CHICAGO JOURNALS

SOLE

The Society of Labor Economists

NORC at the University of Chicago
The University of Chicago

Exploring Differences in Employment between Household and Establishment Data
Author(s): Katharine G. Abraham, John Haltiwanger, Kristin Sandusky, and James R. Spletzer
Source: *Journal of Labor Economics*, Vol. 31, No. 2, The Princeton Data Improvement Initiative (Part 2, April 2013), pp. S129-S172
Published by: The University of Chicago Press on behalf of the Society of Labor Economists and the NORC at the University of Chicago
Stable URL: http://www.jstor.org/stable/10.1086/669062
Accessed: 06/11/2013 11:13

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago, The University of Chicago are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Labor Economics.*

http://www.jstor.org

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Exploring Differences in Employment between Household and Establishment Data

Katharine G. Abraham, *University of Maryland*

John Haltiwanger, *University of Maryland and National Bureau of Economic Research*

Kristin Sandusky, *US Census Bureau*

James R. Spletzer, *US Census Bureau*

Using a large data set that links individual Current Population Survey (CPS) records to employer-reported administrative data, we document substantial discrepancies in basic measures of employment status that persist even after controlling for known definitional differences between the two data sources. We hypothesize that reporting discrepancies should be most prevalent for marginal workers and for marginal or nonstandard jobs, and we find systematic associations between the incidence of reporting discrepancies and observable person and job characteristics that are consistent with this hypothesis. The paper discusses the implications of the reported findings for both micro and macro labor market analysis.

This document reports the results of research and analysis undertaken in part by US Census Bureau and US Bureau of Labor Statistics staff. This work is unofficial and has not undergone the review accorded to official publications of these agencies. All results have been reviewed to ensure that no confidential information is disclosed. The views expressed in the paper are those of the authors and not necessarily those of the Census Bureau or the Bureau of Labor Statistics. The authors

[*Journal of Labor Economics,* 2013, vol. 31, no. 2, pt. 2]
© 2013 by The University of Chicago. All rights reserved.
0734-306X/2013/3102S-0010$10.00

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

I. Background

Labor market analysts rely on data both from household surveys, such as the Current Population Survey (CPS), and employer surveys, such as the Current Employment Statistics (CES) survey. Questions frequently have been raised about whether household-provided and employer-provided data yield a consistent picture of labor market activity. The present study is concerned with discrepancies in the reporting of employment status between individuals and employers. Using CPS records matched to unemployment insurance (UI) wage records for the same people, we document large individual-level discrepancies in employment status between the two sources of data. We also provide evidence of systematic relationships between individuals' personal and job characteristics and the incidence of these discrepancies. These findings have important implications for studies of individual labor market outcomes and for the analysis of aggregate labor market trends.

There is a small literature that uses linked employer-employee microdata to compare household and employer reports of job characteristics. In a seminal study, Mellow and Sider (1983) analyzed data from a January 1977 CPS supplement that collected information from both workers and their employers, with a focus on measurement error in variables such as wages, hours worked, industry, occupation, and union status. Bound and Krueger (1991) used data from a match of CPS respondents to Social Security Administration (SSA) records to study measurement error in earnings data; Bound et al. (1994) used data from the Panel Study of Income Dynamics (PSID) validation study to study measurement error in earnings and other labor market variables; and Roemer (2002) and Abowd and Stinson (2011) used Survey of Income and Program Participation (SIPP) data matched to SSA records to study measurement error in earnings. Each of these studies provides valuable information about discrepancies in the reporting of job characteristics by employed people and their employers, but none addresses discrepancies in the reporting of employment status by households and employers.

Another relevant literature is a set of studies that have compared competing estimates of the effects of government labor market programs on subsequent employment using data from household surveys versus UI

are grateful for financial support from the Princeton Data Improvement Initiative. The current draft is a revised version of a paper circulated and presented at the October 2008 Princeton Data Improvement Initiative Conference. We thank Sasan Bakhtiari, Mary Bowler, David Johnson, Anne Polivka, and Greg Weyland for helpful discussions about data and conceptual issues. The paper has benefited from the comments of Orley Ashenfelter, Kevin Hallock, Lawrence Katz, Alan Krueger, Kristin McCue, Bruce Meyer, and Martha Stinson on earlier drafts. Contact the corresponding author, Katharine G. Abraham, at kabraham@survey.umd.edu.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

wage records. In an early study focused on the employment effects of job training funded under the Job Training Partnership Act (JTPA), Kornfeld and Bloom (1999) found only modest differences between quarterly employment rates computed from survey versus UI data. In a more recent study, Schochet, Burghardt, and McConnell (2008) observed postprogram employment rates that were considerably higher in survey than in UI data, which they interpret as suggesting that UI wage records fail to capture short-duration, informal jobs. Hotz and Scholz (2001) provide a useful overview of this literature, including a discussion of the strengths and weaknesses of survey data versus administrative data for learning about the employment experience of the low-income population. While valuable, all of these studies are by design focused on a narrow group. To our knowledge, no existing study has systematically examined discrepancies in the reporting of employment status in survey data versus administrative data for the population as a whole.

At the macro level, the different behavior of CPS and CES employment series during the late 1990s and early 2000s has attracted considerable attention. CES employment grew markedly faster than CPS employment from 1998 through 2001, opening a large gap between the two series during the economic downturn, and then retreated just as markedly between 2001 and 2003 as the economy improved. A wide variety of hypotheses have been offered about the divergent behavior of CES and CPS employment over the 1998 through 2003 period (see, e.g., Juhn and Potter 1999; Nardone et al. 2003; and Bowler and Morisi 2006), but it remains a puzzle. A possible explanation for the differing behavior of the two series during that period—as well as over the business cycle more generally—lies with the changing importance of what might be termed “marginal” employment and associated discrepancies in the reporting of employment status in household-provided and employer-provided data. The development of this idea is another contribution of our paper.

The paper proceeds as follows. Section II outlines the framework that guides our analysis. Section III describes the linked data set containing CPS information together with UI wage records for the same individuals that we use in our empirical work. We then turn, in Section IV, to an examination of discrepancies between household-reported and employer-reported employment status and the factors that explain those discrepancies. Section V considers the implications of these discrepancies for labor market analysis. Section VI examines the movements over time in aggregate employment estimates derived from household-reported and employer-reported data and the role of reporting discrepancies associated with different personal and job characteristics in explaining the differing behavior of these aggregate series. We end with some concluding observations and suggestions for future research.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

II. Framework for Analysis

We begin by observing that the information about jobs held that a person provides in response to questions on a household survey may or may not agree with the information about jobs held by the same person contained in data based on reports from employers. The possible outcomes with respect to status as a wage and salary worker are illustrated in a two-by-two matrix (table 1). In this matrix, an X represents the number of people in a cell. A person may be recorded as having a wage and salary job in both the household-reported and the employer-reported data (the $X_{H,E}$ group, where H refers to employed in the household-reported data and E refers to employed in the employer-reported data); as having a wage and salary job in neither the household survey nor the employer-reported data (the $X_{NH,NE}$ group, where NH refers to not employed in the household survey and NE refers to not employed in the employer-reported data); as having a wage and salary job in the employer-reported but not the household survey data ($X_{NH,E}$); or as having a wage and salary job in the household survey but not the employer-reported data ($X_{H,NE}$). The number of employed people in the employer-provided data is equal to $X_{NH,E} + X_{H,E}$; the number of employed people in the household-reported data is equal to $X_{H,NE} + X_{H,E}$.

We are most interested in the two off-diagonal cells and, specifically, in how differences in the reporting of employment status by individuals and employers may place people in those cells. Note that there may be other measurement-related reasons for discrepancies in the identification of employment status across data sources—for example, differences in reference periods or Social Security number errors that lead to a job not being counted. These other explanations receive systematic consideration later in the paper; here, however, we focus on the reasons why individuals and employers might report differently about employment.

Consider first the composition of the $X_{NH,E}$ group, those who report no job in the household survey but for whom there is an employer job report. One hypothesis is that marginal workers—by which we mean people who hold jobs but do not view employment as their main activity—should be less likely to report these jobs during a household survey interview.

Table 1
Employment Matrix

		Wage and salary job reported by employer: No	Wage and salary job reported by employer: Yes
Wage and salary job reported by household	No	$X_{NH,NE}$	$X_{NH,E}$
	Yes	$X_{H,NE}$	$X_{H,E}$

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Many people in their teens and early 20s are still in school and seem likely to identify themselves as students rather than workers, even if they hold a job. Older individuals may have concluded long careers and identify themselves as retirees, even if they continue to do some work for pay.[1] We thus would expect people at both the younger and the older ends of the working population to be more likely than prime-age workers to be found in the $X_{NH,E}$ group. Further, whatever the characteristics of the job incumbent, we would expect marginal jobs—by which we mean short-term, low-hour, or low-earnings jobs—to be less likely to be reported in a household survey.[2] To the extent that marginal workers are more likely to hold marginal jobs, these explanations will be at least somewhat overlapping.

Now consider the people found in the $X_{H,NE}$ group, people who report a job in the household survey but for whom no employer job report exists. One reason for a worker's job not to appear in an employer's records is that the worker is being paid *off the books*. Another reason would be that the person is an *independent contractor or consultant* who receives earnings that are reported to the Internal Revenue Service on a Form 1099; such earnings are not considered to be wages and are not subject to unemployment insurance tax. Individuals in any of these categories may regard themselves as employed, since they report to an employer and perhaps work alongside other wage and salary workers, but their earnings do not appear in the UI records. Those found in the $X_{H,NE}$ category might therefore include, for example, both workers with low education working off the books and highly educated people working as independent contractors. To the extent that this sort of nonstandard employment tends to be less stable or less intensive, job discontinuity, low hours of work, and low earnings also may help to explain individuals' presence in the $X_{H,NE}$ cell, although with respect to earnings some independent contractors may be quite highly paid. Working in an industry or occupation in which there are a large number of self-employed workers, suggesting the potential for confusion in the reporting of employment status, also may be predictive of being in the $X_{H,NE}$ group. A final possibility is that some people without a job are embarrassed by this fact and choose to tell the household survey interviewer that they are employed, although having to invent not only the fact of employment but also details about the job should discourage this sort of false reporting.

Another way that household-reported and employer-reported data could differ is that, among those that the two sources agree hold wage and

[1] Ruhm (1990), e.g., finds that many people who report themselves to be retired continue to work in what he terms "bridge jobs."

[2] See Tourangeau, Rips, and Rasinski (2000) for a discussion of the factors that influence the reporting of autobiographical events.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

salary jobs, some could have multiple jobs in one of the data sources but not in the other. A person may be recorded as having a single wage and salary job in both data sets (the $Y_{1H,1E}$ group, where the notation $1H$ refers to one job in the household survey and $1E$ refers to one job in the employer-provided data; see table 2), as having more than one wage and salary job in both data sets (the $Y_{2H,2E}$ group, where the notation $2H$ refers to having two jobs in the household survey and $2E$ refers to having two jobs in the employer-provided data), as having one wage and salary job in the household survey data but more than one in the employer-reported data ($Y_{1H,2E}$), or as having more than one wage and salary job in the household survey data but just one job in the employer-reported data ($Y_{2H,1E}$). If all multiple job holders had exactly two jobs and were classified as employed in both data sets, the total number of jobs recorded in the employer-provided data would equal $X_{NH,E} + X_{H,E} + Y_{1H,2E} + Y_{2H,2E}$ and the total number of jobs in the household-provided data would equal $X_{H,NE} + X_{H,E} + Y_{2H,1E} + Y_{2H,2E}$. Back-of-the-envelope calculations using information on the incidence of multiple job holding indicate that, in the data used for our analysis, the stated sums account for more than 95% of all employment and track movements in actual employment very closely.

For the same reasons that we expect those whose primary jobs are short-lived or involve low hours or low earnings to be more likely to belong to the $X_{NH,E}$ or $X_{H,NE}$ group in the employment matrix, we also expect those whose second jobs have similar characteristics to be more likely to belong to the $Y_{1H,2E}$ or $Y_{2H,1E}$ group in the number of jobs matrix.

III. Data and Measurement

To examine the size and composition of the various cells of the data matrices just described, we require an individual-level data set that includes household-reported and employer-reported information on employment status and number of jobs held, together with information on personal and job characteristics. Our analysis rests on CPS records linked to unemployment insurance wage records for the same people. There has been growing interest in the use of UI wage records for labor market

Table 2
Number of Jobs Matrix

		Number of wage and salary jobs reported by employers: One	Number of wage and salary jobs reported by employers: Two plus
Number of wage and salary jobs reported in the CPS	One	$Y_{1H,1E}$	$Y_{1H,2E}$
	Two plus	$Y_{2H,1E}$	$Y_{2H,2E}$

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

analysis and program evaluation; examples of studies using UI wage record data for labor market analysis include Jacobson, LaLonde, and Sullivan (1993), Schoeni and Dardia (1996), Abowd, McKinney, and Vilhuber (2009), and von Wachter, Handwerker, and Hildreth (2009) on the impact of job displacement; Haltiwanger, Lane, and Spletzer (2007) on the relationship between earnings and worker productivity; Brown, Haltiwanger, and Lane (2006) on the effects of economic turbulence on workers' career paths; Kornfeld and Bloom (1999) on how participation in Job Training Partnership programs affects low-income individuals; and Cancian et al. (1999) on the subsequent earnings of individuals who exited Aid to Families with Dependent Children (AFDC). Given this growing reliance on UI wage data, discrepancies between the CPS and the UI records are of interest in their own right. In addition, since the CES employment estimates are benchmarked to the Quarterly Census of Employment and Wages, which derives from the reports employers must file along with payment of their UI taxes, comparisons of CPS and UI wage record data also should be informative with regard to differences in the behavior of the CPS and CES employment series.

As described in appendix A, the linked individual-level CPS-UI data set we analyze was constructed by the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program with the help of UI wage records provided by the states (see Abowd et al. [2009] for a comprehensive description of the LEHD data infrastructure). CPS records are the starting point for the linked file; information from any UI wage records available for eligible CPS respondents is appended to the CPS records. Each quarter, employers are required to report the quarterly earnings of each of their employees to the agency that administers the UI system in the state where the employer does business. Small agricultural employers typically are excluded from these systems and separate systems cover railroad workers and federal employees. State UI records also generally do not reflect the earnings of unincorporated self-employed workers (see Stevens 2007). CPS respondents may be linked to jobs not only in their state of residence but also to jobs in other states for which UI records are available. Our analysis focuses on what we term in-scope employment—wage and salary employment in the private sector excluding agriculture and private household jobs, plus state and local government employment. Note that, except for the exclusion of federal workers, this definition of in-scope employment is essentially the same as the CES employment definition.

We are able to look at employment in the first quarter of each of the years for which we have data. Protected Identity Keys (PIKs) based on Social Security numbers (SSNs) are used to link individuals' CPS and UI records, but only the March CPS records include these PIKs and they are missing for approximately 20%–30% of March CPS respondents.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Further, because we need to construct a quarterly employment measure from the CPS records for comparison with the quarterly information contained in the UI records, our sample must be restricted to persons who responded to the CPS in all three months of the quarter—January, February, and March. The sample available for analysis consists of approximately 12,000–15,000 individuals per year who are resident in 16 states for which UI wage records are available for the period 1996 through 2003, who responded to a March CPS and have a PIK, and who also responded to the CPS in January and February of the same year. As explained in greater detail in appendix A, we use propensity score methods to reweight our samples to ensure that they represent the March population of the 16 states in each year. For comparability with published CPS employment estimates, we restrict our attention to persons age 16 years and older.

Each quarterly UI wage record identifies an employer, a worker, and the earnings paid by the employer to the worker over the quarter. In the UI data, individuals with at least one report of earnings for an in-scope job held during the first quarter are categorized as employed. Individuals for whom more than one employer provided an in-scope earnings report are categorized as multiple job holders; some of these multiple jobs may have been held simultaneously and others may have been held sequentially, but for our purposes any worker for whom more than one job is reported during the quarter is counted as a multiple job holder. Workers also can be categorized according to the earnings on their highest-paying UI job and, for those holding more than one job over the quarter, the earnings on any other jobs. In addition, using UI earnings records for the previous and the subsequent quarter, we identify workers who hold UI jobs that continue across at least two successive quarters.

In the CPS data, anyone who reports an in-scope main job during the survey reference week in any month of the quarter is categorized as employed. Because only jobs in progress as of the survey reference week are recorded, jobs of very short duration may legitimately not be reported. We cannot always be certain whether a second job reported by a CPS respondent is in-scope or whether a person who is employed in two successive months in the CPS had the same or a different employer. For that reason, we have constructed two measures of whether a person held more than one CPS job during the quarter, one using more restrictive criteria than the other. Additional CPS job variables include measures of the stability of both the main job and any additional jobs, the weekly hours worked on these jobs, and, for the main job, the level of earnings associated with the job. We also capture whether the individual works in an industry and occupation with a high proportion of workers who are self-employed.

In addition to the UI and CPS job variables, we have CPS information on age, education, sex, marital status, race, and whether the individual is

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

foreign born for everyone in the linked sample. We also know whether the individual is the CPS household respondent. Appendix A provides additional details about all of the variables used in the analysis.

In table 3, we compare weighted employment estimates based on our linked sample to estimates based on the full 16-state CPS sample and complete UI wage records database for the same states. Although the main purpose of the table is to help with evaluating the quality of the estimates based on the linked sample, it also reveals that, even though short jobs that do not overlap the survey reference week are not captured in the CPS, the number of people with CPS jobs during the quarter consistently exceeds the number of people with UI jobs. We return to this observation below. Both sets of CPS estimates shown in the top panel of table 3—number of people employed in March and number of people employed in any month during the quarter—are very similar whether we use our more limited

Table 3
Counts of Employed People and Number of Jobs, 1996–2003: CPS Full Sample and UI Universe versus Weighted Linked Sample

A. Number of Employed People, March and First Quarter, CPS Data

	Number Employed in March			Number Employed during First Quarter		
Year	Full Sample	Weighted Linked Sample	Ratio (%)	Full Sample	Weighted Linked Sample	Ratio (%)
1996	53,396,559	53,423,492	100.1	56,380,775	56,333,809	99.9
1997	54,751,688	54,707,231	99.9	57,843,756	57,775,049	99.9
1998	56,620,250	56,552,165	99.9	59,735,205	59,725,372	100.0
1999	57,985,313	57,927,126	99.9	60,720,136	60,652,624	99.9
2000	58,814,505	59,611,816	101.4	62,985,558	62,997,357	100.0
2001	59,689,483	60,688,442	101.7	64,141,704	64,155,358	100.0
2002	59,100,787	60,276,207	102.0	64,021,006	64,051,702	100.0
2003	60,869,144	61,033,141	100.3	64,528,725	64,489,669	99.9
Average	57,653,466	58,027,453	100.6	61,294,608	61,272,618	100.0

B. Number of Employed People and Number of Jobs, First Quarter, UI Data

	Number of People			Number of Jobs		
Year	Universe	Weighted Linked Sample	Ratio (%)	Universe	Weighted Linked Sample	Ratio (%)
1996	50,707,030	49,477,525	97.6	59,880,250	57,968,050	96.8
1997	52,516,172	51,074,482	97.3	62,453,400	60,590,220	97.0
1998	54,479,414	52,154,865	95.7	65,190,010	61,473,350	94.3
1999	55,806,185	53,579,166	96.0	66,784,440	64,217,880	96.2
2000	57,174,841	54,911,649	96.0	68,841,680	66,272,380	96.3
2001	58,378,153	57,268,203	98.1	69,874,490	67,679,090	96.9
2002	57,426,210	56,581,956	98.5	67,373,170	66,278,240	98.4
2003	57,537,936	56,470,108	98.1	67,266,300	66,076,610	98.2
Average	55,503,243	53,939,744	97.2	65,957,968	63,819,478	96.8

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

linked sample or the most inclusive sample for which the estimate in question can be produced. This is a reassuring indication that our estimation weights are performing as intended. In the bottom panel of the table, however, the linked sample estimates of both the number of people holding one or more UI jobs and the total number of UI jobs fall short of the actual numbers calculated from the full UI wage records database. Although the discrepancy is not large, it has implications for our empirical analysis that we consider further below.

IV. Individual-Level Differences in Employment and Multiple-Job-Holding Status in CPS versus UI Records

We turn now to comparisons of employment status (i.e., working versus not working in an in-scope job during the first quarter of the year) and multiple-job-holding status (i.e., holding one versus more than one in-scope job during the first quarter of the year) in individual-level matched CPS-UI data. In addition to looking at weighted estimates of the number of people in the different cells of the employment status and number of jobs matrices described earlier, we also explore the personal and job characteristics that predict membership in the off-diagonal cells of these matrices.

A first-order question is whether status discrepancies for the sample of people in our linked household-reported and employer-reported data are large enough to be of interest. Further, supposing that they are, we would like to understand why these discrepancies occur. Earlier in the paper, we discussed how differences in reporting between households and employers might lead to employment status discrepancies; the sort of reporting differences we hypothesize imply a set of systematic associations between individual and job characteristics and the fact of a discrepancy.

In addition to differences between the reporting behavior of individuals and employers, a number of other measurement factors could cause individuals to appear in the off-diagonal cells of the employment-status and multiple-job-holding matrices. One such measurement factor is that our data set includes UI wage record information for only 16 states, accounting for about half of the US population; to the extent that people in our 16-state sample hold UI jobs in other states for which we do not have UI data, this could lead to our erroneously categorizing some people as not employed in the UI data ($X_{H,NE}$ or $X_{NH,NE}$). As discussed in some detail below, however, we do not believe that this is a serious problem for our findings. Another possible measurement problem is that SSN errors in either the CPS or the UI records could prevent our matching CPS respondents to their UI jobs, causing these individuals to appear in the $X_{H,NE}$ or $X_{NH,NE}$ groups. Again, as described below, the available evidence suggests that this is not a serious problem.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Another potentially important factor is the difference between the reference periods used in collecting CPS and UI data. Whereas individuals are properly counted as employed in the UI data if they received positive earnings at any point during the quarter, individuals in the CPS data records properly are counted as employed only if they were working during the January, February, and/or March survey reference week. This means that the UI data will capture some very short jobs that are legitimately absent from the CPS data, leading to an apparent employment status discrepancy. In our empirical analysis, the variables we use to explain membership in the $X_{NH,E}$ group include an indicator that a UI job was of short duration and an indicator that a UI job had low earnings during the quarter; such jobs are less likely to have been in progress during the CPS reference weeks. In addition to capturing any differential tendency of individuals to overlook short jobs even when they involve work during the survey reference period, the coefficients on these variables also could be capturing the legitimate effects of reference period differences. Another measurement issue is that proxy respondents in the CPS may be poorly informed about the employment status of other household members. In our analysis, we include a variable that captures whether the information for an individual was self-reported or proxy-reported.

One last possible data issue is the difficulty of identifying in-scope second jobs in the CPS. The fact that we obtain very similar results using quite different alternative definitions of CPS multiple job holding, however, suggests this likely is not a major issue (and in any case it would not affect our conclusions concerning discrepancies in employment status).

A. Discrepancies in Employment Status

Table 4 summarizes the distribution of individuals across cells of a two-by-two employment status matrix based on the CPS versus the UI data. The unit of observation here is a person in a particular year. As a reminder, for our purposes, an in-scope job is a nonagricultural private-sector wage and salary job, a state government job, or a local government job. On average over the 8-year period covered by our data, during the first quarter of the year, 49.1% of individuals aged 16 and older are employed in such a job according to both the CPS and the UI data ($X_{H,E}$); 37.1% are not in-scope workers in either the CPS or the UI data ($X_{NH,NE}$);[3] 3.4% are in-scope workers in the UI but not in the CPS ($X_{NH,E}$); and 10.5% are in-scope workers in the CPS but not in the UI ($X_{H,NE}$). Looking at conditional relationships, 6.4% of in-scope UI workers are not in-scope CPS

[3] The upper left cell of the employment status matrix, $X_{NH,NE}$, is the count of individuals not working in either data set. We measure this as the weighted count of individuals in the CPS who are not working and do not have a UI wage record.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Table 4
Discrepancies in Employment Status between CPS and UI Data

	Not In-Scope Worker in UI	In-Scope Worker in UI
Not in-scope worker in CPS	$X_{NH,NE}$	$X_{NH,E}$
Overall share	.371	.034
	(.001)	(.000)
Row share	.917	.083
	(.001)	(.001)
Column share	.779	.064
	(.001)	(.001)
In-scope worker in CPS	$X_{H,NE}$	$X_{H,E}$
Overall share	.105	.491
	(.000)	(.001)
Row share	.176	.824
	(.001)	(.001)
Column share	.221	.936
	(.001)	(.001)

NOTE.—Weighted shares of the CPS-UI overlap sample described in the text. In-scope is defined as wage and salary employment in the private sector excluding agriculture and private household jobs, plus state and local government employment. Pooled data for all years 1996–2003. Standard errors in parentheses.

workers and 17.6% of in-scope CPS workers are not in-scope UI workers. Given the large size of the pooled matched data set, the standard errors of these estimates are low.

The share of CPS workers for whom we can identify no UI job is very large. It seems likely that some of this discrepancy can be explained by the incomplete geographic coverage of our linked data set and by SSN errors in either the CPS data or the UI data, either of which might lead to our failing to identify UI jobs held by CPS respondents. Depending on whether the proportional understatement in $X_{NH,E}$ is larger or smaller than the proportional understatement in $X_{H,E}$, the share of UI workers who appear not to have a CPS job could as a result be either understated or overstated, but this will be a second-order effect. Fortunately, as is shown below, we appear to be missing only a relatively small number of the UI jobs actually held by CPS respondents, and we therefore do not believe that these problems affect any of our qualitative conclusions.

To help understand these issues, recall that the observations in our linked data set are weighted to represent the total population. By construction, the weights reproduce the total CPS population, and they also do a very good job of reproducing total CPS employment. To the extent, however, that we are not successful in locating all of the UI wage records that exist for CPS sample members, estimated UI employment based on the linked data set will fall short of actual UI employment for our 16 states. As shown in table 3, for our 16 states, the number of people with first-quarter UI jobs calculated from the full UI data set (E_{UI}) averages 55.503

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

million over the years 1996–2003; the *estimated* number of UI job holders ($\hat{E}_{UI}$) based on the weighted values from our linked data set averages 53.940 million. The ratio of E_{UI} to $\hat{E}_{UI}$ is 1.029. Given that the 16 states represented in our linked sample and the remaining 35 jurisdictions each account for approximately half of national employment, it seems reasonable to assume that the number of people who live in our 16 states but have UI jobs in one of the other 35 jurisdictions (34 states and the District of Columbia) is approximately equal to the number who live in one of the other 35 jurisdictions but hold UI jobs in one of our 16 states. Under this assumption, we should have found about 2.9% more people with UI jobs than we actually did. Suppose further that UI jobs held by those with and without CPS jobs are equally likely to have been missed. Under these assumptions, accounting for missed UI jobs would not affect our estimate that 6.4% of UI job holders report no CPS job; the share of CPS job holders with no UI job would fall from 17.6% to 15.3%, an effect worth noting, but not one that alters the qualitative story told by table 4.

As a further check on our findings, we replicated our analysis using data only for California, Florida, and Texas, the three largest states in our linked sample. According to data from both the 2000 census and the 2005 American Community Survey (ACS), only about one-half percent of California residents and just over 1% of Florida and Texas residents say they work in a different state, compared to more than 3.5% of people nationwide who say they reside and work in different states. These low percentages reflect not only the size of these states but also the absence of major metropolitan areas that lie near their borders with other states. The results for the three large states are very similar to those shown in table 4. We also examined the quality of the SSNs used to match CPS respondents to their UI records. As described in appendix A, only validated SSNs reported by CPS respondents are retained to form the PIKs used to link the CPS and UI data. Bureau of Labor Statistics (BLS) research (BLS 1997) has found that UI wage records contain only a small proportion of invalid SSNs. All of this additional evidence bolsters our confidence that geographic mismatch and other matching problems can account for no more than a small portion of the many CPS workers with no UI job.

The finding that there are many CPS workers with no UI job is also consistent with evidence based on small samples of employers in selected states that employers misclassify a significant share of their employees as independent contractors or pay them off the books. In an analysis of data from in-depth audits of wage record reports filed in 1987 by a randomly selected sample of 875 Illinois employers, Blakemore et al. (1996) conclude that no UI taxes were paid for 13.6% of workers who should have been classified as employees. About half of these workers were incorrectly treated as independent contractors, with earnings reported on a Form 1099. Despite the increasing attention paid by state authorities to the

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

worker classification issue, a series of more recent reports using data from employer audits that states must conduct to satisfy federal requirements conclude that misclassification of workers as independent contractors continues to be an issue. Carre and Wilson (2004), for example, estimate that about 5% of all workers in Massachusetts were misclassified as independent contractors over the 2001–3 period; Belman and Block (2009) report a similar estimate for Michigan based on data collected from a sample of 894 employers during 2003 and 2004. Extrapolation of this 5% figure to our 16-state sample implies that close to a quarter of individuals in the $X_{H,NE}$ cell in table 4 are independent contractors who should have been reported in the UI wage records. Because state audits generally do not identify off-the-books workers for whom there is no paper trail, these figures almost certainly understate the extent to which employer reports omit workers who should be classified as employees.

A related piece of information is the extent to which people recorded as wage and salary workers in the CPS but absent from the UI wage records—those in the $X_{H,NE}$ cell in table 4—report self-employment income on their federal tax returns as opposed to not showing up in tax data at all. To learn about this, we have merged the records for individuals in the $X_{H,NE}$ off-diagonal in 2002 and 2003 to individual-level records in the Census Bureau's 2002 and 2003 nonemployer database, which originates from IRS Schedule C filings (see Davis et al. [2009] for further discussion). This merging was done at the person level by PIK. In the years we examined, 18% of the people in the $X_{H,NE}$ off-diagonal had a business with positive net receipts in the nonemployer microdata. Whereas these individuals consider themselves to be self-employed for tax purposes, they are classified as wage and salary workers in the CPS.

B. Characteristics of $X_{NH,E}$ and $X_{H,NE}$ Workers

To shed further light on the sources of discrepancies in the reporting of employment, we investigate the personal and job characteristics of those found in the off-diagonal $X_{NH,E}$ and $X_{H,NE}$ cells. The linear probability models in table 5 examine the factors that are associated with holding a UI job but no CPS job (i.e., being in the $X_{NH,E}$ group). (Linear probability model coefficients are reported for ease of interpretation; probit models yield very similar marginal probability effects.) The sample for the estimates in this table includes everyone with a UI job. The simpler model in column 1 includes only demographic variables; the model in column 2 adds UI job characteristics that include whether the individual held any long UI job during the quarter (Any Long Jobs), whether the individual held more than one UI job during the quarter (Two or More UI Jobs), and dummy variables indicating the level of earnings on the individual's primary UI job. Among the demographic variables, being age 65 or older—

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Table 5
Effects of Person and Job Characteristics on the Probability That a UI Worker Is Not a CPS Worker ($X_{NH,E}$)

	Mean	(1)	(2)
Age 16–24	.173	.0511**	−.0205**
		(.0033)	(.0032)
Age 25–34	.249	.0026	−.0047
		(.0025)	(.0024)
Age 55–64	.093	.0257**	.0190**
		(.0036)	(.0034)
Age 65 Plus	.025	.1442**	.0879**
		(.0066)	(.0062)
Less than High School	.138	.0474**	.0066*
		(.0034)	(.0032)
Some College	.299	−.0028	−.0003
		(.0026)	(.0024)
College Graduate	.187	−.0107**	.0017
		(.0030)	(.0029)
More than College	.078	−.0003	.0099*
		(.0041)	(.0040)
Male	.516	−.0088**	.0055**
		(.0020)	(.0019)
Married	.438	−.0012	−.0094**
		(.0024)	(.0022)
Black	.094	.0277**	.0215**
		(.0035)	(.0033)
Other Nonwhite	.059	.0161**	.0088*
		(.0047)	(.0043)
Foreign Born	.151	.0149**	.0259**
		(.0031)	(.0029)
Nonproxy Interview	.307	.0044	.0063**
		(.0024)	(.0022)
Any Long Jobs	.974	. . .	−.2013**
			(.0061)
Two or More UI Jobs	.149	. . .	−.0175**
			(.0027)
Quarter UI Earnings < $1K	.100	. . .	.2803**
			(.0036)
$1K ≤ Quarter UI Earnings < $2.5K	.126	. . .	.0571**
			(.0031)
$12.5K ≤ Quarter UI Earnings < $25K	.126	. . .	−.0118**
			(.0031)
$25K ≤ Quarter UI Earnings	.031	. . .	−.0001
			(.0058)
Observations	. . .	56,027	56,027
R-squared	. . .	.024	.160

NOTE.—Coefficients obtained from linear probability regressions using pooled data for all years 1996–2003 for respondents aged 16 and older. Year dummies included in both models. Weighted means of explanatory variables shown in first column. Standard errors in parentheses.
* Significant at the 5% level.
** Significant at the 1% level.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

one of our proxies for being a marginal worker—has the largest effect, raising the probability of being in the $X_{NH,E}$ category by 14 percentage points in the first model (very similar to the effect in a model that includes only the age variables, shown in app. B, available in the online version of *Journal of Labor Economics*) and by about 9 percentage points after adding UI job characteristics to the model. This is consistent with our expectation that older workers may think of themselves as retirees rather than workers even if they hold a job and, thus, be less likely to report themselves as employed. The variable for the youngest age group—another proxy for being a marginal worker—also assumes a positive coefficient when only demographic effects are included, but this effect becomes negative when job controls are added. The switching of the sign of the effect for young workers is indicative of the strong positive correlation between age and earnings. Beyond the age effects, the other demographic effects tend to be modest, although we find that nonwhite and foreign-born workers are more likely to have UI employment but to report no CPS job.

UI job characteristic variables are added in the second column of table 5.[4] The variable Any Long Jobs measures whether the person held at least one UI job that began in the previous quarter or continued into the subsequent quarter; the large negative coefficient on this variable suggests that short-duration UI jobs are less likely to be reported in the CPS. This could be either because of misreporting or because the UI jobs were not in progress during the CPS reference weeks.[5] The large positive coefficient on the variable indicating that a person's UI earnings for the quarter were less than \$1,000 (Quarter UI Earn < \$1K) reinforces this short-duration finding and the associated conclusion that marginal jobs may be less likely to be reported in the CPS. As an illustration, the coefficients in the second column of table 5 imply that 59% of workers over age 65 who had a short-duration UI job and earned less than \$1,000 during the quarter would not report a job in the CPS, compared to an average of 6.4% for all UI job holders.[6]

Table 6 reports similar models for the probability that a CPS worker does not hold a UI job (i.e., belongs to the $X_{H,NE}$ group). The sample for these models includes everyone with a CPS job. The first model includes

[4] Because the sample used for the models reported in table 5 consists of people with a UI job, not all of whom have a CPS job, only job characteristics based on the UI job records can be included as explanatory variables.

[5] Temporary holiday jobs related to the Christmas season that spill over into the first few days of the new year are an example of jobs for which there might be a first-quarter UI record but that might legitimately not be reported in the CPS first quarter interviews.

[6] Mean values of all other personal and job characteristics were assumed in these calculations.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Table 6
Effects of Person and Job Characteristics on the Probability That a CPS Worker Is Not a UI Worker ($X_{H,NE}$)

	Mean	(1)	(2)
Age 16–24	.164	.0043	−.0401**
		(.0050)	(.0051)
Age 25–34	.247	−.0167**	−.0179**
		(.0037)	(.0037)
Age 55–64	.095	.0336**	.0227**
		(.0053)	(.0053)
Age 65 Plus	.026	.1423**	.0792**
		(.0096)	(.0096)
Less than High School	.135	.0345**	.0075
		(.0051)	(.0051)
Some College	.298	.0052	.0046
		(.0039)	(.0039)
College Graduate	.191	.0119**	.0152**
		(.0044)	(.0045)
More than College	.083	.0521**	.0527**
		(.0059)	(.0061)
Male	.525	.0170**	.0242**
		(.0030)	(.0031)
Married	.430	−.0040	−.0085*
		(.0035)	(.0035)
Black	.089	−.0164**	−.0173**
		(.0053)	(.0053)
Other Nonwhite	.060	.0124	.0046
		(.0068)	(.0068)
Foreign Born	.155	.0360**	.0420**
		(.0046)	(.0045)
Nonproxy Interview	.305	−.0061	−.0023
		(.0035)	(.0035)
Work Discontinuity	.109	. . .	.0735**
			(.0055)
Probability of Being a Contractor	.037	. . .	.0723**
			(.0079)
Any Full Time Jobs	.818	. . .	−.0272**
			(.0057)
CPS Earnings under \$1K	.028	. . .	.1718**
			(.0098)
\$1K ≤ CPS Earnings < \$2.5K	.091	. . .	.0602**
			(.0061)
\$12.5K ≤ CPS Earnings < \$25K	.116	. . .	−.0092
			(.0050)
\$25K ≤ CPS Earnings	.019	. . .	.0311**
			(.0112)
CPS Earnings Missing	.106	. . .	.1710**
			(.0054)
Observations	. . .	63,901	63,901
R-squared	. . .	.010	.045

NOTE.—Coefficients obtained from linear probability regressions using pooled data for all years 1996–2003 for respondents aged 16 and older. Year dummies included in both models. Weighted means of explanatory variables shown in first column. Standard errors in parentheses.
* Significant at the 5% level.
** Significant at the 1% level.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

only demographic variables (identical to the first column of table 5); the model in the second column adds job measures based on the CPS data.[7] These include whether the individual had a work discontinuity (i.e., was not employed at the time of one or two of the monthly CPS interviews during the quarter, labeled in table 6 as Work Discontinuity); whether the respondent works in an occupation and industry with a high percentage of self-employed workers (Probability of Being a Contractor); whether any of the CPS jobs held during the quarter were full-time jobs (Any Full Time Jobs); and a set of earnings dummies based on earnings in the main job as reported in the individual's outgoing rotation month (March or April). Again, those over age 65 are especially likely to be found in the off-diagonal cell; the 14 percentage point effect in the model containing other demographic variables is very similar to that in a model that includes only age variables (shown in online app. B) and drops to about 8 percentage points when CPS job characteristic variables are included in the model. Other demographic characteristics that raise the probability of membership in the $X_{H,NE}$ cell are being male, nonblack, or foreign born and having either a low or a very high level of education. These effects are not sensitive to the addition of other controls. Some of these demographic characteristics seem more likely to be associated with holding an off-the-books job (e.g., being foreign born and having a low level of education), while others seem more likely to be associated with working as an independent contractor (e.g., being older, being male, and having a very high level of education). Introducing CPS job characteristic variables adds considerable explanatory power to the model. CPS workers who were not employed at the time of one or two of the monthly CPS interviews during the quarter, who are in an industry and occupation with a high proportion of self-employed workers, who had a job with very low earnings, or who reported no earnings in the CPS outgoing rotation month are substantially more likely not to have a UI job. The CPS job characteristic coefficients in this model seem primarily to reflect the presence of off-the-books jobs, in that they imply that it is workers with intermittent and lower paying jobs who are more likely to appear in the $X_{H,NE}$ cell. As an illustration, the coefficients in the second column of table 6 imply that 57% of CPS job holders over age 65 who worked in an occupation and industry with many self-employed workers, reported employment in some but not all months of the quarter, and had earnings for the quarter under $1,000 would not report a UI job, compared to an average of 17.6% for everyone who reports a CPS job.[8]

[7] Because the sample used for the models reported in table 6 consists of people with a CPS job, not all of whom have a UI job, only job characteristics based on the CPS data can be included as explanatory variables.

[8] Mean values of all other personal and job characteristics were assumed in these calculations.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

One empirical finding common to both table 5 and table 6 is that adding job characteristics to the classification discrepancy model dramatically increases the R-squared, relative to a specification with demographic coefficients only. Additional analyses reported in online appendix B show that the earnings variables drive much of this increase in explanatory power. When adding just the earnings variables, the R-squared in table 5 jumps from .024 to .143, and the R-squared in table 6 jumps from .010 to .041. In both cases, the coefficient on the low earnings dummy is very large.[9] The UI variable Any Long Jobs also has a large coefficient and adds explanatory power in table 5 (the R-squared increases from .024 to .064 with just this variable); the CPS variable Work Discontinuity has a large coefficient and adds explanatory power in table 6 (the R-squared increases from .010 to .022 when adding just this variable). The Any Long Jobs and Work Discontinuity variables are alternative measures of short-duration jobs.

Additional specifications of the models in tables 5 and 6 that control for industry also are reported in online appendix B; the inclusion of industry dummies generally has little effect on the coefficients for other variables, with the notable exception that, in table 6, the coefficient on the variable that captures the probability of being a contractor falls (from 0.0723 to 0.0495). This decrease is not surprising, given that the variable in question is created using industry and occupation data.[10]

In sum, tables 5 and 6 provide evidence that, in the cross section, the UI workers who report no CPS employment have person and job characteristics consistent with their being marginal workers or holding marginal jobs, while the CPS workers for whom no UI employment can be identified have person and job characteristics consistent with their holding off-the-books jobs or being employed as an independent contractor or consultant.

C. Discrepancies in Multiple-Job-Holding Status

In table 7, we explore discrepancies in multiple job status, looking at employed individuals who report more than one job in the UI but not the CPS data (the $Y_{1H,2E}$ group) or who report more than one job in the CPS but not the UI data (the $Y_{2H,1E}$ group). Because of the difficulty of identifying workers in the CPS who held more than one job during the

[9] In table 5, a person with quarterly UI earnings below $1,000 is 28% more likely not to have a CPS job (relative to someone in the omitted group of $2,500–$12,500). In table 6, a person with quarterly CPS earnings below $1,000 is 17% more likely not to have a UI job (relative to someone in the omitted group of $2,500–$12,500).

[10] We also report results including CPS occupation as well as industry dummies in online appendix B. Adding occupation dummies also has a relatively modest impact on the results, with R-squared increasing from .045 without industry and occupation dummies to .056 with just industry dummies and to .061 with both industry and occupation dummies included.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Table 7
Discrepancies in Multiple Job Status between CPS and UI Data

	One In-Scope Job in UI	Two or More In-Scope Jobs in UI
One in-scope job in CPS	$Y_{1H,1E}$	$Y_{1H,2E}$
Overall share	.813	.104
	(.001)	(.001)
Row share	.887	.113
	(.001)	(.001)
Column share	.956	.692
	(.000)	(.005)
Two or more in-scope jobs in CPS	$Y_{2H,1E}$	$Y_{2H,2E}$
Overall share	.037	.046
	(.000)	(.000)
Row share	.446	.554
	(.007)	(.007)
Column share	.044	.308
	(.000)	(.005)

NOTE.—Weighted shares of persons in the CPS-UI overlap sample described in the text who both sources agree have an in-scope job. Pooled data for all years 1996–2003. Estimates based on the more restrictive set of criteria described in the data appendix to identify persons holding multiple jobs in the CPS. Standard errors in parentheses.

quarter, it is less straightforward to identify discrepancies along the single-versus-multiple-jobs dimension than along the whether-employed dimension and, as described in appendix A, we have defined alternate indicators of whether a person held multiple CPS jobs to test whether any conclusions drawn are sensitive to the precise definition employed. The results reported in the text make use of the more restrictive of the two sets of criteria for identifying the multiple job group in the CPS, but our main findings are not affected by that choice. The at-risk group in table 7 is individuals who hold both a CPS and a UI job, the $X_{H,E}$ group from table 4. Some 81.3% of such workers have just one job both in the CPS and in the UI data, 4.6% have two or more jobs both in the CPS and in the UI, 10.4% have two or more UI jobs but only one CPS job, and 3.7% have two or more CPS jobs but only one UI job. Conditional percentages are again instructive. Some 69.2% of workers with two or more in-scope UI jobs have only one in-scope CPS job. Conversely, conditional on having two or more in-scope CPS jobs, 44.6% of workers have only one in-scope UI job. The standard errors for all of these estimates are small. The magnitude of the discrepancies between the two data sources with regard to multiple job holding is striking.

D. Characteristics of $Y_{1H,2E}$ and $Y_{2H,1E}$ Workers

Table 8 reports the results of models that seek to identify the factors associated with holding more than one UI job but only a single CPS job (i.e.,

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Table 8
Effects of Person and Job Characteristics on the Probability That a UI Multiple Job Holder Has a Single CPS Job ($Y_{1H,2E}$)

	Mean	(1)	(2)
Age 16–24	.231	−.0063	.0023
		(.0161)	(.0161)
Age 25–34	.276	.0176	.0313*
		(.0133)	(.0131)
Age 55 to 64	.066	.0269	.0350
		(.0223)	(.0219)
Age 65 Plus	.012	.0316	.0621
		(.0492)	(.0483)
Less than High School	.130	.0143	.0125
		(.0183)	(.0179)
Some College	.328	−.0482**	−.0508**
		(.0137)	(.0134)
College Graduate	.187	−.0344*	−.0595**
		(.0161)	(.016)
More than College	.074	−.0288	−.0822**
		(.0224)	(.0225)
Male	.495	.0171	.0069
		(.0108)	(.0107)
Married	.509	−.0383**	−.0278*
		(.0122)	(.0120)
Black	.131	.0857**	.0972**
		(.0160)	(.0156)
Other Nonwhite	.064	.0044	.0027
		(.0238)	(.0233)
Foreign Born	.155	.0487**	.0574**
		(.0164)	(.0161)
Nonproxy Interview	.292	−.0186	−.0255*
		(.0125)	(.0122)
Any Long Second Jobs	.655	. . .	−.1103**
			(.0144)
Three or More UI Jobs	.162	. . .	−.0016
			(.0115)
UI < $1K (Second Job)	.444	. . .	−.0136
			(.0139)
$1K ≤ UI < $2.5K (Second Job)	.248	. . .	−.1572**
			(.0150)
$12.5K ≤ UI < $25K (Second Job)	.048	. . .	.2097**
			(.0264)
$25K ≤ UI (Second Job)	.011	. . .	.2499**
			(.052)
Observations	. . .	7,442	7,442
R-squared	. . .	.015	.058

NOTE.—Coefficients obtained from linear probability regressions using pooled data for all years 1996 to 2003 for respondents aged 16 and older. Year dummies included in both models and the more restrictive set of criteria described in the data appendix for identifying CPS multiple job holders applied. Weighted means of explanatory variables shown in first column. Standard errors in parentheses.
* Significant at the 5% level.
** Significant at the 1% level.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

being in the $Y_{1H,2E}$ group). The model in the first column contains only demographic factors; characteristics of UI jobs beyond the first are added in the second column. For those with multiple UI jobs, the likelihood of having only a single CPS job is higher for black and foreign-born workers and lower for more educated and married workers. Adding job characteristic variables, perhaps surprisingly, workers with the highest-paying second UI jobs are much more likely than those with lower-paying second jobs to be $Y_{1H,2E}$ workers. One possible interpretation is that some highly compensated individuals with multiple sources of earnings think of themselves as having a single job. Actors, skilled construction trades workers, college faculty, or even doctors and lawyers all are examples of people who might fall into this category. Results very similar to those reported in table 8 are obtained when the less restrictive criteria are used to identify multiple job holders in the CPS.

Table 9 reports results for models of the probability that people with more than one CPS job hold only a single UI job (i.e., are found in the $Y_{2H,1E}$ cell). Personal characteristics that raise this probability include being highly educated, age 55 or older (although the effect is not statistically significant for those age 65 and older), or male. The addition of job characteristics has little effect on the demographic variable coefficients but raises the explanatory power of the model. Those who hold two simultaneous CPS jobs in one or more months, or work at least 16 hours a week on their second job(s) are much less likely to be $Y_{2H,1E}$ workers. Much the same coefficient patterns hold when the more expansive definition of CPS multiple job holders is applied.

The results for multiple job holders largely mimic those found for employment status—worker and job characteristics consistent with marginal jobs on the one hand and informal or nonstandard jobs on the other hand contribute substantially and significantly to the discrepancies in multiple-job-holder status between the UI and the CPS data.

V. Implications for Employment Rates and Labor Market Analysis

We have established that having a UI job but not a CPS job or vice versa is not random but rather is highly systematic with respect to both demographic and job characteristics. The characteristics that predict membership in the $X_{H,NE}$ cell (CPS job but no UI job) and $X_{NH,E}$ cell (UI job but no CPS job) are broadly similar but not identical; further, the $X_{H,NE}$ group is much larger than the $X_{NH,E}$ group. As a result, there are also systematic differences in the characteristics of CPS workers versus UI workers. Table 10 reports the marginal effects of personal characteristics on the probability that a person in the matched data set is employed, first using the UI employment measure and then using the CPS employment mea-

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Table 9
Effects of Person and Job Characteristics on the Probability That a CPS Multiple Job Holder Has a Single UI Job ($Y_{2H,1E}$)

	Mean	(1)	(2)
Age 16–24	.218	−.0786**	−.1168**
		(.0228)	(.0225)
Age 25–34	.268	−.0139	−.0331
		(.0187)	(.0183)
Age 55 to 64	.074	.0873**	.0868**
		(.0297)	(.0290)
Age 65 Plus	.013	.0426	.0124
		(.0679)	(.0664)
Less than High School	.117	.0310	.0143
		(.0271)	(.0265)
Some College	.339	−.0099	.0036
		(.0197)	(.0194)
College Graduate	.203	.0368	.0450*
		(.0227)	(.0223)
More than College	.091	.0969**	.1167**
		(.0294)	(.0291)
Male	.515	.0538**	.0468**
		(.0153)	(.0150)
Married	.507	−.0252	−.0178
		(.0172)	(.0169)
Black	.094	−.0365	−.0378
		(.0258)	(.0253)
Other Nonwhite	.060	−.0337	−.0367
		(.0341)	(.0333)
Foreign Born	.141	.0450	.0310
		(.0238)	(.0234)
Nonproxy Interview	.291	−.0571**	−.0403*
		(.0178)	(.0175)
Simultaneous Multiple Jobs	.532	. . .	−.1020**
			(.0210)
Multiple Jobs All 3 Months	.275	. . .	−.0601**
			(.0220)
16+ Hours per Week Second Job(s)	.227	. . .	−.1236**
			(.0220)
Observations	. . .	4,352	4,352
R-squared	. . .	.025	.070

NOTE.—Coefficients obtained from linear probability regressions using pooled data for all years 1996–2003 for respondents aged 16 and older. Year dummies included in both models and the more restrictive set of criteria described in the data appendix for identifying CPS multiple job holders applied. Weighted means of explanatory variables shown in first column. Standard errors in parentheses.
* Significant at the 5% level.
** Significant at the 1% level.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Table 10
Effects of Person Characteristics on the Probability of UI Employment and CPS Employment

	Mean	UI Employment	CPS Employment
Age 16–24	.163	−.042**	−.082**
		(.005)	(.004)
Age 25–34	.190	.052**	.042**
		(.004)	(.004)
Age 55–64	.111	−.182**	−.202**
		(.005)	(.004)
Age 65 Plus	.155	−.514**	−.585**
		(.004)	(.004)
Less than High School	.206	−.143**	−.162**
		(.004)	(.004)
Some College	.269	.024**	.033**
		(.004)	(.003)
College Graduate	.158	.047**	.071**
		(.004)	(.004)
More than College	.070	.038**	.088**
		(.006)	(.006)
Male	.487	.042**	.066**
		(.003)	(.003)
Married	.442	.019**	.020**
		(.003)	(.003)
Black	.091	−.002	−.032**
		(.005)	(.005)
Other Nonwhite	.061	−.058**	−.064**
		(.006)	(.006)
Foreign Born	.160	−.019**	−.007
		(.004)	(.004)
Nonproxy Interview	.334	−.027**	−.037**
		(.003)	(.003)
Observations	. . .	107,731	107,731
R-squared	. . .	.18	.25

NOTE.—Coefficients obtained from linear probability regressions using pooled data for all years 1996–2003 for respondents aged 16 and older. Year dummies included in both models. Weighted means of explanatory variables shown in first column. Standard errors in parentheses.
* Significant at the 5% level.
** Significant at the 1% level.

sure.[11] In both cases, prime age, educated, male, married, white, and US-born individuals are more likely to be employed, but the estimated effects differ in magnitude. The positive marginal effect of having an advanced degree on employment, for example, is about 5 percentage points larger in the CPS than in the UI data, and the negative effect of having less than a

[11] Table 10 is an unconditional regression with an unweighted sample size of 107,731. In contrast, the sample in table 5 is conditional on having a UI job (with an unweighted sample size of 56,027), and the sample in table 6 is conditional on having a CPS job (with an unweighted sample size of 63,901).

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

high school education also is more pronounced (by about 2 percentage points) in the CPS than in the UI data.

These findings translate into differences in probabilities of being employed in the CPS and the UI that can vary markedly. To take two extreme examples, for a black female aged 16–24 with less than a high school education (and other characteristics evaluated at the mean of their distribution in the data), the predicted probability of employment is very similar across the two data sources (41% based on the CPS model versus 42% based on the UI model). On the other hand, for a white male aged 35–54 with an advanced degree (and mean values for other characteristics), the predicted probability of employment is 84% based on the CPS model but just 69% based on the UI model. Depending on the question and the groups of interest, the conclusions to be drawn from a labor market analysis could look quite different depending on whether employer-provided or individually reported information was used to determine employment status.

To illustrate the potential importance of the employment rate discrepancies we have detected for labor market analysis, consider the question of how worker displacement affects subsequent employment, a topic that has been a focus of research using UI administrative records (see, e.g., Jacobson et al. 1993; Schoeni and Dardia 1996; Abowd, McKinney, and Vilhuber 2009; von Wachter et al. 2009) and of a parallel stream of research based on data from the CPS Displaced Worker Supplement (DWS; see, e.g., Farber 1993, 1997, 2005). While a thorough analysis would be beyond the scope of this paper, we have constructed a few simple tabulations to illustrate the potential importance of the discrepancies in the reporting of employment status that have been our focus for conclusions about the consequences of worker displacement. Specifically, we link person-level records from the 2002 DWS to our matched CPS-UI data file and then calculate both CPS and UI employment rates for the resulting sample of displaced workers.[12]

The first row of table 11 shows that, among individuals identified as displaced in the three years preceding the 2002 DWS, current employment rates are higher in the CPS than in the UI data. This is not terribly surprising given what we already know about employment rates in the two data sources. More interesting is how this discrepancy differs by time since reported displacement. A typical finding in the displaced worker literature using UI data is that the earnings of displaced workers fall sharply in the period immediately following their displacement and then

[12] Although in principle all workers in our linked sample who were age 20 and older in 2002 should have been asked the DWS questions, some provided no response, and we have re-weighted the linked DWS-CPS-UI file, using the same approach we use for the original matched sample, to account for DWS nonresponse.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Table 11
Probability of UI and CPS Employment for Recently Displaced Workers, First Quarter 2002

	CPS Employment Rate	UI Employment Rate
All workers displaced 1999–2001	67.1	60.5
	(2.3)	(2.8)
Workers displaced 2001	55.3	54.3
	(4.0)	(4.0)
Workers displaced 1999 or 2000	80.3	67.0
	(3.3)	(3.9)

NOTE.—Weighted employment rates for individuals who reported in the January 2002 Displaced Worker Supplement (DWS) that they were displaced from a previous job in 1999, 2000, or 2001 and met criteria for inclusion in linked CPS-UI sample. Standard errors in parentheses.

begin to rebound in the following years. Moreover, the adverse immediate impact of displacement is very much associated with low employment rates (and high turnover of those who do find a job). As a simple exercise along these lines, the second and third rows of table 11 report employment rates as of the first quarter of 2002 for those whose displacement occurred in the immediate prior year (in 2001) and those whose displacement occurred 2–3 years earlier (in 1999 or 2000). Both the CPS and the UI employment rates exhibit the pattern typically found in the literature: employment rates are lower in the year just after displacement than 2–3 years later. Interestingly, however, the magnitudes of the differences in employment rates over time differ substantially across the two data sources. The CPS data show those displaced more than a year ago to have an employment rate that is about 25 percentage points higher than those displaced in the last year (the difference between 80.3 and 55.3). In contrast, the UI employment rate is only about 13 percentage points higher for those displaced more than one year ago relative to those displaced in the last year. Further work is clearly needed here, but these results suggest that conclusions about the impact of displacement over time may vary substantially depending on whether one tracks employment using household-reported survey data or employer-reported administrative data.

VI. Aggregate Time Series Patterns in the Discrepancy between Household and Employer Based Employment

We have established that there are large individual-level discrepancies in employment status and number of jobs recorded between the CPS and UI data. Moreover, these discrepancies are not random but are systematically related to person and job characteristics, with potentially important implications for disaggregated labor market analyses. In this section, we consider the aggregate time series implications of these individual-level findings.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

A. Are Aggregate Discrepancies Cyclical? Evidence and Conceptual Underpinnings

The different time series behavior of CPS and CES employment over the period from 1998 through 2003 that we referred to at the start of the paper is displayed in figure 1. The official CPS employment measure includes some workers who are not counted in the CES—self-employed individuals, unpaid family workers, agricultural and related workers, private household workers, and workers with a job but not at work during the survey reference week—and, because it is a person count rather than a job count, does not reflect that some individuals hold multiple jobs (see Bowler and Morisi [2006] for details). Even after the CPS data have been adjusted by the Bureau of Labor Statistics to be more comparable in scope and concept to the CES data, CES employment grew considerably more rapidly than CPS employment during the last few years of the expansion that ended in 2001 and then fell sharply from 2001 to 2003 while CPS employment remained more level.

While the period from 1998 through 2003 is unusual in the degree to which CES and CPS employment growth diverged, the data displayed in figure 2 suggest that divergence between CES and CPS employment is a cyclical phenomenon. The figure displays a 60-year history of the ratio of CES employment to CPS nonagricultural wage and salary employment. In contrast to the adjusted CPS series displayed in figure 1, the CPS data

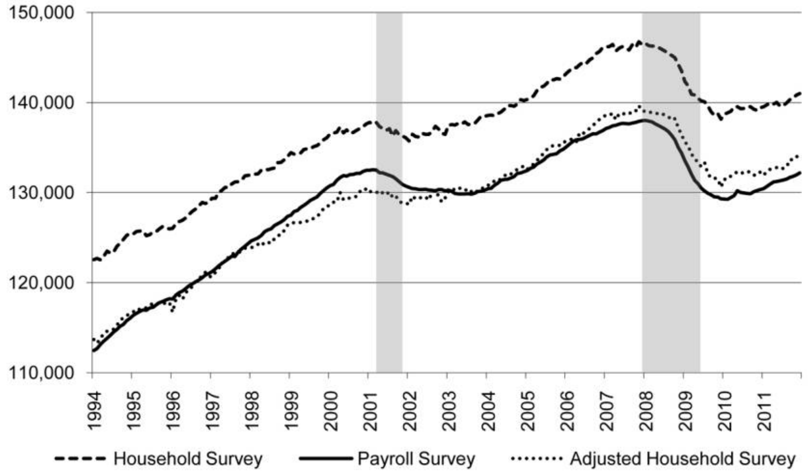

FIG. 1.—Household and payroll survey employment, in thousands, seasonally adjusted, 1994–2011. Both household series are smoothed for population control revisions. The employment concept for the adjusted household series is similar to the payroll survey concept. Source: www.bls.gov.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

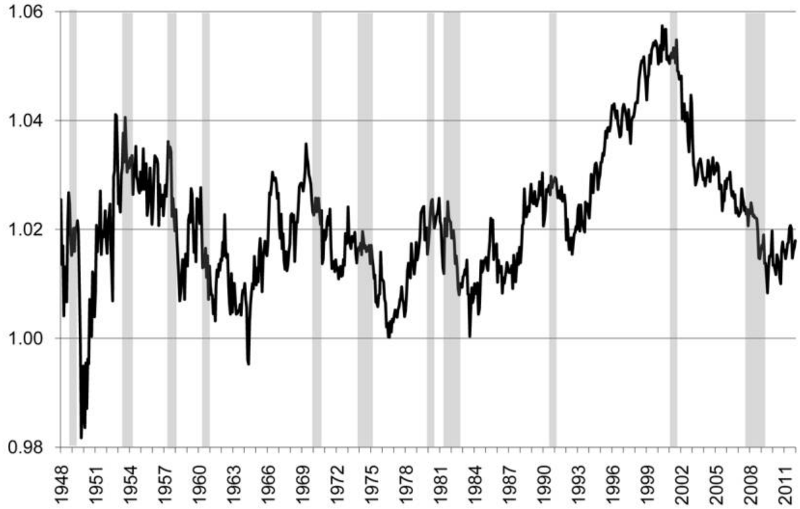

FIG. 2.—Ratio of establishment survey employment to household survey nonagricultural wage and salary employment, 1948–2011. The household series is smoothed for population control revisions. Source: www.bls.gov.

displayed here are on a person basis rather than a jobs basis, whereas the CES data in both figures are on a jobs basis. It is nonetheless notable that, over this longer period, establishment survey employment typically increases relative to household survey employment during business cycle expansions, then falls in relative terms during recessions and the early part of the subsequent recovery period. Figure 2 suggests strongly that whatever story we tell about the differing behavior of the CES and CPS employment series should have a cyclical dimension.

One relevant hypothesis is that, in tight labor markets, there may be a growing number of marginal jobs that are not reported in the CPS, leading to growth in the number of people in the $X_{NH,E}$ and/or $Y_{1H,2E}$ cells. As economic activity strengthens, employers may become more inclined to hire extra help to cover peak workloads, raising the number of short-duration wage and salary jobs. For example, the owner of a retail store might decide to hire a larger than usual number of temporary staff over the Christmas holidays. To the extent that short-duration jobs are less likely to be reported by CPS respondents, either because the respondent fails to report a short-duration job that was in progress during the CPS reference week or because the job does not overlap the CPS reference week, this might lead us to expect an increase in $X_{NH,E}$ as the economy tightens (i.e., to expect that $X_{NH,E}$ will be procyclical). A similar dynamic might be in play for $Y_{1H,2E}$. Increases in $X_{NH,E}$ and/or $Y_{1H,2E}$ should be associated with increases in CES employment relative to CPS employment.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Off-the-books or Form 1099 employment also may be cyclical. As economic activity strengthens and labor markets become tighter, people tend to leave informal jobs not recorded on employer payrolls for formal jobs. Alternatively, during periods of stronger economic activity, employers might "regularize" more of their jobs, converting them from off-the-books, independent contractor, or consultant positions to jobs for which the employer pays applicable employment taxes. As labor markets tighten, we might thus expect $X_{H,NE}$ to fall and perhaps also for $Y_{2H,1E}$ to fall. Put another way, we might expect both $X_{H,NE}$ and $Y_{2H,1E}$ to be countercyclical. Decreases in $X_{H,NE}$ and/or $Y_{2H,1E}$ during an economic upturn also should be associated with a growing gap between CES and CPS employment.

In order for our linked CPS-UI data to be useful for understanding the cyclical behavior of the CPS and CES employment series, it must be the case that the series derived from the linked data set mimic the published employment series. The two sets of numbers are not strictly comparable. The linked sample includes data only for 16 states and, to match the coverage of the UI wage records, we have focused on jobs in the private sector, state government, and local government. Another difference is that the series based on the linked sample are quarterly measures for the first quarter of the year rather than monthly measures that cover the entire year. Finally, the linked sample estimates are based on many fewer observations than the published estimates. Despite these differences, the linked sample CPS and UI employment estimates display the same puzzling pattern as the published CPS and CES employment estimates, with the UI-based employment estimate rising faster than the CPS-based estimate between the first quarter of 1998 and the first quarter of 2001, and then falling toward the CPS series. Additional details may be found in appendix A, but the similarity of the patterns observed in employment estimates based on our linked sample to those based on published estimates leads us to think that our linked sample data can be useful for understanding the trend discrepancies between CPS and CES employment.

B. Aggregate Time Series Patterns in Marginal Workers and Marginal or Nonstandard Jobs

The question we would like to answer is whether the person and job characteristics that help to explain individual-level discrepancies between household-reported and employer-reported employment status also can help to explain discrepancies in the time series behavior of the corresponding aggregate employment measures. While having only 8 years of data covering one cyclical episode limits the conclusions we can draw, there is nonetheless useful information to be gleaned from an investigation of changes in the number of marginal ($X_{NH,E}$, $Y_{1H,2E}$) and off-the-

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

books or Form 1099 ($X_{H,NE}$, $Y_{2H,1E}$) jobs over the period covered by our matched sample.

Using estimates derived from the weighted matched CPS-UI sample, figure 3 displays the trend both in the number of people employed in the UI but not the CPS ($X_{NH,E}$) and in the number of people employed in the CPS but not in the UI ($X_{H,NE}$). All else the same, any increase (decrease) in $X_{NH,E}$ relative to $X_{H,NE}$ will be associated with growth (decline) in UI employment relative to CPS employment. Over the period from 1996 to 2001, $X_{NH,E}$ and $X_{H,NE}$ fluctuate relative to one another, but not in a consistent fashion. Over the 2001–3 period, $X_{H,NE}$ grew by about 900,000 workers while $X_{NH,E}$ fell by about 300,000 workers, both movements that would have contributed to the relative increase in CPS employment over this period. The large increase in $X_{H,NE}$ is consistent with marked growth in the number of off-the-books or independent contractor jobs. The modest decline in $X_{NH,E}$ is consistent with a shrinking number of marginal (short-duration or low earnings) jobs. The combined swing of about 1.1 million jobs is substantial and is an important factor in the shrinking discrepancy between the household and employer job counts over this subperiod.

Figure 4 displays the trend in the number of people categorized as holding more than one in-scope job in the UI data but a single in-scope job in the CPS data ($Y_{1H,2E}$) and the trend in the number of people categorized as holding more than one in-scope job in the CPS data but a single job in the UI data ($Y_{2H,1E}$). As in the previous sections, we report results based on the more restrictive criteria for defining multiple job holding in the CPS, but all of the main conclusions are robust to which set of criteria we use. In the results displayed, the number of people in the $Y_{1H,2E}$ category

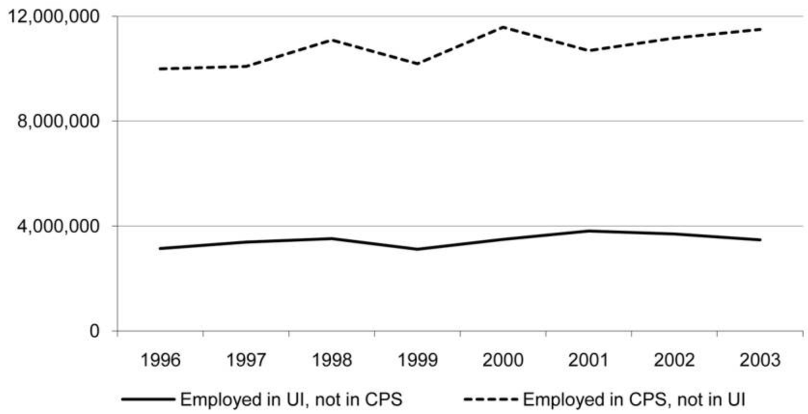

FIG. 3.—Estimated number of people in off-diagonal employment status cells, 1996–2003.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

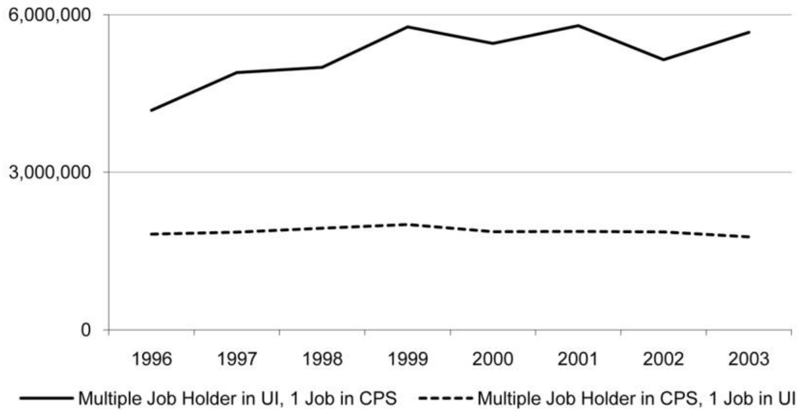

FIG. 4.—Estimated number of people in off-diagonal multiple job status cells, 1996–2003.

grew by about 1.6 million people between 1996 and 1999, leveling off thereafter. In contrast, the number of people holding multiple jobs in CPS but not in UI ($Y_{2H,1E}$) grew by about 200,000. On net, the difference between $Y_{1H,2E}$ and $Y_{2H,1E}$ increased by about 1.4 million people over the 1996–99 period, contributing significantly to the increase in the UI job count relative to the CPS job count over these years.

Looking at figures 3 and 4 together suggests that different components of the off-diagonal elements of the employer and household data play a role in the different subperiods. One of the reasons that the employer job count grew so rapidly relative to the household job count over the 1996–99 period appears to be that the number of employed people holding multiple jobs increased faster in the employer than in the household data. Then in the downturn from 2001 through 2003, employment status plays a bigger role, with growth especially in the number of individuals identifying themselves as employed in the household data but not in the employer data.

The combined effects are illustrated in figure 5, which shows the estimated number of CPS jobs that do not appear in the UI data ($X_{H,NE}$ + $Y_{2H,1E}$) and the estimated number of UI jobs not reported in the CPS ($X_{NH,E}$ + $Y_{1H,2E}$). Jobs counted in the UI but not found in the CPS grew by 2.3 million between 1996 and 2001, while the number of jobs counted in the CPS but not in the UI grew by much less—about 600,000—over the same period. These patterns are consistent with the more rapid growth of the employer-based employment measure than of the household-based employment measure over these years shown in figure 1. In contrast, from 2001 to 2003, the number of jobs counted in the CPS but not found in the UI grew by 800,000, while the number of jobs counted in the UI but not found in the CPS fell by about 500,000. Again, this pattern corresponds

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

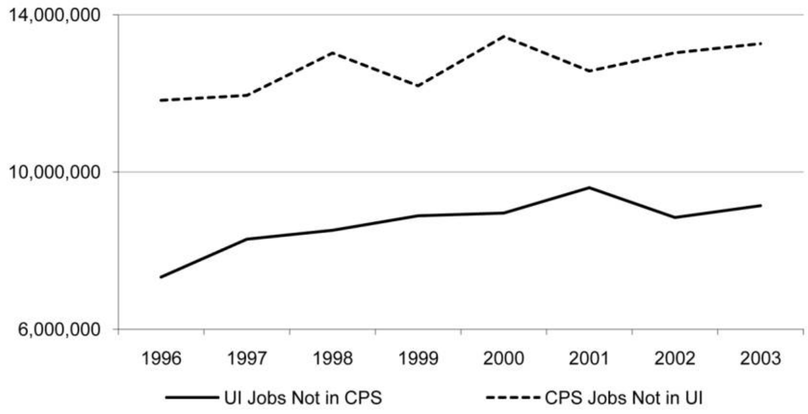

FIG. 5.—Estimated number of jobs in combined off-diagonal cells (number of persons working plus number of multiple job holders), 1996–2003.

well with the decline in the employer-based measure of employment relative to the household-based measure that is shown in figure 1.

C. Simulating the Fluctuations in Observed Employment

The next question we address is whether changes over time in the characteristics of workers and jobs can account for the patterns in figures 3 and 4. We use information on the composition of workers and jobs in each year together with estimated coefficients from the linear probability models including person and job characteristics variables reported in tables 5, 6, 8, and 9 to simulate aggregate values for each of the off-diagonal cells in the employment status and multiple-job-holding matrices. To illustrate, to simulate annual values for $X_{NH,E}$, we first simulate the $X_{NH,E}$ share of UI employment, using the estimated coefficients from the model in the second column of table 5. Each year's share is equal to the sum of the equation's intercept plus the average of the year dummy coefficients plus the vector product of the characteristic coefficient estimates and the mean characteristic values for that year. These simulated shares are then multiplied by the actual number of UI workers in the same year. If we included the estimated year effects then by construction we would match exactly the year-to-year aggregate fluctuations. Since we are using the average of the year dummy coefficients, the time-series variation for the prediction $X_{NH,E}$ is coming from the interaction of the predicted change in the share in $X_{NH,E}$ with the actual number of UI workers in the same year.

Simulated values for $X_{H,NE}$, $Y_{1H,2E}$, and $Y_{2H,1E}$ are constructed similarly using the coefficient estimates from tables 6, 8, and 9 and the actual numbers of CPS workers, UI multiple job holders, or CPS multiple job holders, as appropriate. To avoid noise in the simulated values attributable to noise in the estimated mean characteristic values by year, we use the largest

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

possible data set to estimate those means. Specifically, for the demographic and CPS job characteristics we use all of the records for people in our 16 states who completed CPS interviews in January, February, and March in the year in question, rather than the more restricted CPS sample for which a PIK was available. For the UI job characteristics, we use all of the UI wage records available for our 16 states.

Figure 6 displays the actual and predicted changes for the 1996–2001 and 2001–3 period. For brevity of exposition, we focus on the combined number of jobs in the off-diagonal cells as in figure 5. That is, we define the number of UI jobs not found in the CPS as the sum of $X_{NH,E}$ and $Y_{1H,2E}$ and the number of CPS jobs not found in the UI as the sum of $X_{H,NE}$ and $Y_{2H,1E}$. We find that the changes in the number of UI jobs not found in the CPS from 1996 to 2001 and from 2001 to 2003 are captured reasonably well by the predicted values from the demographic and job characteristics. Specifically, the predicted values capture 68% of the increase from 1996 to 2001 and 81% of the decline from 2001 to 2003. The predicted numbers of CPS jobs not found in the UI, however, do not capture the actual changes well—the predictions substantially overshoot the actual change from 1996 to 2001 and imply a modest decline from 2001 to 2003 rather than the increase actually observed. The success in accounting for the changes in the UI jobs not found in the CPS is about equally shared by the demographic and job characteristics.

VII. Concluding Remarks

Using a large data set that matches CPS respondents to UI wage records, we find large discrepancies in employment status and in number of jobs held at the individual level. Our basic results show that 17.6% of

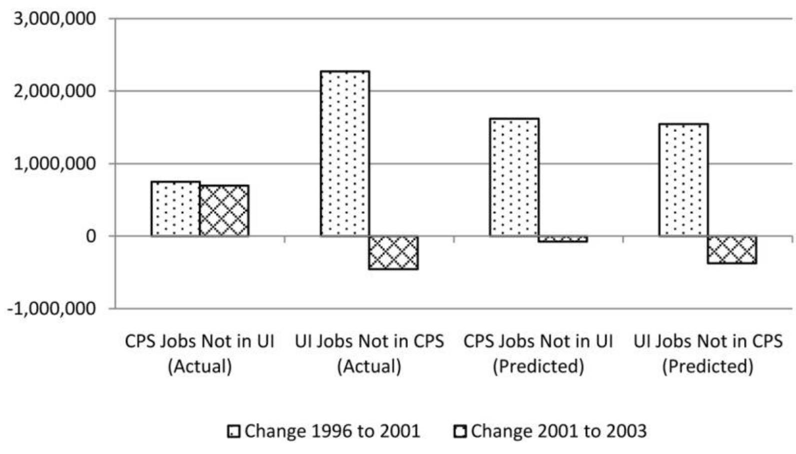

FIG. 6.—Actual and predicted changes in number of jobs in combined off-diagonal cells (number of persons working plus number of multiple job holders).

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

CPS workers who are working in a job that should be covered by unemployment insurance cannot be found in the UI records and that 6.4% of UI workers report no corresponding job in the CPS. Even larger discrepancies are found for multiple job holders. Some 69.2% of multiple job holders in the UI data have only one CPS job and 44.6% of multiple job holders in the CPS only have one UI job.

We find further that the large off-diagonal discrepancies are not random but rather are systematically associated with observable person and job characteristics. UI workers not found in the CPS tend to have person and job characteristics consistent with marginal employment that the worker might not consider their main activity and thus fail to report. CPS workers not found in the UI data have person and job characteristics consistent with off-the-books or independent contractor relationships. Similar remarks apply to discrepancies for multiple job holders. Taken together, our findings suggest that it is workers and jobs on the margin for which household survey data and employer-provided data are most likely to yield discrepant results.

These findings have implications for microeconomic labor market analyses as well as for discrepancies in aggregate time series patterns of employment from household and employer data. At the micro level, the estimated impact of events such as worker displacement and policy interventions such as job training programs on employment outcomes may vary depending on whether household or employer data are used, especially for workers with the person and job characteristics that are associated with the largest discrepancies across data sets. In some exploratory analysis, we find that the time series patterns of employment rates following displacement differs substantially depending on whether for the same person employment is measured using CPS or UI data. At the aggregate level, our findings suggest that the cyclicality of marginal workers and marginal jobs may be central to understanding the differing movements of household-based versus employer-based employment statistics.

To further our understanding of the implications of these measurement discrepancies, it would be desirable to extend the time period and the geography covered by the matched CPS and UI data. Among other advantages, a national database should yield sufficient regional variation that differences in the cyclical pattern of aggregate employment estimates based on household-reported versus employer-reported data could be investigated, taking advantage of differences in the cycle across regional labor markets. It also would be useful to update the matched UI-CPS data sets on an ongoing basis. A timely and thorough CPS-UI microdata match would facilitate economic research into employment and wages. Particularly given the role of the CPS as the workhorse data set relied upon by both academic researchers and policy makers to answer many questions about the labor market, having a better understanding of how the CPS captures employment status would be of high value.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Appendix A
Data

Analysis Sample

The data file used in our analysis is a linked sample for which CPS records for the first quarter of the year have been matched with state unemployment insurance (UI) wage records for the same period for the same individuals. The identifier used to link these records is a Protected Identity Key (PIK), the person ID used internally at Census to process and integrate person-level data. Social Security numbers (SSNs) are collected for CPS respondents who complete the Annual Social and Economic Supplement (ASEC) conducted as part of the March CPS. Information on age, race, and sex is compared with the information in the Social Security Administration's records to validate reported SSNs. In addition, the same information together with name and address may be used to assign SSNs to respondents who agree to have their survey responses matched to administrative records but do not know or cannot remember their SSN. All UI wage records also contain an SSN. Files containing SSNs received by the Census Bureau are maintained and protected in a secure environment within an administrative records division. The administrative records division validates the SSNs and replaces the SSNs with a Protected Identity Key (PIK).

PIKs are available in the CPS only for March, and they are missing for 20%–30% of March CPS respondents. In addition, our analysis requires CPS reports from each sample member for all three months of the quarter (January, February, and March). This means that our sample includes only individuals for whom March was their third, fourth, seventh, or eighth month in the CPS sample, cutting the sample otherwise available roughly in half. Some additional sample is lost due to nonresponse in one or more of the three months in the quarter.

Although the LEHD program is approaching national coverage, data are available in all years from the mid-1990s through 2003, the end point of our sample, only for 17 states: California, Colorado, Florida, Idaho, Illinois, Kansas, Maryland, Minnesota, Missouri, Montana, New Mexico, North Carolina, Oregon, Pennsylvania, Texas, Washington, and Wisconsin. In most states, American Community Survey data show that 95% or more of the workers resident in the state also are employed in the state, but more than 15% of workers resident in Maryland (where they are coded in the CPS) actually work in another state or the District of Columbia (where their employers pay unemployment insurance taxes). Because we do not have UI wage records for Virginia, West Virginia, or the District of Columbia, we chose to drop Maryland from our analysis. The 16 states remaining account for roughly 50% of employment nationwide.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

To make the age range of our sample comparable to the age range of the population covered by published CPS employment estimates, we restrict our analysis sample to persons aged 16 years and older.

Estimation Weights

To account for the restrictions imposed on our linked sample, we modify the March CPS estimation weights using a two-step propensity scoring procedure.

Step 1: Our sample includes only those March respondents who also responded to the January and February monthly surveys. Starting with the full March sample aged 16 and above, we estimate the probability that an individual was interviewed in all three months of the first quarter as a function of age group, gender, race, education, marital status, foreign-born status, and an indicator for whether the person had in-scope employment in March. For each observation in the three-month sample, we then construct a weight adjustment factor equal to the inverse of this predicted probability. The average value of this propensity score adjustment factor is roughly 2.5.

Step 2: Roughly 20%–30% of March records do not have a PIK. For each record for a person age 16 and above living in one of our 16 states who was interviewed in January, February, and March, we estimate the probability that the record has a PIK as a function of the same traits listed above, but using an indicator for in-scope employment at any time during the quarter rather than during March. The weight used in the estimation of this stage is the adjusted weight constructed by applying the weight adjustment factor from step 1 to the March basic estimation weights. For each record with a PIK, we then apply a second weight adjustment factor equal to the inverse of the predicted probability of having a PIK. This adjustment raises the average value of the estimation weights by 25%–40%, depending on the year.

UI Variable Construction

Each quarterly UI wage record identifies an employer, a worker, and the earnings paid by the employer to the worker over the quarter.[13] We use information on the employer's industry to exclude private household and agricultural jobs and information on establishment ownership to exclude jobs outside of the private sector, state government, or local government. The wage records are used to construct an indicator of whether a person

[13] In the UI wage records, the employer ID is a state unemployment insurance account number for the business. For multi-establishment businesses, this employer ID is typically at a level above the establishment but below the firm, generally representing the activity of the firm at an industry-state level. For details about the UI wage record data, see Abowd et al. (2009).

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

had positive in-scope UI earnings during the quarter (i.e., was employed) and whether the person received positive earnings during the quarter from more than one employer (i.e., was a multiple job holder).

We should be clear that the reference period we use in this paper is "during a quarter" rather than "at a point in time," and as such, we count anyone who has two wage records during a quarter as a multiple job holder, whether those jobs were held simultaneously or sequentially.

We also construct variables that capture earnings during the quarter on the worker's primary job, defined as the highest-earning job, and, for those who hold more than one job during the quarter, variables that capture earnings on any additional jobs. If we observe that a worker is employed at a particular business in the first quarter and also worked for that same employer in the preceding and/or following quarter, the job is coded as a long-duration job.

CPS Variable Construction

CPS employment is collected monthly and refers to employment during the survey reference week. Although the CPS records up to four jobs in each month, except in the outgoing rotation months (CPS month-in-sample four and eight), the class-of-worker information needed to determine whether a job is in-scope is collected only for the main job. Class-of-worker information also is collected for the second job, if there is one, in the outgoing rotation months. No class-of-worker information is collected for other jobs.

We define a CPS respondent to be an in-scope worker in the first quarter if they worked at a main job in any one of the three months that was a nonagricultural private sector wage-and-salary job or a wage-and-salary job in state or local government.[14] To count the number of unique in-scope jobs held over a quarter, we would need to know both the number of jobs held each month and the number of employer changes that may have occurred across months. Because we know the class of worker for second jobs only in the outgoing rotation months and have no class-of-worker information for additional jobs, we cannot be certain about the number of in-scope jobs held by those who report holding multiple jobs at any point in time. Further, the monthly CPS questionnaire probes only for changes in main job over time, and even this information is not complete. Rather than attempting to count the number of in-scope jobs that

[14] This definition excludes individuals whose primary job is out-of-scope but who have a second job that is in-scope. Except in an individual's outgoing rotation month, we cannot say with certainty whether a second job is in-scope. In data for the March outgoing rotation groups covering the years 1996 to 2003, adding those with out-of-scope primary jobs but in-scope second jobs to the weighted count of in-scope workers would raise the total number of in-scope CPS workers less than 1% on average.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

a worker holds during the quarter, we instead construct indicator variables for whether a worker holds one in-scope job or more than one in-scope job during the quarter.

Simultaneous jobholders: If a worker has two or more jobs in January, February, and/or March, has two or more jobs in his or her outgoing rotation month (March or April), and both of the jobs held in the outgoing rotation month are in-scope jobs, we can be reasonably confident that the worker held two simultaneous in-scope jobs at some point during the quarter. In each year, roughly 2% of CPS workers fall into this group, which we term "multiple job group 1" (MJ1).

If a worker has two or more jobs in January, February, and/or March and has two or more jobs in the outgoing rotation month, but class-of-workers detail is missing for at least one of these outgoing rotation jobs, the worker could have held simultaneous in-scope jobs at some point during the quarter, but this is less certain. The same is true of workers who have two or more jobs in January, February, and/or March but only one job in the outgoing rotation group. These people are assigned to multiple job group 2 (MJ2). In each year, between 4% and 5% of workers fall into this category.

Job changers: For anyone who was employed in the prior month and reports having a job in the current month, the basic CPS questionnaire includes a question about whether the respondent has changed employer. The answers to this question should identify job changers, but it applies only to the main job and is not asked consistently.[15]

Most respondents who were employed in both January and February were asked in February if they were still working for the same employer on their main job. Similarly, most respondents employed in both February and March were asked this same question in March. Respondents who report a change in job from one in-scope employer to another are assigned to job change group 1 (JC1). Depending on the year, between 2% and 3% of workers belong to this group.

If a worker holding an in-scope job either in both January and February or in both February and March does not report a job change directly, either responding negatively to the job change question or not having been asked the question, but does report a change in industry, occupation, or class of worker between months, he or she is assigned to job change group 2 (JC2). In addition, if a worker held fewer in-scope jobs in February than in both January and March and the worker was not on layoff in February, they are assigned to group JC2. Again depend-

[15] Roughly 7% of workers who are employed in adjacent months have a missing value for the answer to the job change question. The question is a screener to determine whether industry and occupation need to be updated, and we were told that interviewers have the discretion just to re-ask those questions if they have any reservations about the quality of the information collected in the previous interview.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

ing on the year, between roughly 4% and 6% of workers belong to this group.

Multiple job indicator: We use the indicator variables just described to construct both a more restrictive and a less restrictive measure of which CPS respondents hold more than one in-scope job during the quarter. The more restrictive measure counts only workers assigned to groups MJ1 and/or JC1 as multiple job holders. The less restrictive measure includes all workers assigned to groups MJ1, MJ2, JC1, and/or JC2.

Other job characteristics: We also construct several other job characteristic measures based on the CPS data. One captures whether the person was without an in-scope job in at least one month of the quarter.[16] For those holding more than one job during the quarter, two others capture whether they ever held multiple jobs simultaneously and whether this was so in all three months of the quarter. Another indicates whether a person is in an industry and occupation in which the share of self-employed workers for the industry/occupation cell is in the top 20% for all employment, as measured in the full March 1999 CPS sample. The rationale behind this variable is that, in jobs with many self-employed workers, a higher share of respondents are at risk of reporting erroneously that they are a wage and salary worker. Information on hours of work is used to distinguish those who worked full time in at least one month from those who never worked full time and to distinguish those who worked 16 hours or more on a second job in at least one month from those whose second jobs always involved fewer than 16 hours. Finally, for the main job, we use the weekly earnings reported at the time of the outgoing rotation interview, converted to a quarterly equivalent at the same weekly earnings rate, to construct a set of earnings dummies as similar as possible to those constructed for the UI data. One of the earnings dummies captures those for whom no earnings information is available (approximately 11% of those with some work during the quarter).

Relating the Trends in Published Employment Estimates to the Linked Sample Estimates

As noted in the body of the paper, in order for our linked CPS-UI data to be useful for understanding the cyclical behavior of the CPS and CES employment series, it must be the case that CPS and UI employment series derived from the linked data set behave similarly to the published CPS and CES employment series. The charts displayed in figure A1 show how accounting in turn for each of the differences between the published CPS and CES estimates affects the behavior of the resulting employment se-

[16] This "Work Discontinuity" indicator is related to the monthly CPS gross flows data that are analyzed by authors such as Frazis et al. (2005). The work discontinuity indicator captures CPS respondents who work only one or two months of the first quarter, as opposed to working all three months.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

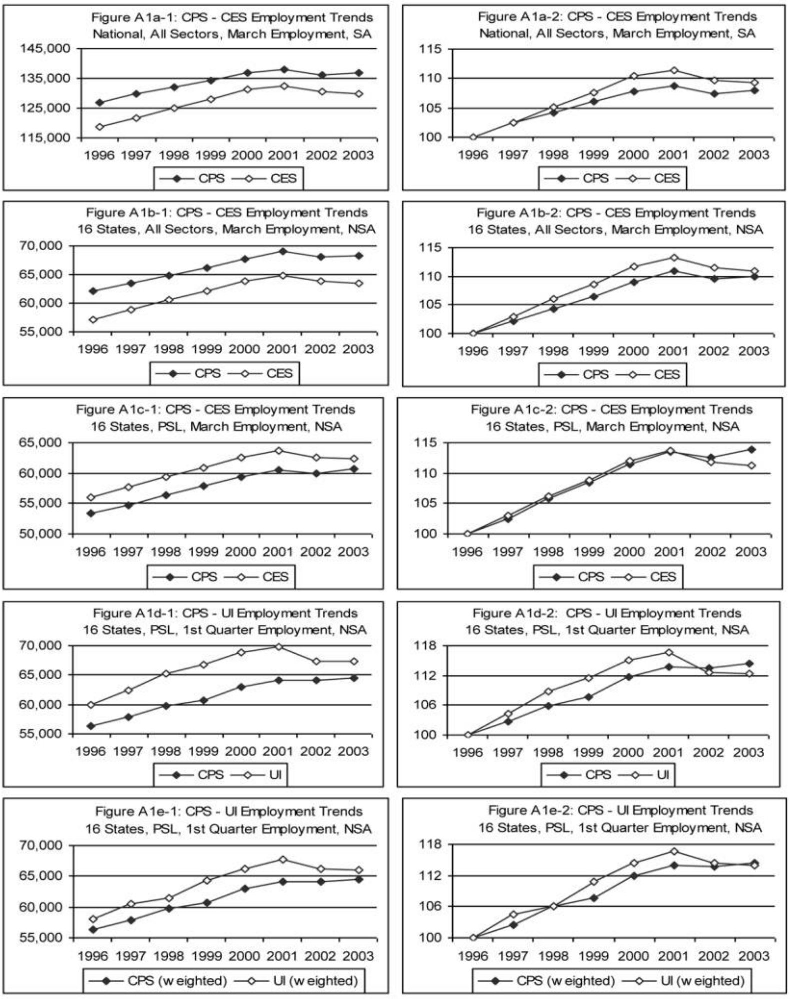

FIG. A1.—Effects of sample restrictions and employment concepts on CPS, CES, and UI employment trends, 1996–2003, in levels in the charts on the left and in index form in the charts on the right.

ries over time. In each of the panels in the figure, the CPS series measures the number of individuals employed and the CES or UI series measures the number of jobs. The top panel in the figure, figure A1*a*, shows the trends in seasonally adjusted employment for the month of March taken from figure 1, in levels in the chart on the left and in index form in the chart on the right. The gap between the CPS and CES employment estimates that emerges between 1998 and 2001 and then closes between 2001 and 2003 is very apparent in these data. These charts are the benchmark to which each of the other pairs of charts ultimately is compared.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

The next panel, figure A1*b*, narrows the geographic scope of the estimates, reporting seasonally unadjusted numbers for the 16 states represented in our linked sample. Seasonally adjusted series for the same states are not shown but look very similar. The CPS estimates were calculated from the microdata, and the CES estimates were downloaded from the BLS website. The employment trend discrepancies are essentially the same in our 16 states as in the national data.

Figure A1*c* further narrows the scope of the estimates reported to jobs in the private sector, state government, or local government (denoted as PSL), removing the self-employed, agricultural workers, private household workers, and federal government workers from the CPS data and federal government workers from the CES data. These restrictions are needed so that the scope of the CPS microdata we analyze conforms to the scope of the UI wage records. CPS and CES data for this measure of employment show essentially the same trend growth rate between 1996 and 2001, although because estimated CES employment starts in 1996 at a higher level than CPS employment, it is still the case that CES job gains exceed CPS job gains over this period, by more than half a million over the 5 years in question. Further, as in the earlier figures, CES employment falls off sharply after 2001, while CPS employment falls more modestly between 2001 and 2002 and then rises between 2002 and 2003.

In figure A1*d*, we switch from the published CES data to the UI wage records microdata. For the reasons already explained, this requires a switch from the familiar CPS monthly employment concept to a first-quarter employment concept (employed during January, February, or March). This increases the level of employment, since more people are employed at some point during the quarter than in any one month, but the growth trends in figure A1*d* look very similar to those in the top panel.

Finally, in figure A1*e*, we switch from estimates based on the full UI database to estimates based on our linked sample of approximately 12,000–15,000 individuals per year. We use the adjusted March CPS basic estimation weights already described to create 16-state employment totals for both the CPS and the UI data. The resulting CPS employment estimates are very close to those in the previous panel; because of the missing UI jobs issue discussed in the body of the paper, the UI employment estimates are a few percentage points lower. The employment trends in this last panel, however, are strongly similar.

References

Abowd, John M., Kevin L. McKinney, and Lars Vilhuber. 2009. The link between mass layoffs, human capital and firm deaths. In *Producer dynamics*, ed. Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts. Chicago: University of Chicago Press.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2009. The LEHD infrastructure files and the creation of the quarterly workforce indicators. In *Producer dynamics*, ed. Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts. Chicago: University of Chicago Press.

Abowd, John M., and Martha H. Stinson. 2011. Estimating measurement error in SIPP annual job earnings: A comparison of census survey and SSA administrative data. SEHSD Working Paper no. 2011–19, US Census Bureau, Washington, DC. http://www.census.gov/sipp/workpapr/abowd-stinson-ME-ss-20110707-sehsdworking.pdf.

Belman, Dale L., and Richard Block. 2009. The social and economic costs of employee misclassification in Michigan. Report of the Indiana, Illinois, Iowa Foundation for Fair Contracting. http://www.iiiffc.org/pdf/Employee_Classification/MI%20-Social%20&%20Economic%20Costs.2009.pdf.

Blakemore, Arthur E., Paul L. Burgess, Stuart A. Low, and Robert D. St. Louis. 1996. Employer tax evasion in the unemployment insurance program. *Journal of Labor Economics* 14, no. 2:210–30.

Bowler, Mary, and Teresa L. Morisi. 2006. Understanding the employment measures from the CPS and CES survey. *Monthly Labor Review* 129, no. 2:23–38.

Bound, John, Charles Brown, Greg J. Duncan, and Willard L. Rodgers. 1994. Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics* 11, no. 3:345–68.

Bound, John, and Alan B. Krueger. 1991. The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics* 9, no. 1:1–24.

Brown, Clair, John Haltiwanger, and Julia Lane. 2006. *Economic turbulence: Is a volatile economy good for America?* Chicago: University of Chicago Press.

BLS (Bureau of Labor Statistics). 1997. Quality improvement project: Unemployment insurance wage records. Unpublished report, Bureau of Labor Statistics, Washington, DC.

Cancian, Maria, Robert Haveman, Thomas Kaplan, and Barbara Wolfe. 1999. Post-exit earnings and benefit receipt among those who left AFDC in Wisconsin. Special Report no. 75, Institute for Research on Poverty, University of Wisconsin. http://www.irp.wisc.edu/publications/sr/pdfs/sr75.pdf.

Carre, Francoise, and Randall Wilson. 2004. The social and economic costs of employee misclassification in construction. Report of the Construction Policy Research Center, Harvard University. http://www.law.harvard.edu/programs/lwp/Misclassification%20Report%20Mass.pdf.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

Davis, Steven J., John Haltiwanger, Ronald S. Jarmin, C. J. Krizan, Javier Miranda, Alfred Nucci, and Kristin Sandusky. 2009. Measuring the dynamics of young and small businesses: Integrating the employer and nonemployer universes. In *Producer dynamics*, ed. Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts. Chicago: University of Chicago Press.

Farber, Henry. 1993. The incidence and costs of job loss: 1982–1991. *Brookings Papers on Economic Activity: Microeconomics*, 73–132.

———. 1997. The changing face of job loss in the United States, 1981–1995. *Brookings Papers on Economic Activity: Microeconomics*, 55–128.

———. 2005. What do we know about job loss in the United States? Evidence from the Displaced Workers Survey, 1981–2004. *Federal Reserve Bank of Chicago Economic Perspectives* 29, no. 2:13–28.

Frazis, Harley J., Edwin L. Robison, Thomas D. Evans, and Martha A. Duff. 2005. Estimating gross flows consistent with stocks in the CPS. *Monthly Labor Review* 128, no. 9:3–9.

Haltiwanger, John C., Julia I. Lane, and James R. Spletzer. 2007. Wages, productivity, and the dynamic interaction of businesses and workers. *Labour Economics* 14, no. 3:575–602.

Hotz, V. Joseph, and John Karl Scholz. 2001. Measuring employment and income for low-income populations with administrative and survey data. In *Studies of welfare populations: Data collection and research issues*, ed. Michele Ver Ploeg, Robert A. Moffitt, and Constance F. Citro. Washington, DC: National Academy Press.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. Earnings losses of displaced workers. *American Economic Review* 83, no. 4:685–709.

Juhn, Chinhui, and Simon Potter. 1999. Explaining the recent divergence in payroll and household employment growth. *Federal Reserve Bank of New York Current Issues in Economics and Finance* 5, no. 16:1–6.

Kornfeld, Robert, and Howard S. Bloom. 1999. Measuring program impacts on earnings and employment: Do unemployment insurance wage reports from employers agree with surveys on individuals? *Journal of Labor Economics* 17, no. 1:168–97.

Mellow, Wesley, and Hal Sider. 1983. Accuracy of response in labor market surveys: Evidence and implications. *Journal of Labor Economics* 1, no. 4:331–44.

Nardone, Thomas, Mary Bowler, Jurgen Kropf, Katie Kirkland, and Signe Wetrogan. 2003. Examining the discrepancy in employment growth between the CPS and the CES. Report prepared for the Federal Economic Statistics Advisory Committee, Washington, DC. http://www.bls.gov/bls/fesacp2101703.pdf.

Roemer, Marc. 2002. Using administrative earnings records to assess wage data quality in the March Current Population Survey and the Survey

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions

of Income and Program Participation. Technical Paper no. TP-2002-22, Longitudinal Employer-Household Dynamics, US Census Bureau, Washington, DC. http://lehd.did.census.gov/led/library/techpapers/tp-2002-22.pdf.

Ruhm, Christopher J. 1990. Bridge jobs and partial retirement. *Journal of Labor Economics* 8, no. 4:482–501.

Schochet, Peter Z., John Burghardt, and Sheena McConnell. 2008. Does Job Corps work? Impact findings from the National Job Corps Study. *American Economic Review* 98, no. 5:1864–86.

Schoeni, Robert F., and Michael Dardia. 1996. Wage losses of displaced workers in the 1990s. Labor and Population Working Paper no. 96–14, Rand Corporation, Santa Monica, CA. http://www.rand.org/content/dam/rand/pubs/drafts/2008/DRU1474.pdf.

Stevens, David W. 2007. Employment that is not covered by state unemployment insurance laws. Technical Paper no. TP-2007-04, Longitudinal Employer-Household Dynamics, US Census Bureau, Washington, DC. http://lehd.did.census.gov/led/library/techpapers/tp-2007-04.pdf.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The psychology of survey response*. Cambridge: Cambridge University Press.

von Wachter, Till M., Elizabeth Weber Handwerker, and Andrew K. G. Hildreth. 2009. Estimating the "true" cost of job loss: Evidence using matched data from California, 1991–2000. Unpublished manuscript, Department of Economics, Columbia University. http://www.columbia.edu/~vw2112/papers/dwsui_vonwahandwhild.pdf.

This content downloaded from 129.2.28.46 on Wed, 6 Nov 2013 11:13:22 AM
All use subject to JSTOR Terms and Conditions