Measuring firm-level displacement events with administrative data - draft[1]

Lars Vilhuber [2] Kalyani Raghunathan [3]

DRAFT - October 31, 2013

[1]The computations for this research where done while at the New York Census Research Data Centers in New York City, NY, and Ithaca, NY. Support by NSF Grant SES-0820349 and the Sloan Foundation is gratefully acknowledged.

[2]Cornell University and Labor Dynamics Institute, lars.vilhuber@cornell.edu

[3]Cornell University, kr328@cornell.edu

Abstract

1 Introduction

The recession of 2008-2009 has triggered a resurgence in the interest in the causes and consequences of mass layoffs (Couch and Placzek; 2010; Sullivan and von Wachter; 2009a,b). The seminal work of Jacobson et al. (1993a,b) [JLS] laid the groundwork for the use of evaluation methodology and administrative data as tools to understand mass layoffs. However, there has been little research into understanding some of the core measurement issues. Some previous work has focussed on how workers remember and report displacement (Evans and Leighton; 1995; Hildreth et al.; 2005), but little effort has been expended to consider how mass layoffs are reasonably defined, with most papers using some variation of the JLS definition of a mass layoff event and the classification of workers into the treatment or comparison group.

While most of the literature that uses administrative data has focussed on net employment declines, survey measures have focussed on questions that correlate more naturally with layoffs. In this paper, we will use a number of datasets with broad national coverage to explore different definitions of a significant personnel event, including net employment declines, spikes in layoffs, and spikes in labor turnover.

2 Data

LEHD Data

The core data used in this paper are derived from the Longitudinal Employer-Household Dynamics (LEHD) Infrastructure Files. The LEHD Infrastructure Files provide a worker's quarterly earnings history (derived from unemployment insurance wage records), basic demographic information, and, most importantly, identify a person's employer. The fact that we know the history of the firm and the employees at that firm over time allows us to estimate displacement events as well as provide a richer characterization of the employees at each firm. In particular, we can compute all relevant measures from the same core data (see also Abowd and Vilhuber; 2011).

We use data from xx states The choice of these states is driven primarily by data availability: data for these states covered the largest available time period. For some of these states, data was available from the early 1990s, and for all of these states, data from the mid-1990s through 2012Q1 was available

Table 2 shows the distribution of industry employment in these states relative to the national distribution. [1]

Longitudinal Business Database

We further use data derived from the Census Bureau's Longitudinal Business Database (LBD) (and SynLBD, see Kinney et al.; 2011). In contrast to the LEHD data, which relies on a model to assign workers to establishments, the LBD directly provides establishment-level employment declines. The LBD is frequently used to study employment dynamics Miranda and Jarmin (2002) but has not been used to explore mass layoffs.

We contrast time-series derived from these two data sources to publicly available data sources that have been previously used: Current Population Survey (CPS) Displaced Worker Supplements and the Mass Layoff Statistics (MLS), and explore in particular the behavior of these different data across the business cycle.[2]

Mass Layoff Statistics

The Bureau of Labor Statistics (BLS)' MLS are an invaluable source to obtain the number of events and initial claimants.

[1] The distribution data are derived from public-use QWI data. They are thus comparable in coverage and scope.

[2] We note that the analysis of an alternative to the MLS is particularly germane, since the program has been eliminated as of March 2013, with the final publication occurring in June 2013.

Current Population Survey

While the LEHD and LBD data can illustrate the number of firm or establishment level events, the LEHD and the CPS can be used to obtain an estimate of the number of people affected.

3 Defining displacement events for firms

The dataset of displacement events and workers affected by it is created from the raw wage record employment history, in several steps. We will need some definition to properly define some of the concepts. Wage records only denote that an individual received earnings at some point during the quarter. To get to meaningful definitions, we need some auxiliary definitions (Abowd et al.; 2009).

We will denote worker-level concepts by lower-case letters, and firm level concepts with upper-case letters. Worker-level concepts will have indices i for the worker, j for the firm the worker is working at, and t for the time period.

The presence of a wage record for worker i at firm j in period t is denoted by $m_{ijt} = 1$. A worker is deemed employed at the beginning of the quarter, $b_{ijt} = 1$ if $m_{ijt} = 1$ and $m_{ijt-1} = 1$. This provides us with a point-in-time employment definition. A firm's employment at the beginning of the period is thus $B_{jt} = \sum_i b_{ijt}$. This will be our fundamental employment concept. Equivalently, end-of-period employment $e_{ijt} = 1$ if $m_{ijt} = 1$ and $m_{ijt+1} = 1$. Note that $e_{ijt-1} = b_{ijt}$. Conceptually, B_{jt} captures persistent jobs, whereas M_{jt} captures all jobs, including transitory and very short-term jobs.

A worker is observed to separate from the firm if he is present at the firm at the start of the quarter, and absent at the end of the quarter: $b_{ijt} = 1$ and $e_{ijt} = 0$. A wage record is still present at t ($m_{ijt} = 1$), but not in the next. Separations are denoted as $s_{ijt} = 1$ if $m_{ijt} = 1$ and $m_{ijt+1} = 0$. and the firm aggregation - the total flow of workers out of the firm is S_{jt}. Equivalently, accessions (hires) occur when $m_{ijt-1} = 0$ and $m_{ijt} = 1$, and are denoted by a_{ijt}. While these measures capture all jobs, including transitory jobs, the measures cs_{ijt} and ca_{ijt} are defined as

$$ca_{ijt} = 1 \iff b_{ijt} = 0 \wedge e_{ijt} = 1$$

and

$$cs_{ijt} = 1 \iff b_{ijt} = 1 \wedge e_{ijt} = 0$$

Note that

$$B_{jt} + A_{jt} - S_{jt} = E_{jt} = B_{jt+1}. \tag{1}$$

We will use several different point-in-time definitions that capture different components of the workforce: transitory, short-term, and medium-term. Of course, long-term attachment, limited only by data availability can also be computed.

The core of the paper centers around using different ways of aggregating these data and determining when a "mass layoff" event occurs. We thus define several definitions for a mass layoff, each of which has an intuitive appeal and often an empirical or historical counterpart.

An straightforward candidate for a "mass layoff" definition is when there are "many" of layoffs:

$$D^a_{jt} = \mathbf{1}\{S_{jt} > S^*_j\} \tag{2}$$

where one might consider S^*_j as the threshold beyond which we consider an event to have occurred at time t. S^*_j could be a large fraction, say 30 percent, of the highest employment level observed at the firm ($S^*_j = 0.3 \max_t B_{jt}$). Other baselines could be used, and will impact the observed frequency of such events. This definition has obvious intuitive appeal, but as we will see, also some empirical problems. Foremost among them is the fact that for some firms, high turnover, and thus a high separation reate, is the dominant modus operandi. This will be reflected by firms with repeated events. We will attempt to correct for this by investigating a subgroup of firms for whom this type of event occurs only rarely and infrequently.

Alternatively, a (rapid) net decline in firm employment might be considered:

$$D^b_{jt} = \mathbf{1}\{E_{jt} - B_{jt} > N^*_j\} \tag{3}$$

where again, N_j^* is defined as a function of the firm's previous or overall employment, for instance $N_j^* = 0.3max_t B_{jt}$. Note that generally, D^b implies larger separation flows than D^a unless no hiring ($A_{jt} = 0$) is occurring at these firms.

Note that the definitions above are defined at the quarterly level. However, much of the literature has focussed on events defined over a whole year (Couch and Placzek; 2010; Hildreth et al.; 2009; Jacobson et al.; 1993b; Schoeni and Dardia; 1996). The annual equivalents of the above events are somewhat more tricky to time: does one denote the start of the decline in employment or the increase in separations, or at the end of the decline? How does one define the end of the decline - by using a local minimum, a global minimum?

For starters, consider the following definition:

$$YD_{jt}^b = \mathbf{1}\{E_{jt+4} - B_{jt} > YN_j^*\} \tag{4}$$

This measures year-on-year decline in net employment, and fixing the event at the start of the four-quarter period. This corresponds to the definition used by Hildreth et al. (2009, pg. 9): a displaced "worker was declared part of a mass layoff if the separators firms employment in the year following their departure was 30 percent or more". Note that the decline could continue beyond the initial decline of YN/dB percent, and extend to several periods. Here, YD_t^b denotes the start of a decline that could stretch on for a long time. Thus, this definition can exclude a significant part of the sample that leave in the latter part of a mass layoff if the outflow lasts for more than a quarter. Couch and Placzek (2010) address this by fixing the mass layoff event at the end of the first period at which employment has declined significantly,[3] and including in their sample workers who left the firm one year before, but also after the event. Thus, their event is defined by

$$YD_{jt}^c = \mathbf{1}\{E_{jt} - B_{jt-4} > YN_j^*\} \tag{5}$$

which captures the point in time when the decline for the first time surpassed the cutoff point, but does not necessarily capture the start of the decline. Their worker sample selection criterion based on this event, however, differs from the criterion

[3] "If the individuals change in employment occurred within a year (before or after) of a drop in the firms employment to 30 percent or more below its maximum level prior to 1999, it is considered a displacement due to mass layoff."(Couch and Placzek; 2010, pg. 580)

used by Hildreth et al. (2009), and may or may not capture workers leaving after the employment decline has stopped or been reversed.

Equivalently to D^a, we can also define

$$YD^a_{jt} = \mathbf{1}\{\sum_{k=0}^{3} S_{jt+k} > YS^*_j\} \tag{6}$$

so that we are looking at cumulative outflows that are larger than some baseline measure. We will use again $YS^*_j = 0.3 \max_t B_{jt}$ for consistency, but alternate definitions could make YS^*_j a function of previous separation levels ("2 standard deviations above the average separation rate of the previous x quarters").

Finally, we can consider measures in absolute terms:

$$D^e_{jt} = \mathbf{1}\{S_{jt} > S^{abs}\} \tag{7}$$

where S^{abs} does not depend on firm size. One particular interesting measure is $S^{abs} = 50$, which approximately corresponds to the BLS' definition of a mass layoff for the MLS (http://www.bls.gov/MLS/), except that the MLS is defined in terms of UI claimants, and not simple layoffs.

It is very likely that by one or more of the above measures, a firm experiences a sequence of events defined formally as above. For example, consider $YN^*_j = 0.3 \max_t B_{jt}$, $max_t B_{jt} = B_{j1}$, and a firm with a 50% decline in net employment occurring over 5 quarters starting in $t = 1$ (and no growth or decline otherwise). By this definition, $YD_{b1} = YD_{b2} = YD_{b3} = 1$. Thus, we will also need to define whether we consider the first of a sequence of events to be the event of interest, or the last. In what follows, we will typically identify the "event" as the start of a sequence of such events, when we let multiple events occur. It is not clear how the literature has treated multiple consecutive events.[4]

Finally, the above definitions, if using cutoffs that are dependent on the firm's prior employment levels, may fail to capture firms ceasing to operate. For in-

[4]This is a different problem from when workers experience multiple layoffs (Stevens; 1997).

stance, let "death" be defined by

$$DTH_{jt} = 1 \wedge \sum_{k=1}^{4} M_{j(t+k)} = 0 \tag{8}$$

where the firm's (possibly temporary) death is taken to occur if no positive employment, however transitory, is measured at firm j in the four subsequent periods. However, if $B_{jt-1} < B_j^*$ because the firm had experienced a long decline from its highest employment levels, then the last layoff - the one occurring when the firm shuts down - will not be captured in period t.

Further variation

While the above is already a fairly exhaustive list of possible definitions, additional elements can expand on that rule. For one, the 30% rule may seem arbitrary.[5] Figure 1 shows a hockey stick graph Davis et al. (2006), plotting net employment changes ($\frac{E_{jt}-B_{jt}}{\frac{1}{2}(E_{jt}+B_{jt})}$) against separation rates (green, $\frac{S_{jt}}{\frac{1}{2}(E_{jt}+B_{jt})}$) and accession rates (red, $\frac{A_{jt}}{\frac{1}{2}(E_{jt}+B_{jt})}$). [6] Figure 2 shows the same graph, for yearly measures. While the graph is noisy, it also highlights that there is no obvious break at any particular point in the graph. The noisiness in the left and right regions of the graph comes in part from the scarcity of such observations: very large employment declines or increases are rare events. However, as Figures 3 (for quarterly events) and 4 (for yearly events) illustrate, there is no particular breakpoint visible that would suggest a particular number, 30% or otherwise. We will explore, for some of the definitions, how appropriate the commonly used 30% rule may be.

A further concern is the unit of observation for the "firm" laying off workers. Administrative data rely on administrative identifiers, but different levels of identifiers exist. At the core of the Quarterly Census of Employment and Wages (QCEW) is an establishment (a "reporting unit", called "SEINUNIT" in

[5] Hildreth et al. (2009) report private correspondence with Daniel Sullivan pointing to a histogram-driven approach in defining the breakpoint, but find no corresponding breakpoint in the California sample they used.

[6] Each underlying data point is a firm-quarter observation on each of the three variables, smoothed into 100 bins across the range of the rate of net employment change, a Törnquist index.

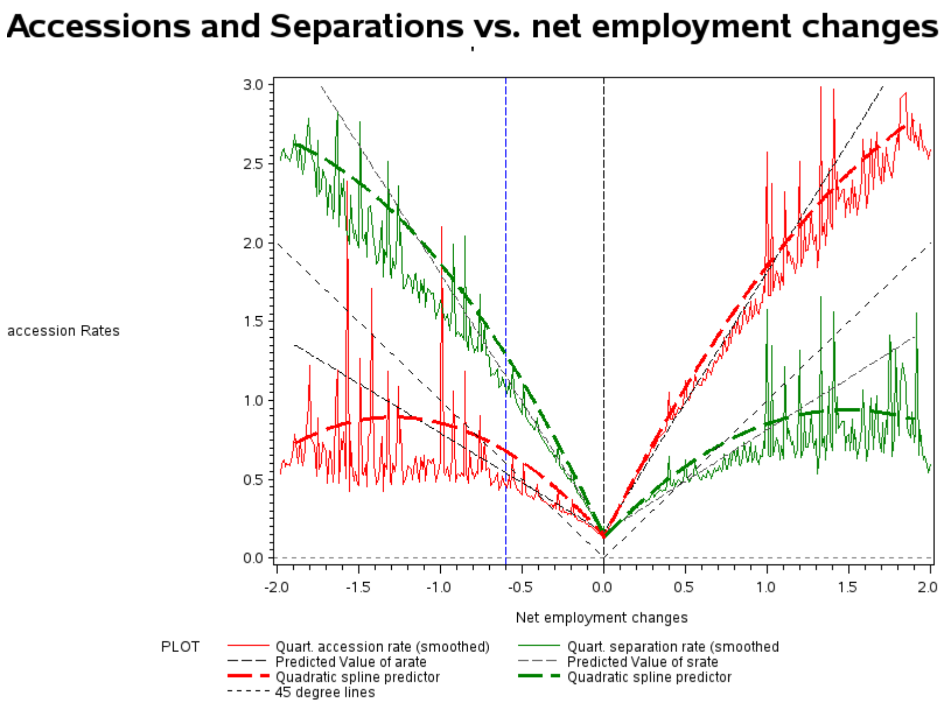

Figure 1: Hockey stick graph for quarterly measures

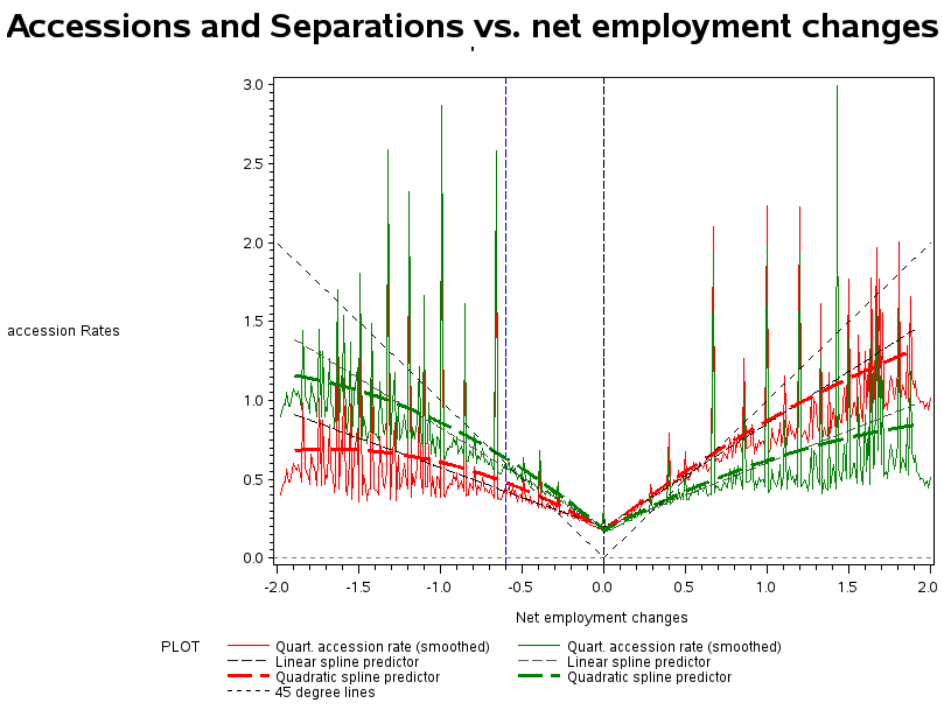

Figure 2: Hockey stick graph for yearly measures

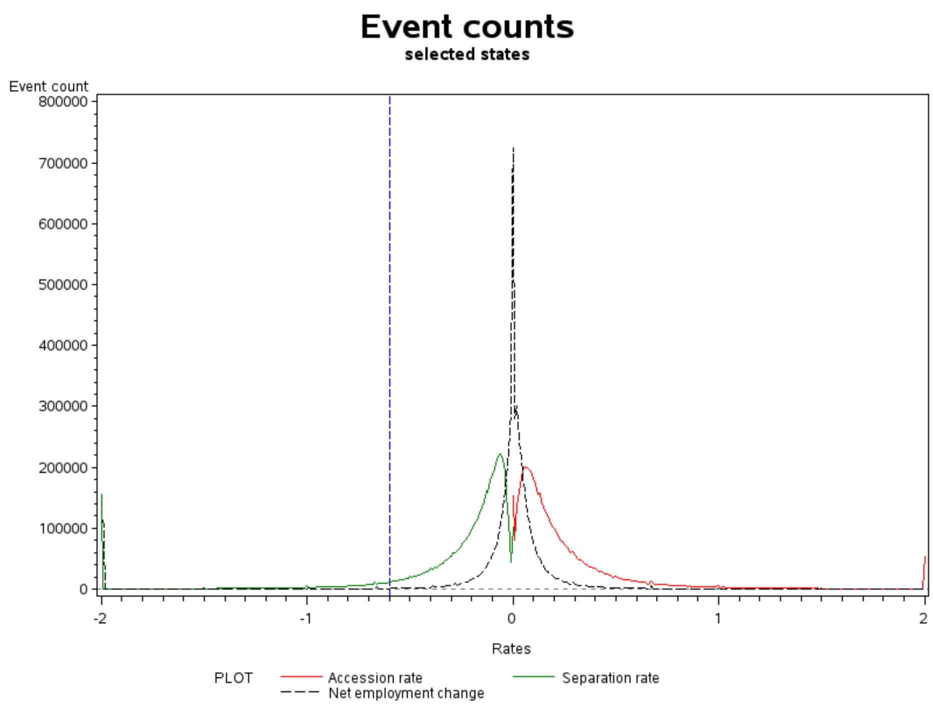

Figure 3: Distribution of quarterly separations, accessions, and net employment changes.

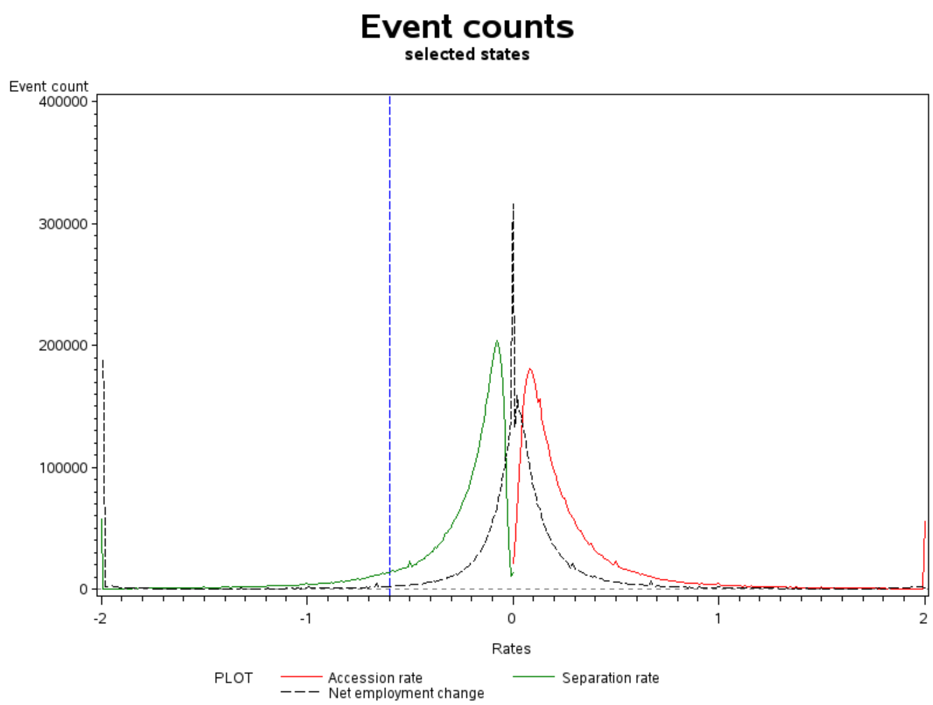

Figure 4: Distribution of yearly separations, accessions, and net employment changes.

the LEHD Infrastructure files) subordinated to a state-level identifier ("SEIN" in the LEHD files), i.e., the concept of a "firm" is limited to the political state boundaries. However, not all firms file their Multiple Worksite Report (MWR). unemployment insurance (UI) wage (worker-level) records are almost never reported at the establishment level, except for the state of Minnesota. Finally, when using federal rather than state-level data, the usual identifier is the Employer Identification Number (EIN), which can identify a firm ("company") in multiple states, without breaking out where each worker's employment occurs (Song et al.; 2009).

Statistics based on worker flows (Definitions (7), (6)) require the use of UI wage records and thus are dependent on the smallest level of firm identity on those files, i.e., a firm identifier at the state level. However, even other definitions (Net employment declines), when linked to the worker (Couch and Placzek; 2010; Hildreth et al.; 2009; Jacobson et al.; 1993b)) rely on the firm-within-state definition in order to perform that link, even if lower-level firm statistics ("establishment") are available. Official MLS numbers, as well as the LBD rely on establishment-level reporting. Finally, some recent papers (Song et al.; 2009) rely on a federal identifier of a firm, localized to the region, to identify the denominator of the breakpoint computation. relies on establishment identifiers.

The level of aggregation matters, since the economic decisions of the firm, in particular for larger, multi-establishment firms, may encompass arbitrary aggregations of establishments across state boundaries. We will attempt to address this partially, by being able to assess the change of incidence in some definitions (YD^b,D^b) when changing the unit of aggregation from SEINUNIT to SEIN to EIN.

Previous work

McKinney and Vilhuber (2006) explore the difference in firm characteristics between firms with a slow or a fast decline, using Definition (3) as the sample selector for firms that survive beyond the sample period, and then considering the length over which the decline occurs. They found that a more than half of firms that experience a single decline in employment do so over a prolonged period of time, and that some distinct differences in the pre-displacement phase exist between these two types of firms.

Dostie et al. (2009) extend the previous analysis by using the annual Definition (4), and again found some differences between firms with an extended decline vs. those with short and rapid declines, even when considering year-on-year declines.

Bowlus and Vilhuber (2002); Lengermann and Vilhuber (2002) explore the composition of worker flows in the period before events according to Definition (7), and find significant differences in the types of outflows before the event even occurs, consistent with conjectures by Farber (2003) and others.

These exploratory papers highlight the need to consider carefully the definition and timing of the event under study.

4 Identifying spurious events

Describe here the two methods of ”cleaning the data”, and baseline statistics (time-series effects of cleaning are reported in the next section)

Longitudinal linking

When computing longitudinal statistics (computations that require data from multiple time-periods), the appropriate longitudinal link is important. The method of ascertaining the appropriate link, however, are often not made explicit, or are simply not performed. Different data sources do this kind of data cleaning in a variety of ways Vilhuber (2007).

LEHD’s Quarterly Workforce Indicators use a flow-based approach to linking firms longitudinally when identifiers are ambiguous or change for spurious reasons Benedetto et al. (2007). Large flows of workers between firms (above 80%) are taken as an indicator that an administrative, but not an economic change occurred at a firm that has two different administrative identifiers. Similar approaches are used elsewhere Hethey and Schmieder (2010). Because it focusses precisely on the kind of flows that would otherwise be flagged as large displacement flows, this approach seems particularly important when studying mass layoffs. For this paper, we have re-estimated the flow links using the original code from the LEHD Production system. This allows us to modulate the strength of the required link [we may not actually do this].

[table here on the importance of SPF]

The LBD, on the other hand, uses a name-and-address matching procedure to link establishments over time Jarmin and Miranda (2002). Thus, the focus there is on the continued presence of the same physical establishment. The impact on large net employment declines is difficult to ascertain (comparison to underlying BR would be necessary).

Accounting for non-filers

Due to the filing process underlying all administrative data, updates to the underlying databases for enforcement purposes may occur over several years, but may not be reflected in the extracts performed for statistical purposes, such as those filed with the Census Bureau for the Business Register (BR) (source of the LBD) and the UI wage records (the key source for the LEHD data). In the LEHD data, the resulting lack of wage records can cause problems for a study of mass layoffs, in particular when wage records for an entire firm (the primary filer) are missing.

We apply a statistical procedure to account for these "holes," by identifying the firm-level earnings patterns of the workers.

[model here for the hole filling]

[table here for the effect of this procedure]

5 A time-series of displacement events

6 Results

7 Conclusion

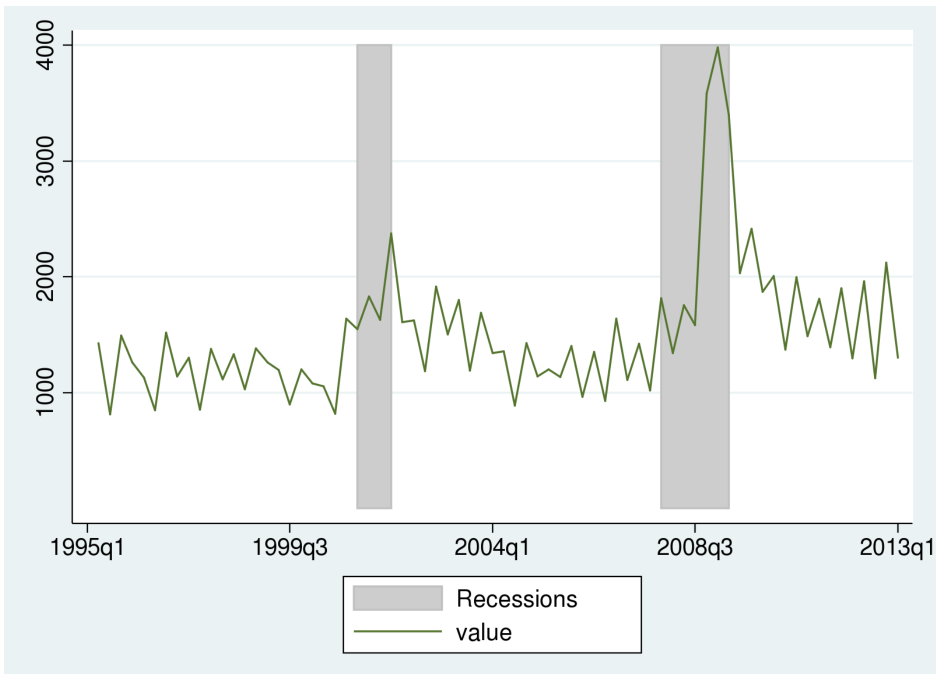

Figure 5: Quarterly number of MLS events

References

Abowd, J. M., Stephens, B. E., Vilhuber, L., Andersson, F., McKinney, K. L., Roemer, M. and Woodcock, S. D. (2009). The LEHD infrastructure files and the creation of the Quarterly Workforce Indicators, The University of Chicago Press for the National Bureau of Economic Research.

Abowd, J. M. and Vilhuber, L. (2011). National estimates of gross employment and job flows from the quarterly workforce indicators with demographic and industry detail, *Journal of Econometrics* **161**: 82–99.

Benedetto, G., Haltiwanger, J., Lane, J. and McKinney, K. (2007). Using worker flows in the analysis of the firm, *Journal of Business and Economic Statistics* **25**(3): 299–313.

Bowlus, A. and Vilhuber, L. (2002). Displaced workers, early leavers, and re-employment wages, *Technical paper TP-2002-18*, LEHD, U.S. Census Bureau.

Couch, K. A. and Placzek, D. W. (2010). Earnings losses of displaced workers revisited, *American Economic Review* **100**(1): 572–589.
URL: *http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.1.572*

Davis, S. J., Faberman, R. J. and Haltiwanger, J. (2006). The flow approach to labor markets: New data sources and micro-macro links, *Journal of Economic Perspectives* **20**(3).
URL: *http://www.nber.org/papers/w12167*

Dostie, B., McKinney, K. L. E. and Vilhuber, L. (2009). Using linked employer-employee data to investigate the speed of adjustment in downsizing firms in canada and the us, *International Census Research Data Center Conference*, Ithaca, NY.

Evans, D. S. and Leighton, L. S. (1995). Retrospective Bias in the Displaced Worker Surveys, *Journal of Human Resources* **30**(2): 386–396.

Farber, H. S. (2003). Job Loss in the United States, 1981-2001, *working paper 9707*, NBER.

Hethey, T. and Schmieder, J. F. (2010). Using worker flows in the analysis of establishment turnover : Evidence from German administrative data, *FDZ Methodenreport 201006_en*, Institute for Employment Research, Nuremberg, Ger-

many.
URL: *http://ideas.repec.org/p/iab/iabfme/201006_en.html*

Hildreth, A. K. G., von Wachter, T. M. and Weber Handwerker, E. (2005). Estimating the true cost of job loss: Evidence using matched data from california, 1991-2000, *mimeo*, Columbia University.

Hildreth, A. K. G., von Wachter, T. M. and Weber Handwerker, E. (2009). Estimating the true cost of job loss: Evidence using matched data from california, 1991-2000, *Discussion Paper 09-14*, Center for Economic Studies.

Jacobson, L., LaLonde, R. and Sullivan, D. (1993a). *The Costs of Worker Dislocation*, W.E. Upjohn Institute, Kalamazoo, Michigan.

Jacobson, L. S., LaLonde, R. J. and Sullivan, D. G. (1993b). Earnings Losses of Displaced Workers, *American Economic Review* **83**(4): 685–709.

Jarmin, R. S. and Miranda, J. (2002). The Longitudinal Business Database, *Center for Economic Studies Working Paper 02-17*, US Census.

Kinney, S. K., Reiter, J. P., Reznek, A. P., Miranda, J., Jarmin, R. S. and Abowd, J. M. (2011). Towards unrestricted public use business microdata: The synthetic longitudinal business database, *International Statistical Review* **79**(3): 362–384.
URL: *http://ideas.repec.org/a/bla/istatr/v79y2011i3p362-384.html*

Lengermann, P. A. and Vilhuber, L. (2002). Abandoning the sinking ship: The composition of worker flows prior to displacement, *Technical paper TP-2002-11*, LEHD, U.S. Census Bureau.

McKinney, K. L. E. and Vilhuber, L. (2006). Using linked employer-employee data to investigate the speed of adjustment in downsizing firms, *Conference on the Analysis of Firms and Employees (CAFE)*, Nuremberg, Germany.

Miranda, J. and Jarmin, R. (2002). The longitudinal business database, *Discussion Paper CES-WP-02-17*, U.S. Census Bureau, Center for Economic Studies.

Schoeni, R. F. and Dardia, M. (1996). Wage losses of displaced workers in the 1990s, *Labor and Population Program Working paper 96-14*, RAND.

Song, J., Manchester, J. and von Wachter, T. (2009). Long-term earnings losses due to mass-layoffs during the 1982 recession: An analysis using U.S. administrative data from 1974 to 2004, *Technical report*, Columbia University.

Stevens, A. H. (1997). Persistent effects of job displacement: The importance of multiple job losses, *Journal of Labor Economics* **15**(1): 165–88.

Sullivan, D. and von Wachter, T. (2009a). Average earnings and long-term mortality: Evidence from administrative data, *American Economic Review* **99**(2): 133–138.

Sullivan, D. and von Wachter, T. (2009b). Job displacement and mortality: An analysis using administrative data, *Quarterly Journal of Economics* **124**(3): 1265–1306.

Vilhuber, L. (2007). Adjusting imperfect data: Overview and case studies, *Working Paper 12977*, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w12977*

A Tables

State	Earliest date		Workers ever present	Firms ever present
	used	available		
CA	1991Q3		34,012,258	2,461,870
FL	1992Q4		18,252,091	1,195,785
ID	1990Q1		1,730,508	98,495
IL	1993Q3	(1990Q1)	13,875,529	738,173
MD	1985Q2		7,235,448	356,187
MN	1994Q3		5,019,721	245,420
MO	1995Q1	(1990Q1)	6,750,881	422,286
MT	1993Q1		1,065,111	73,274
NC	1992Q4	(1991Q1)	9,630,526	441,795
NJ	1996Q1		7,895,006	494,277
OR	1991Q1		4,374,229	269,631
PA	1997Q1	(1991Q1)	11,983,407	616,370
TX	1995Q1		19,411,438	1,070,548
WA	1990Q1		7,438,184	561,270

Table 1: Data availability

Table 2: Industry employment distribution in percent of total employment

NAICS major industry.	National	Selected states
11	1.03	1.54
21	0.55	0.49
22	0.62	0.54
23	5.87	5.98
31-33	15.31	13.49
42	5.38	5.71
44-45	14.06	13.86
48-49	3.77	3.82
51	2.99	3.21
52	5.38	5.34
53	1.90	2.05
54	6.01	6.54
55	1.63	1.57
56	6.63	7.07
61	1.75	1.70
62	12.28	11.99
71	1.66	1.75
72	9.26	9.15
81	3.92	4.20

National statistics derived from Abowd and Vilhuber (2011).

Selected states are listed in the text.

B Acronyms

BLS Bureau of Labor Statistics

BR Business Register

CPS Current Population Survey

EIN (federal) Employer Identification Number

LBD Longitudinal Business Database

LEHD Longitudinal Employer-Household Dynamics

MLS BLSMass Layoff Statistics

MWR Multiple Worksite Report

QCEW Quarterly Census of Employment and Wages, managed by the Bureau of Labor Statistics (BLS)

UI unemployment insurance