reports themselves. The final layer of confidentiality protection is not applied to the full-quarter estimates (employment, flows and average earnings) because there is no comparable estimate produced by the BLS from ES-202 data.

7 The Final product

7.1 Summary Variable Definitions

Timing and Category Variables

Timing and categorical variables are used to describe the population and time period that the content variables cover. The first such variable is STATE, which is the two-digit FIPS code for the state upon which the employment dynamics estimates are based. The next two variables (YEAR and QUARTER) refer to the calendar year and quarter covered by the content variables. The COUNTY variable (county-level data file) is the three-digit FIPS code for the county (within the state). The SIC_DIVISION variable (sic-division-level data file) is the one-character SIC (1987) major industry group. The SEX variable indicates whether the data cover men or women. The AGEGROUP variable indicates which of the eight age categories the data cover.

Content Variables

The quarterly employment estimates for beginning of quarter employment are contained the variable B and the estimates for end of quarter employment are found in the variable E. Accessions are reported in the variable A. New hires are in H and recalls are reported in R. Separations are reported in the variable S.

Because of the confidentiality protection system used for the Employment Dynamics Estimates, the estimate of beginning-of-quarter employment for both sexes (SEX=0) and all age groups (AGEGROUP=0) is exactly equal to the BLS-published Covered Employment and Wages estimate of employment on the 12th day of the first month of the quarter for the relevant geographic and industrial category. For example, in California the EDE estimate for beginning-of-quarter employment in the entire state in 1999:3 is 14,440,000 (B=14,440,000 for STATE="06", YEAR=1999, QUARTER=3, COUNTY="000" (or SIC_DIVISION=(blank)),SEX=0, AGEGROUP=0), which exactly equals the BLS CEW estimate for month 1 in 1999:3 for the entire state, combining all establishment sizes and all ownership categories. Similarly, the EDE estimate of end-of-quarter employment is controlled for the category both sexes (SEX=0) and all age groups (AGEGROUP=0) to equal the BLS-published CEW estimate of employment on the 12th day of the first month of the succeeding quarter. Again, considering California, the EDE estimate for end-of-quarter employment in 1999:3 is 14,660,000, which exactly equals the BLS CEW estimate for month 1 in 1999:4 for the entire state (E=14,660,000 for STATE="06", YEAR=1999, QUARTER=3, COUNTY="000" (or SIC_DIVISION=(blank)), SEX=0, AGEGROUP=0).

Quarterly employment estimates are also provided on a full-quarter basis. These estimates are reported in the variable F. Full-quarter accessions are reported in FA. Full-quarter separations are reported in FS. Full-quarter new hires are in H3. The raking step of the EDE confidentiality protection system used to disclosure proof the variables B and E (and related variables) does not affect the estimates of full-quarter employment and related flows.

Job creations and destructions are reported in the variables JC and JD, respectively. Net job flows are reported in the variable JF. Full-quarter job creations and destructions are reported in FJC and FJD, respectively. Full-quarter net job flows are in FJF.

Average earnings of full-quarter employees can be found in Z_W3. Average earnings of full-quarter new hires are reported in Z_WH3.

Status Flag Variables

Every variable in EDE data files has an associated status flag. These variables are called [varname]_status. The status flag variables are also shown in the contents tables at the end of this primer. The status flag has three distinct values:

$*$ indicates significant distortion is necessary to preserve confidentiality

d indicates an estimate is based on < 3 employees in the at-risk group.

n indicates an estimate is not defined because no employees are in the relevant category

7.2 Data structure

Appendix B on page 30 describes the contents of a typical output statistics file, in this case for the state of Texas, and aggregated to the county level. A similar file exists at the industry aggregation level, and the same pair of files is constructed for every available state.

7.3 Example statistics

Table 6 shows the extreme case of statistics for a small county and for one gender only.Two types of confidentiality measures are implemented within the same table. Ten cells have been distorted through the injection of noise described earlier. Furthermore, one cell has been suppressed because the statistics in that cell is based on too few individuals.

Table 6: Example table: ALEXANDER, IL, Men only, by age group

	Beginning-of-period employment		Job creation		Job destruction		New hires		Average earnings of full-quarter employees		Average earnings of full-quarter new hires	
All Ages	1,266		67		76		192		6,764		3,945	
14-18	21	*	10	*	1	*	22		1,794	*	921	*
19-21	44		9		5		21		3,835		3,645	*
22-24	74		12		14		19		4,020		3,425	
25-34	286		26		32		58		6,041		5,128	
35-44	356		22		27		38		6,477		4,827	
45-54	278		12		12		25		8,644		1,861	*
55-64	158		2		9		6		8,592		4,207	*
65+	50		1		4			d	3,379		1,216	*

* indicates significant distortion is necessary to preserve confidentiality
d indicates an estimate is based on less than 3 employees in the at-risk group
n indicates an estimate is not defined because no employees are in the relevant category

Figure 1 on the next page presents the data in a different way. The graph shows net job creation rates by county, for both genders, and for the youngest workers. It shows large dispersion across the state, and highlights the value added from joining demographic and firm-level data to form new statistics.

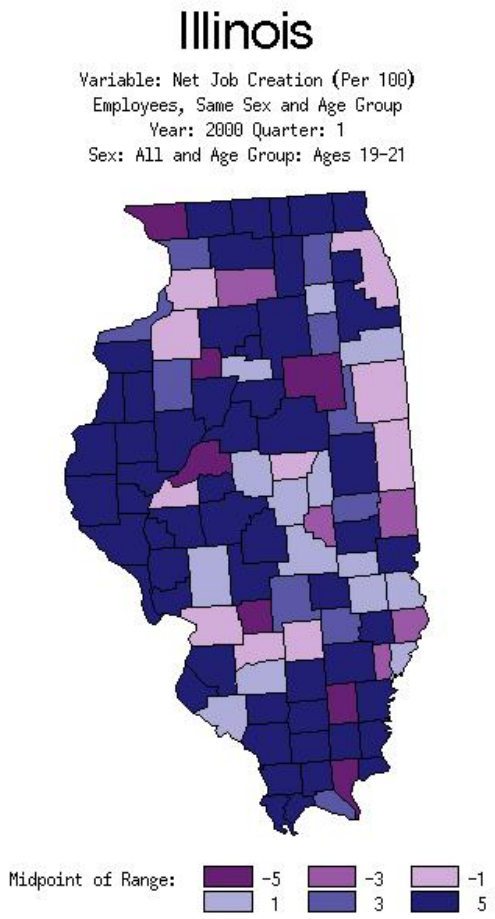

Figure 1: Example graph

Bibliography

Abowd, J. M., Corbel and Kramarz, F. (1999). The entry and exit of workers and the growth of employment: An analysis of French establishments, *Review of Economics and Statistics* **81**(2): 170–87.

Davis, S. and Haltiwanger, J. (1999). Gross job creation and destruction, *in* O. Ashenfelter and D. Card (eds), *Handbook of labor economics*, Vol. 3, North-Holland, Amsterdam, New York.

Davis, S. J., Haltiwanger, J. C. and Schuh, S. (1996). *Job creation and destruction*, MIT Press, Cambridge, MA.

McKinney, K. (2002). Successor/predecessor firms, *Technical paper TP-2002-04*, LEHD, U.S. Census Bureau.

Vilhuber, L. (2001). SSN editing of California UI files: Results, *Technical paper TP-2001-01*, LEHD, U.S. Census Bureau.

Vilhuber, L. (2002a). SSN editing of North Carolina UI files: Results, *Technical paper TP-2002-07*, LEHD, U.S. Census Bureau.

Vilhuber, L. (2002b). SSN editing of Texas UI files: Results, *Technical paper TP-2002-01*, LEHD, U.S. Census Bureau.

A Definitions of Job Flow, Worker Flow, and Earnings Statistics

A.1 Overview and basic data processing conventions

For internal processing the variable t refers to the sequential quarter. The variable t runs from $qmin$ to $qmax$, regardless of the state being processed. The quarters are numbered sequentially from 1 (1985:1) to the latest available quarter. These values are $qmin = 1$ (1985:1) and $qmax = 64$ (2000:4), as of January 17, 2002. For publication, presentation, and internal data files, all dates are presented as (year:quarter) pairs, *e.g.* (1990:1) for first quarter 1990. The variable $qfirst$ refers to the first available sequential quarter of data for a state (*e.g.*, $qfirst = 21$ for Illinois). The variable $qlast$ refers to the last available sequential quarter of data for a state (*e.g.*, $qlast = 62$ for Illinois). Unless otherwise specified a variable is defined for $qfirst \leq t \leq qlast$. Statistics are produced for both sexes combined, as well as separately, for all age groups, ages 14-18, 19-21, 22-24, 25-34, 35-44, 45-54, 55-64, 65+. An individual's age is measured as of the last day of the quarter.

A.2 Individual concepts

Flow employment (m): for $qfirst \leq t \leq qlast$, individual i employed (matched to a job) at some time during period t at employer j

$$m_{ijt} = \begin{cases} 1, \text{ if } i \text{ has positive earnings at employer } j \text{ during quarter } t \\ 0, \text{ otherwise.} \end{cases} \tag{1}$$

Beginning of quarter employment (b): For $qfirst < t$, individual i employed at the end of $t-1$, beginning of t

$$b_{ijt} = \begin{cases} 1, \text{ if } m_{ijt-1} = m_{ijt} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{2}$$

End of quarter employment (e): For $t < qlast$, individual i employed at j at the end of t, beginning of $t+1$

$$e_{ijt} = \begin{cases} 1, \text{ if } m_{ijt} = m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{3}$$

Accessions (a_1): For $qfirst < t$, individual i acceded to j during t

$$a_{1ijt} = \begin{cases} 1, \text{ if } m_{ijt-1} = 0 \ \& \ m_{ijt} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{4}$$

Separations (s_1): For $t < qlast$, individual i separated from j during t

$$s_{1ijt} = \begin{cases} 1, \text{ if } m_{ijt} = 1 \ \& \ m_{ijt+1} = 0 \\ 0, \text{ otherwise.} \end{cases} \tag{5}$$

Full quarter employment (f): For $qfirst < t < qlast$, individual i was employed at j at the beginning and end of quarter t (full-quarter job)

$$f_{ijt} = \begin{cases} 1, \text{ if } m_{ijt-1} = 1 \ \& \ m_{ijt} = 1 \ \& \ m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{6}$$

New hires (h_1): For $qfirst + 3 < t$, individual i was newly hired at j during period t

$$h_{1ijt} = \begin{cases} 1, \text{ if } m_{ijt-4} = 0 \ \& \ m_{ijt-3} = 0 \ \& \ m_{ijt-2} = 0 \ \& \ m_{ijt-1} = 0 \ \& \ m_{ijt} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{7}$$

Recalls (r_1): For $qfirst + 3 < t$, individual$i$ was recalled from layoff at j during period t

$$r_{1ijt} = \begin{cases} 1, \text{ if } m_{ijt-1} = 0 \ \& \ m_{ijt} = 1 \ \& \ h_{ijt} = 0 \\ 0, \text{ otherwise.} \end{cases} \tag{8}$$

Accessions to consecutive quarter status (a_2): For $qfirst < t < qlast$, individual i transited from accession to consecutive-quarter status at j at the start of $t + 1$ (accession in t and still employed at the end of the quarter)

$$a_{2ijt} = \begin{cases} 1, \text{ if } a_{1ijt} = 1 \ \& \ m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{9}$$

Accessions to full quarter status (a_3): For $qfirst + 1 < t < qlast$, individual i transited from consecutive-quarter to full-quarter status at j at the start of $t + 1$ (accession in $t - 1$ and employed for the full quarter in t)

$$a_{3ijt} = \begin{cases} 1, \text{ if } a_{2ijt-1} = 1 \ \& \ m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{10}$$

New hires to consecutive quarter status (h_2): For $qfirst + 3 < t < qlast$, individual i transited from newly hired to consecutive-quarter hired status at j at the start of $t + 1$ (hired in t and still employed at the end of the quarter)

$$h_{2ijt} = \begin{cases} 1, \text{ if } h_{1ijt} = 1 \ \& \ m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{11}$$

New hires to full quarter status (a_3): For $qfirst + 4 < t < qlast$, individual i transited from consecutive-quarter hired to full-quarter hired status at j at the start of $t + 1$ (hired in $t - 1$ and full-quarter employed in t)

$$h_{3ijt} = \begin{cases} 1, \text{ if } h_{2ijt-1} = 1 \ \& \ m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{12}$$

Recalls to consecutive quarter status (r_2): For $qfirst + 3 < t < qlast$, individual i transited from recalled to consecutive-quarter recalled status at j at the start of $t + 1$ (recalled in t and still employed at the end of the quarter)

$$r_{2ijt} = \begin{cases} 1, \text{ if } r_{1ijt} = 1 \ \& \ m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{13}$$

Recalls to full quarter status (r_3): For $qfirst + 4 < t < qlast$, individual i transited from consecutive-quarter recalled to full-quarter recalled status at j at the start of $t + 1$ (recalled in $t - 1$ and full-quarter employed in t)

$$r_{3ijt} = \begin{cases} 1, \text{ if } r_{2ijt-1} = 1 \ \& \ m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{14}$$

Separations from consecutive quarter status (s_2): For $qfirst < t < qlast$, individual i separated from j during t with consecutive-quarter status at the start of t

$$s_{2ijt} = \begin{cases} 1, \text{ if } s_{1ijt} = 1 \ \& \ m_{ijt-1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{15}$$

Separations from full-quarter status (s_3): For $qfirst + 1 < t < qlast$, individual i separated from j during t with full-quarter status during $t-1$

$$s_{3ijt} = \begin{cases} 1, \text{ if } s_{2ijt} = 1 \ \ \& \ \ m_{ijt-2} = 1 \\ 0, \text{ otherwise} \end{cases} \tag{16}$$

Total earnings during the quarter (w_1): for $qfirst \leq t \leq qlast$, earnings of individual i at employer j during period t

$$w_{1ijt} = \sum \text{all } UI \text{ covered earnings by } i \text{ at } j \text{ during } t \tag{17}$$

Earnings of end-of-period employees at employer j during period t

$$w_{2ijt} = \begin{cases} w_{1ijt}, \text{ if } e_{ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{18}$$

Earnings of full-quarter individual i at employer j during period t

$$w_{3ijt} = \begin{cases} w_{1ijt}, \text{ if } f_{ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{19}$$

For $qfirst \leq t \leq qlast$, total earnings of individual i during period t

$$w_{1i\bullet t} = \sum_{j \text{ employs } i \text{ during } t} w_{1ijt} \tag{20}$$

Total earnings of end-of-period employees i during period t

$$w_{2i\bullet t} = \begin{cases} w_{1i\bullet t}, \text{ if } e_{ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{21}$$

Total earnings of full-quarter employees i during period t

$$w_{3i\bullet t} = \begin{cases} w_{1i\bullet t}, \text{ if } f_{ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{22}$$

For $qfirst < t$, change in total earnings of individual i between periods $t-1$ and t. The goal is to produce statistics based on:

$$\Delta w_{1i\bullet t} = w_{1i\bullet t} - w_{1i\bullet t-1} \tag{23}$$

Earnings of accessions to employer j during period t

$$wa_{1ijt} = \begin{cases} w_{1ijt}, \text{ if } a_{1ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{24}$$

Earnings of consecutive-quarter accessions to employer j during period t

$$wa_{2ijt} = \begin{cases} w_{1ijt}, \text{ if } a_{2ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{25}$$

Earnings of full-quarter accessions to employer j during period t

$$wa_{3ijt} = \begin{cases} w_{1ijt}, \text{ if } a_{3ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{26}$$

Earnings of full-quarter new hires to employer j during period t

$$wh_{3ijt} = \begin{cases} w_{1ijt}, \text{ if } h_{3ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{27}$$

Total earnings change for accessions to employer j during t

$$\Delta wa_{1ijt} = \begin{cases} \Delta w_{1i\bullet t}, \text{ if } a_{1ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{28}$$

Total earnings change for full-quarter accessions to employer j during t

$$\Delta wa_{3ijt} = \begin{cases} \Delta w_{1i\bullet t}, \text{ if } a_{3ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{29}$$

Earnings of separations from employer j during period t

$$ws_{1ijt} = \begin{cases} w_{1ijt}, \text{ if } s_{1ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{30}$$

Earnings of full-quarter separations to employer j during period t

$$ws_{3ijt} = \begin{cases} w_{1ijt}, \text{ if } s_{3ijt+1} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{31}$$

Total earnings change for separations from employer j during t

$$\Delta ws_{1ijt} = \begin{cases} \Delta w_{1i\bullet t+1}, \text{ if } s_{1ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{32}$$

Total earnings change for full-quarter separations from employer j during t

$$\Delta ws_{3ijt} = \begin{cases} \Delta w_{1i\bullet t+1}, \text{ if } s_{3ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{33}$$

Periods of non-employment prior to an accession by i at employer j during t during the previous four quarters (defined for $qfirst + 3 < t$)

$$na_{ijt} = \begin{cases} \sum\limits_{1 \leqslant s \leqslant 4} n_{it-s}, \text{ if } a_{1ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{34}$$

where $n_{it} = 1$ if $m_{ijt} = 0 \; \forall j$.

Periods of non-employment prior to a new hire by i at employer j during t during the previous four quarters

$$nh_{ijt} = \begin{cases} \sum_{1 \leqslant s \leqslant 4} n_{it-s}, \text{ if } h_{1ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{35}$$

Periods of non-employment prior to a recall by i at employer j during t during the previous four quarters

$$nr_{ijt} = \begin{cases} \sum_{1 \leqslant s \leqslant 4} n_{it-s}, \text{ if } r_{1ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{36}$$

Periods of non-employment following a separation by i from employer j during t during the next four quarters, (defined for $t < qlast - 3$)

$$ns_{ijt} = \begin{cases} \sum_{1 \leqslant s \leqslant 4} n_{it+s}, \text{ if } s_{1ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{37}$$

A.3 Employer concepts

For statistic x_{cijt} denote the sum over i during period t as $x_{c \cdot jt}$. For example, beginning of period employment for firm j is written as:

$$b_{\cdot jt} = \sum_i b_{ijt} \tag{38}$$

All individual statistics generate employer totals according to the formula above. The key employer statistic is the average end-of-period employment growth rate for employer j, the components of which are defined here.

Beginning-of-period employment (number of jobs)

$$B_{jt} = b_{\cdot jt} \tag{39}$$

End-of-period employment (number of jobs)

$$E_{jt} = e_{\cdot jt} \tag{40}$$

Employment any time during the period (number of jobs)

$$M_{jt} = m_{\cdot jt} \tag{41}$$

Full-quarter employment

$$F_{jt} = f_{\cdot jt} \tag{42}$$

Net job flows (change in employment) for employer j during period t

$$JF_{jt} = E_{jt} - B_{jt} \tag{43}$$

Average employment for employer j between periods $t-1$ and t

$$\bar{E}_{jt} = \frac{(B_{jt} + E_{jt})}{2} \tag{44}$$

Average employment growth rate for employer j between periods $t-1$ and t

$$G_{jt} = \frac{JF_{jt}}{\bar{E}_{jt}} \tag{45}$$

Job creation for employer j between periods $t-1$ and t

$$JC_{jt} = \bar{E}_{jt} \max\left(0, G_{jt}\right) \tag{46}$$

Average job creation rate for employer j between periods $t-1$ and t

$$JCR_{jt} = \frac{JC_{jt}}{\bar{E}_{jt}} \tag{47}$$

Job destruction for employer j between periods $t-1$ and t

$$JD_{jt} = \bar{E}_{jt} \operatorname{abs}\left(\min\left(0, G_{jt}\right)\right) \tag{48}$$

Average job destruction rate for employer j between periods $t-1$ and t

$$JDR_{jt} = \frac{JD_{jt}}{\bar{E}_{jt}} \tag{49}$$

Net change in full-quarter employment for employer j during period t

$$FJF_{jt} = F_{jt} - F_{jt-1} \tag{50}$$

Average full-quarter employment for employer j during period t

$$\bar{F}_{jt} = \frac{F_{jt-1} + F_{jt}}{2} \tag{51}$$

Average full-quarter employment growth rate for employer j between $t-1$ and t

$$FG_{jt} = \frac{FJF_{jt}}{\bar{F}_{jt}} \tag{52}$$

Full-quarter job creations for employer j between $t-1$ and t

$$FJC_{jt} = \bar{F}_{jt} \max\left(0, FG_{jt}\right) \tag{53}$$

Average full-quarter job creation rate for employer j between $t-1$ and t

$$FJCR_{jt} = FJC_{jt} \,/\, \bar{F}_{jt} \tag{54}$$

Full-quarter job destruction for employer j between $t-1$ and t

$$FJD_{jt} = \bar{F}_{jt} abs\left(\min\left(0, FG_{jt}\right)\right) \tag{55}$$

Average full-quarter job destruction rate for employer j between $t-1$ and t

$$FJDR_{jt} = FJD_{jt} \,/\, \bar{F}_{jt} \tag{56}$$

Accessions for employer j during t

$$A_{jt} = a_{1 \cdot jt} \tag{57}$$

Average accession rate for employer j during t

$$AR_{jt} = A_{jt} \Big/ \bar{E}_{jt} \tag{58}$$

Separations for employer j during t

$$S_{jt} = s_{1 \cdot jt} \tag{59}$$

Average separation rate for employer j during t

$$SR_{jt} = S_{jt} \Big/ \bar{E}_{jt} \tag{60}$$

New hires for employer j during t

$$H_{jt} = h_{1 \cdot jt} \tag{61}$$

Full Quarter New hires for employer j during t

$$H_{3jt} = h_{3 \cdot jt} \tag{62}$$

Recalls for employer j during t

$$R_{jt} = r_{1 \cdot jt} \tag{63}$$

Flow into full-quarter employment for employer j during t

$$FA_{jt} = a_{3 \cdot jt} \tag{64}$$

New hires into full-quarter employment for employer j during t

$$FH_{jt} = h_{3 \cdot jt} \tag{65}$$

Average rate of flow into full-quarter employment for employer j during t

$$FAR_{jt} = FA_{jt} \Big/ \bar{F}_{jt} \tag{66}$$

Flow out of full-quarter employment for employer j during t

$$FS_{jt} = s_{3 \cdot jt} \tag{67}$$

Average rate of flow out of full-quarter employment for employer j during t

$$FSR_{jt} = FS_{jt} \Big/ \bar{F}_{jt} \tag{68}$$

Flow into consecutive quarter employment for employer j during t

$$CA_{jt} = a_{2 \cdot jt} \tag{69}$$

Flow out of consecutive quarter employment for employer j during t

$$CS_{jt} = s_{2 \cdot jt} \tag{70}$$

Total payroll of all employees

$$W_{1jt} = w_{1 \cdot jt} \tag{71}$$

Total payroll of end-of-period employees

$$W_{2jt} = w_{2 \cdot jt} \tag{72}$$

Total payroll of full-quarter employees

$$W_{3jt} = w_{3 \cdot jt} \tag{73}$$

Total payroll of accessions

$$WA_{jt} = wa_{1 \cdot jt} \tag{74}$$

Change in total earnings for accessions

$$\Delta WA_{jt} = \sum_{i \in \{J(i,t)=j\}} \triangle wa_{1ijt} \tag{75}$$

Total payroll of transits to consecutive-quarter status

$$WCA_{jt} = wa_{2 \cdot jt} \tag{76}$$

Total payroll of transits to full-quarter status

$$WFA_{jt} = wa_{3 \cdot jt} \tag{77}$$

Total payroll of new hires to full-quarter status

$$WFH_{jt} = wh_{3 \cdot jt} \tag{78}$$

Change in total earnings for transits to full-quarter status

$$\Delta WFA_{jt} = \sum_{i \in \{J(i,t)=j\}} \Delta wa_{3ijt} \tag{79}$$

Total periods of non-employment for accessions

$$NA_{jt} = na_{\cdot jt} \tag{80}$$

Total periods of non-employment for new hires (last four quarters)

$$NH_{jt} = nh_{\cdot jt} \tag{81}$$

Total periods of non-employment for recalls (last four quarters)

$$NR_{jt} = nr_{\cdot jt} \tag{82}$$

Total earnings of separations

$$WS_{jt} = ws_{1 \cdot jt} \tag{83}$$

Total change in total earnings for separations

$$\Delta WS_{jt} = \sum_{i \in \{J(i,t)=j\}} \Delta ws_{1ijt} \tag{84}$$

Total earnings of separations from full-quarter status (most recent full quarter)

$$WFS_{jt} = ws_{3 \cdot jt} \tag{85}$$

Total change in total earnings for full-quarter separations

$$\Delta WFS_{jt} = \sum_{i \in \{J(i,t)=j\}} \Delta ws_{3ijt} \tag{86}$$

Total periods of non-employment for separations

$$NS_{jt} = ns_{\cdot jt} \tag{87}$$

Average earnings of end-of-period employees

$$ZW_{2jt} = W_{2jt} \,/\, E_{jt} \tag{88}$$

Average earnings of full-quarter employees

$$ZW_{3jt} = W_{3jt} \,/\, F_{jt} \tag{89}$$

Average earnings of accessions

$$ZWA_{jt} = WA_{jt} \,/\, A_{jt} \tag{90}$$

Average change in total earnings for accessions

$$Z\Delta WA_{jt} = \Delta WA_{jt} \,/\, A_{jt} \tag{91}$$

Average earnings of transits to full-quarter status

$$ZWFA_{jt} = WFA_{jt} \,/\, FA_{jt} \tag{92}$$

Average earnings of new hires to full-quarter status

$$ZWFH_{jt} = WFH_{jt} \,/\, FH_{jt} \tag{93}$$

Average change in total earnings for transits to full-quarter status

$$Z\Delta WFA_{jt} = \Delta WFA_{jt} \,/\, FA_{jt} \tag{94}$$

Average periods of non-employment for accessions

$$ZNA_{jt} = NA_{jt} \,/\, A_{jt} \tag{95}$$

Average periods of non-employment for new hires (last four quarters)

$$ZNH_{jt} = NH_{jt} \,/\, H_{jt} \tag{96}$$

Average periods of non-employment for recalls (last four quarters)

$$ZNR_{jt} = NR_{jt} \,/\, R_{jt} \tag{97}$$

Average earnings of separations

$$ZWS_{jt} = WS_{jt} \,/\, S_{jt} \tag{98}$$

Average change in total earnings for separations

$$Z\Delta WS_{jt} = \Delta WS_{jt} \,/\, S_{jt} \tag{99}$$

Average earnings of separations from full-quarter status (most recent full quarter)

$$ZWFS_{jt-1} = WFS_{jt-1} \,/\, FS_{jt} \tag{100}$$

Average change in total earnings for full-quarter separations

$$Z\Delta WFS_{jt} = \Delta WFS_{jt} \,/\, FS_{jt} \tag{101}$$

Average periods of non-employment for separations

$$ZNS_{jt} = NS_{jt} \,/\, S_{jt} \tag{102}$$

End-of-period employment (number of workers) [Aggregate concept not related to a business]

$$N_t = n_{.t} \tag{103}$$

A.4 Identities

Identity 1 Employment at beginning of period t equals end of period $t-1$

$$B_{jt} = E_{jt-1}$$

Identity 2 Evolution of end of period employment

$$E_{jt} = B_{jt} + A_{jt} - S_{jt}$$

Identity 3 Evolution of average employment

$$\bar{E}_{jt} = B_{jt} + (A_{jt} - S_{jt}) \,/\, 2$$

Identity 4 Job flow identity

$$JF_{jt} = JC_{jt} - JD_{jt}$$

Identity 5 Creation-destruction identity

$$E_{jt} = B_{jt} + JC_{jt} - JD_{jt}$$

Identity 6 Creation-destruction/accession-separation identity

$$A_{jt} - S_{jt} = JC_{jt} - JD_{jt}$$

Identity 7 Evolution of full-quarter employment

$$F_{jt} = F_{jt-1} + FA_{\ jt} - FS_{jt}$$

Identity 8 Full-quarter creation-destruction identity

$$F_{jt} = F_{jt-1} + FJC_{jt} - FJD_{jt}$$

Identity 9 Full-quarter job flow identity

$$FJF_{jt} = FJC_{jt} - FJD_{jt}$$

Identity 10 Full-quarter creation-destruction/accession-separation identity

$$FA_{jt} - FS_{jt} = FJC_{jt} - FJD_{jt}$$

Identity 11 Employment growth rate identity

$$G_{jt} = JCR_{jt} - JDR_{jt}$$

Identity 12 Creation-destruction/accession-separation rate identity

$$JCR_{jt} - JDR_{jt} = AR_{jt} - SR_{jt}$$

Identity 13 Full quarter employment growth rate identity

$$FG_{jt} = FJCR_{jt} - FJDR_{jt}$$

Identity 14 Full quarter creation-destruction/accession-separation rate identity

$$FJCR_{jt} - FJDR_{jt} = FAR_{jt} - FSR_{jt}$$

Identity 15 Total payroll identity

$$W_{1jt} = W_{2jt} + WS_{jt}$$

Identity 16 Payroll identity for consecutive-quarter employees

$$W_{2jt} = W_{1jt} - WCA_{jt} - WS_{jt}$$

Identity 17 Full-quarter payroll identity

$$W_{3jt} = W_{2jt} - WCA_{jt}$$

Identity 18 New hires/recalls identity

$$A_{jt} = H_{jt} + R_{jt}$$

Identity 19 Periods of non-employment identity

$$NA_{jt} = NH_{jt} + NR_{jt}$$

Identity 20 Worker-jobs in period t are the sum of accessions and beginning of period employment.

$$M_{jt} = A_{jt} + B_{jt}$$

Identity 21 Worker-jobs in period t are the sum of accessions to consecutive quarter status, separations, and full quarter workers.

$$M_{jt} = CA_{jt} + S_{jt} + F_{jt}$$

Identity 22 Consecutive quarter accessions in period $t-1$ are the sum of consecutive quarter separations in period t and full quarter accessions in period t

$$CA_{jt-1} - CS_{jt} = FA_{jt} - FS_{jt}$$

A.5 Aggregation of flows

The rate of growth is equal to the ratio of net job flows to total employment:

$$G_{jt} = JF_{jt} \,/\, \bar{E}_{jt} \tag{104}$$

So, to impute the aggregate growth rate in a county (or sic) for some group of firms, let

$$G_{kt} = \frac{\sum\limits_{j \in \{K(j)=k\}} \bar{E}_{jt} \times G_{jt}}{\bar{E}_{kt}} \tag{105}$$

for county k where the function $K(j)$ indicates the classification associated with firm j.

We calculate the aggregate job flow as

$$JF_{kt} = \sum_{j \in \{K(j)=k\}} JF_{jt}. \tag{106}$$

Substitution yields

$$JF_{kt} = \sum_{j} (\bar{E}_{jt} \times G_{jt}) = G_{kt} \times \bar{E}_{kt}, \tag{107}$$

so the aggregate job flow, as computed, is equivalent to the aggregate growth rate times aggregate employment. Gross job creation/destruction are related to job creation/destruction rates by similar logic (Davis et al.; 1996, p. 189 for details).

B Example data structure

The CONTENTS Procedure

Data Set Name:	STATE.IL_COUNTY_V23	**Observations:**	116802
Member Type:	DATA	**Variables:**	60
Engine:	V8	**Indexes:**	0
Created:	18:27 Thursday, May 16, 2002	**Observation Length:**	288
Last Modified:	18:27 Thursday, May 16, 2002	**Deleted Observations:**	0
Protection:		**Compressed:**	NO
Data Set Type:		**Sorted:**	NO
Label:			

—–Variables Ordered by Position—–

#	Variable	Type	Len	Pos	Label
1	state	Char	2	216	FIPS State
2	year	Num	3	275	Year
3	quarter	Num	3	278	Quarter
4	county	Char	3	218	FIPS county
5	sex	Num	3	281	Sex

(cont.)

——Variables Ordered by Position——

#	Variable	Type	Len	Pos	Label
6	agegroup	Num	3	284	Age group
7	A	Num	8	0	Accessions
8	B	Num	8	8	Beginning-of-period employment
9	E	Num	8	16	End-of-period employment
10	F	Num	8	24	Full-quarter employment
11	FA	Num	8	32	Flow into full-quarter employment
12	FJC	Num	8	40	Full-quarter job creation
13	FJD	Num	8	48	Full-quarter job destruction
14	FJF	Num	8	56	Net change in full-quarter employment
15	FS	Num	8	64	Flow out of full-quarter employment
16	H	Num	8	72	New hires
17	H3	Num	8	80	Full-quarter new hires
18	JC	Num	8	88	Job creation
19	JD	Num	8	96	Job destruction
20	JF	Num	8	104	Net job flows
21	R	Num	8	112	Recalls
22	S	Num	8	120	Separations
23	Z_NA	Num	8	128	Average periods of non-employment for accessions
24	Z_NH	Num	8	136	Average periods of non-employment for new hires
25	Z_NR	Num	8	144	Average periods of non-employment for recalls
26	Z_NS	Num	8	152	Average periods of non-employment for separations
27	Z_W2	Num	8	160	Average earnings of end-of-period employees
28	Z_W3	Num	8	168	Average earnings of full-quarter employees
29	Z_WFA	Num	8	176	Average earnings of transits to full-quarter status
30	Z_WFS	Num	8	184	Average earnings of separations from full-quarter status
31	Z_WH3	Num	8	192	Average earnings of full-quarter new hires
32	Z_dWA	Num	8	200	Average change in total earnings for accessions
33	Z_dWS	Num	8	208	Average change in total earnings for separations
34	A_status	Char	2	221	Status: accessions
35	B_status	Char	2	223	Status: beginning-of-period employment
36	E_status	Char	2	225	Status: end-of-period employment
37	F_status	Char	2	227	Status: full-quarter employment
38	FA_status	Char	2	229	Status: flow into full-quarter employment
39	FJC_status	Char	2	231	Status: full-quarter job creation
40	FJD_status	Char	2	233	Status: full-quarter job destruction
41	FJF_status	Char	2	235	Status: net change in full-quarter employment
42	FS_status	Char	2	237	Status: flow out of full-quarter employment
43	H_status	Char	2	239	Status: new hires

(cont.)

——Variables Ordered by Position——

#	Variable	Type	Len	Pos	Label
44	H3_status	Char	2	241	Status: full-quarter new hires
45	JC_status	Char	2	243	Status: job creation
46	JD_status	Char	2	245	Status: job destruction
47	JF_status	Char	2	247	Status: net job flows
48	R_status	Char	2	249	Status: recalls
49	S_status	Char	2	251	Status: separations
50	Z_NA_status	Char	2	253	Status: average periods of non-employment for accessions
51	Z_NH_status	Char	2	255	Status: average periods of non-employment for new hires
52	Z_NR_status	Char	2	257	Status: average periods of non-employment for recalls
53	Z_NS_status	Char	2	259	Status: average periods of non-employment for separations
54	Z_W2_status	Char	2	261	Status: average earnings of end-of-period employees
55	Z_W3_status	Char	2	263	Status: average earnings of full-quarter employees
56	Z_WFA_status	Char	2	265	Status: average earnings of transits to full-quarter status
57	Z_WFS_status	Char	2	267	Status: average earnings of separations from full-quarter status
58	Z_WH3_status	Char	2	269	Status: average earnings of full-quarter new hires
59	Z_dWA_status	Char	2	271	Status: average change in total earnings for accessions
60	Z_dWS_status	Char	2	273	Status: average change in total earnings for separations

Illinois

The FREQ Procedure

FIPS State				
state	Frequency	Percent	CumulativeFrequency	CumulativePercent
17 ILLINOIS	116802	100.00	116802	100.00

FIPS county				
county	Frequency	Percent	CumulativeFrequency	CumulativePercent
000 ILLINOIS	1134	0.97	1134	0.97
001 ADAMS	1134	0.97	2268	1.94
003 ALEXANDER	1134	0.97	3402	2.91
005 BOND	1134	0.97	4536	3.88
007 BOONE	1134	0.97	5670	4.85
009 BROWN	1134	0.97	6804	5.83
011 BUREAU	1134	0.97	7938	6.80
013 CALHOUN	1134	0.97	9072	7.77
015 CARROLL	1134	0.97	10206	8.74
017 CASS	1134	0.97	11340	9.71
019 CHAMPAIGN	1134	0.97	12474	10.68
021 CHRISTIAN	1134	0.97	13608	11.65
023 CLARK	1134	0.97	14742	12.62
025 CLAY	1134	0.97	15876	13.59
027 CLINTON	1134	0.97	17010	14.56
029 COLES	1134	0.97	18144	15.53
031 COOK	1134	0.97	19278	16.50
033 CRAWFORD	1134	0.97	20412	17.48
035 CUMBERLAND	1134	0.97	21546	18.45
037 DEKALB	1134	0.97	22680	19.42
039 DE WITT	1134	0.97	23814	20.39
041 DOUGLAS	1134	0.97	24948	21.36
043 DUPAGE	1134	0.97	26082	22.33
045 EDGAR	1134	0.97	27216	23.30
047 EDWARDS	1134	0.97	28350	24.27
049 EFFINGHAM	1134	0.97	29484	25.24
051 FAYETTE	1134	0.97	30618	26.21
053 FORD	1134	0.97	31752	27.18
055 FRANKLIN	1134	0.97	32886	28.16

(cont.)

FIPS county				
county	**Frequency**	**Percent**	**CumulativeFrequency**	**CumulativePercent**
057 FULTON	1134	0.97	34020	29.13
059 GALLATIN	1134	0.97	35154	30.10
061 GREENE	1134	0.97	36288	31.07
063 GRUNDY	1134	0.97	37422	32.04
065 HAMILTON	1134	0.97	38556	33.01
067 HANCOCK	1134	0.97	39690	33.98
069 HARDIN	1134	0.97	40824	34.95
071 HENDERSON	1134	0.97	41958	35.92
073 HENRY	1134	0.97	43092	36.89
075 IROQUOIS	1134	0.97	44226	37.86
077 JACKSON	1134	0.97	45360	38.83
079 JASPER	1134	0.97	46494	39.81
081 JEFFERSON	1134	0.97	47628	40.78
083 JERSEY	1134	0.97	48762	41.75
085 JO DAVIESS	1134	0.97	49896	42.72
087 JOHNSON	1134	0.97	51030	43.69
089 KANE	1134	0.97	52164	44.66
091 KANKAKEE	1134	0.97	53298	45.63
093 KENDALL	1134	0.97	54432	46.60
095 KNOX	1134	0.97	55566	47.57
097 LAKE	1134	0.97	56700	48.54
099 LA SALLE	1134	0.97	57834	49.51
101 LAWRENCE	1134	0.97	58968	50.49
103 LEE	1134	0.97	60102	51.46
105 LIVINGSTON	1134	0.97	61236	52.43
107 LOGAN	1134	0.97	62370	53.40
109 MCDONOUGH	1134	0.97	63504	54.37
111 MCHENRY	1134	0.97	64638	55.34
113 MCLEAN	1134	0.97	65772	56.31
115 MACON	1134	0.97	66906	57.28
117 MACOUPIN	1134	0.97	68040	58.25
119 MADISON	1134	0.97	69174	59.22
121 MARION	1134	0.97	70308	60.19
123 MARSHALL	1134	0.97	71442	61.17
125 MASON	1134	0.97	72576	62.14
127 MASSAC	1134	0.97	73710	63.11
129 MENARD	1134	0.97	74844	64.08
131 MERCER	1134	0.97	75978	65.05

(cont.)

FIPS county				
county	**Frequency**	**Percent**	**CumulativeFrequency**	**CumulativePercent**
133 MONROE	1134	0.97	77112	66.02
135 MONTGOMERY	1134	0.97	78246	66.99
137 MORGAN	1134	0.97	79380	67.96
139 MOULTRIE	1134	0.97	80514	68.93
141 OGLE	1134	0.97	81648	69.90
143 PEORIA	1134	0.97	82782	70.87
145 PERRY	1134	0.97	83916	71.84
147 PIATT	1134	0.97	85050	72.82
149 PIKE	1134	0.97	86184	73.79
151 POPE	1134	0.97	87318	74.76
153 PULASKI	1134	0.97	88452	75.73
155 PUTNAM	1134	0.97	89586	76.70
157 RANDOLPH	1134	0.97	90720	77.67
159 RICHLAND	1134	0.97	91854	78.64
161 ROCK ISLAND	1134	0.97	92988	79.61
163 ST.CLAIR	1134	0.97	94122	80.58
165 SALINE	1134	0.97	95256	81.55
167 SANGAMON	1134	0.97	96390	82.52
169 SCHUYLER	1134	0.97	97524	83.50
171 SCOTT	1134	0.97	98658	84.47
173 SHELBY	1134	0.97	99792	85.44
175 STARK	1134	0.97	100926	86.41
177 STEPHENSON	1134	0.97	102060	87.38
179 TAZEWELL	1134	0.97	103194	88.35
181 UNION	1134	0.97	104328	89.32
183 VERMILION	1134	0.97	105462	90.29
185 WABASH	1134	0.97	106596	91.26
187 WARREN	1134	0.97	107730	92.23
189 WASHINGTON	1134	0.97	108864	93.20
191 WAYNE	1134	0.97	109998	94.17
193 WHITE	1134	0.97	111132	95.15
195 WHITESIDE	1134	0.97	112266	96.12
197 WILL	1134	0.97	113400	97.09
199 WILLIAMSON	1134	0.97	114534	98.06
201 WINNEBAGO	1134	0.97	115668	99.03
203 WOODFORD	1134	0.97	116802	100.00

Sex

sex	Frequency	Percent	CumulativeFrequency	CumulativePercent
0 : All	38934	33.33	38934	33.33
1 : Men	38934	33.33	77868	66.67
2 : Women	38934	33.33	116802	100.00

Age group

agegroup	Frequency	Percent	CumulativeFrequency	CumulativePercent
0 : All Ages	12978	11.11	12978	11.11
1 : 14-18	12978	11.11	25956	22.22
2 : 19-21	12978	11.11	38934	33.33
3 : 22-24	12978	11.11	51912	44.44
4 : 25-34	12978	11.11	64890	55.56
5 : 35-44	12978	11.11	77868	66.67
6 : 45-54	12978	11.11	90846	77.78
7 : 55-64	12978	11.11	103824	88.89
8 : 65+	12978	11.11	116802	100.00

Table of year by quarter

year(Year)	quarter(Quarter)				Total
	1	2	3	4	
1990	2781	2781	2781	2781	11124
1991	2781	2781	2781	2781	11124
1992	2781	2781	2781	2781	11124
1993	2781	2781	2781	2781	11124
1994	2781	2781	2781	2781	11124
1995	2781	2781	2781	2781	11124
1996	2781	2781	2781	2781	11124
1997	2781	2781	2781	2781	11124
1998	2781	2781	2781	2781	11124
1999	2781	2781	2781	2781	11124
2000	2781	2781	0	0	5562
Total	30591	30591	27810	27810	116802