NBER WORKING PAPER SERIES

EXPLORING DIFFERENCES IN EMPLOYMENT BETWEEN HOUSEHOLD AND
ESTABLISHMENT DATA

Katharine G. Abraham
John C. Haltiwanger
Kristin Sandusky
James Spletzer

Exploring Differences in Employment between Household and Establishment Data
Katharine G. Abraham, John C. Haltiwanger, Kristin Sandusky, and James Spletzer
NBER Working Paper No. 14805
March 2009
JEL No. C80,J21

ABSTRACT

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 marginal jobs, and 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.

Katharine G. Abraham
Joint Program in Survey Methodology
1218 LeFrak Hall
University of Maryland
College Park, MD 20742
and NBER
kabraham@survey.umd.edu

John C. Haltiwanger
Department of Economics
University of Maryland
College Park, MD 20742
and NBER
haltiwan@econ.umd.edu

Kristin Sandusky
U.S. Bureau of Census
LEHD Program
4700 Silverhill Road
Suitland, MD 20233
lee.k.sandusky@census.gov

James Spletzer
Bureau of Labor Statistics
2 Massachusetts Avenue, NE, Suite 4945
Washington, DC 20212
Spletzer.Jim@bls.gov

I. Background

Labor economists and others interested in the labor market rely both on data from household surveys, such as the Current Population Survey (CPS), and on data from employer surveys, such as the Current Employment Statistics Survey (CES). Questions frequently have been raised about whether household-provided and employer-provided data yield a consistent picture of labor market activity. The differing trends in CPS employment and CES employment leading up to and following the 2001 recession, for example, provoked a great deal of discussion about whether the two sets of numbers could be reconciled and, if not, which survey should be believed. 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 individuals, we document the large discrepancies in individuals' employment status in 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 individual-level labor market analysis and for the analysis of aggregate labor market trends.

There is an existing 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 both from workers and from their employers, with a focus on the extent and implications of measurement error in conventional labor market 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) earnings records to study measurement error in earnings data, and Bound et al (1994) used data from the PSID validation study to study measurement error in earnings and other labor market variables. Roemer (2002) and Abowd and Stinson

(2002) use SIPP data matched to SSA earnings data to study measurement error in earnings. Each of these studies provides valuable information about discrepancies in the reporting of job characteristics for people who say they are employed and for whom employer-reported job information can be obtained, but none of them addresses discrepancies in the reporting of employment status by households and employers.

Another relevant literature is a set of studies that have compared the effects of labor market programs estimated using survey data collected from individuals to the effects estimated from administrative UI wage records obtained from employers. In these studies, the incidence of employment in the study population features prominently in the analysis. Kornfeld and Bloom (1999) is an early example; in a comparison of the impact of job training funded under the Job Training Partnership Act (JTPA), Kornfeld and Bloom find only modest differences between quarterly employment rates computed from survey data versus UI wage records. In a recent study of how Job Corps participation affects subsequent employment outcomes, Schochet et al (2008) find that quarterly employment rates are considerably higher in the survey data relative to the UI wage records. Schochet et al's interpretation of this finding is that the UI wage records may not be capturing short duration informal jobs. Hotz and Scholz (2001) provide a useful survey of studies that use UI wage record data to study the low income population, focusing on the strengths and weaknesses of survey data versus administrative data for learning about the employment experience of this population. While of considerable value, these studies are by design focused on a narrow segment of the population. To our knowledge, despite their potentially important implications for a variety of labor market analyses, no existing study has systematically examined discrepancies in the reporting of employment status for the population as a whole. This is an important contribution of our analysis.

At the macro level, the different behavior of CPS and CES employment series during the late 1990s and early 2000s has attracted considerable attention. This CES-CPS employment trend discrepancy is evident in Figure 1. Two CPS series are shown in the figure, one the published CPS employment figures and the other an adjusted series that makes the CPS numbers as comparable as possible to the CES numbers. As can be seen, the adjusted CPS series matches the CES series well in many years, but CES employment grew markedly faster than CPS employment from 1998 through 2001, opening a large gap between the two series, and then retreated just as markedly between 2001 and 2003. As discussed by Bowler and Morisi (2006), a wide variety of hypotheses have been offered about the divergent behavior of CES and CPS employment over the 1998 through 2003 period, but no compelling explanation has been documented.[1] We consider the possibility that the explanation lies with the changing importance of "fringe" employment and associated discrepancies in the reporting of employment status between household-provided and employer-provided data that change in importance over time. 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 dataset containing CPS information together with UI wage records for the same individuals that is used in our empirical work. We then turn, in Section IV, to an examination of discrepancies between household-reported and employer-reported employment status, the factors that explain those discrepancies, and their implications for labor market analysis. Section V considers 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

---

[1] Earlier studies that have examined alternative explanations for the differing trends in the two series include Juhn and Potter (1999) and Nardone et al (2003).

the differing behavior of these aggregate series. We end with some concluding observations and

suggestions for future research.

II. Framework for Analysis

 We begin by asking why it is that discrepancies might arise between what individuals say

about their employment status and the information about their employment status found in

employer records. With respect to the classification of employment status for a particular

individual, four different outcomes are possible. A person may be recorded as having a wage

and salary job in neither data set (X_1); as having a wage and salary job in the employer-reported

but not the CPS data (X_2); as having a wage and salary job in the CPS but not the employer-

reported data (X_3); or as having a wage and salary job in both data sets (X_4). These outcomes are

illustrated in the table below. The number of employed people in the employer-provided data is

		Wage and salary job reported by employers	
		No	**Yes**
Wage and salary job reported in the CPS	**No**	X_1	X_2
	Yes	X_3	X_4

equal to $X_2 + X_4$; the number of employed people in the CPS data is equal to $X_3 + X_4$. Note that,

even if there are substantial discrepancies in the way that individuals are categorized in the

household and the employer data, the total number of employed persons in the two sources will

be similar if the magnitudes of X_2 and X_3 are similar.

We are most interested in the X_2 and X_3 cells and the reasons for the reporting discrepancies that place people in those cells. We hypothesize that people who hold jobs that they do not view as their main activity – positions that might be termed *marginal* – should be especially likely to show up in employers' records but not be reported during a household survey interview, that is, especially likely to appear in the X_2 cell. People who hold *off-the-books* jobs for which the employer does not keep formal records or, alternatively, people who do regular work for a firm but are considered by those firms to be *independent contractors or consultants* may report themselves as employed in a household survey interview but will not show up in employer records, making them likely to appear in the X_3 cell. To the extent that these hypotheses about how people end up in the off-diagonal cells are valid, they should have implications for the characteristics of both the workers and the jobs found in these cells.

When an individual who is attending school or an individual who has retired from a career job is asked whether they are working (or when someone else in their household is asked whether they are working), employment activity may not be reported because the individual considers him- or herself to be "a student" or "retired" rather than "a worker." This suggests that young adults and people near or past retirement age should be relatively more likely to be counted in the X_2 category. Whatever the characteristics of the job incumbent, we also would expect short-term, low-hour or low-earnings jobs to be more likely than other jobs that are present in administrative data not to be reported in a household survey, either because the people in these jobs do not consider working to be their primary activity or because they are not working during the CPS reference week.

Individuals who have earned income from an off-the-books job or from working as an independent contractor or consultant may regard themselves as employed, since they report to an

employer and perhaps work alongside other wage and salary workers. Money paid to an off-the-books worker typically will not appear anywhere in an employer's records. Independent contractors may receive earnings that are reported to the Internal Revenue Service on a Form 1099, but such earnings are not considered to be wages and are not subject to unemployment insurance taxes. Because they have no reported UI earnings, all such workers who tell the CPS interviewer that they are employed would be counted in the X_3 category. Those more likely to appear in the X_3 category might include, for example, workers with low education who may be more likely to work off the books, highly educated people who may be more likely to work as independent contractors, or anyone 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. To the extent that this sort of employment also tends to be less stable or less intensive, measures of job continuity, hours of work, and earnings also may help to explain individuals' presence in the X_3 cell.

Another way that household-reported and employer-reported data could differ is that, among those the two sources agree hold wage and salary jobs, some could have multiple jobs in one of the data sources but not in the other. We consider four possibilities. A person may be recorded as having a single wage and salary job in both data sets (Y_1), as having one wage and salary job in the CPS but more than one in the employer-reported data (Y_2), as having more than one wage and salary job in the CPS data but just one job in the employer-reported data (Y_3) or as having more than one wage and salary job in both data sets (Y_4). If all multiple job holders had

	Number of wage and salary jobs reported by employers	
	One	Two plus

Number of wage and salary jobs reported in the CPS	One	Y_1	Y_2
	Two plus	Y_3	Y_4

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 $X_2 + Y_2 + X_4 + Y_4$ and the total number of jobs in the household-provided data would equal to $X_3 + Y_3 + X_4 + Y_4$.[2] The difference between the employer-based and the household-survey-based job count thus will depend on the difference between $X_2 + Y_2$ and $X_3 + Y_3$. For the same reasons that we expect those whose primary jobs are short-lived or involve low hours or low earnings to be likely to appear in the X_2 or X_3 cell in the employment matrix, we also expect those whose second jobs have similar characteristics to be more likely to appear in the Y_2 or Y_3 cell in the number-of-jobs matrix.

---

[2] Because they exclude some jobs, these sums provide lower bound estimates of the total number of jobs. In the data we examine later in the paper, which cover the years 1996 to 2003, approximately 15 percent of persons who are categorized as employed in both the CPS and the UI data hold more than one UI job; on average, these individuals hold 2.2 UI jobs, not 2.0 jobs. We are not able to estimate reliably the number of jobs held by CPS respondents, so cannot provide a similar figure for CPS multiple job holders. In addition, there are some people who have no job in one data source but hold multiple jobs in the other. Individuals recorded as employed in the UI data but not employed in the CPS (the X_2 cell) account for less than 6 percent of persons with multiple jobs in the UI data; individuals recorded as employed in the CPS but not in the UI data (the X_3 cell) account for only about 10 percent of persons with multiple jobs in the CPS using a restrictive definition and 15 percent using a more lenient definition. Taking both sources of omission into account would increase the total number of UI jobs by about 3.5 percent and, assuming that CPS multiple job holders also hold an average of 2.2 jobs, increase the total number of CPS jobs by between 4.0 and 5.0 percent. While the sums shown thus understate actual employment by a few percentage points, they appear to track movements in actual employment very closely.

III. Data and Measurement

To examine the levels and composition of the various X's and Y's in the cells of the data tables just described, we require direct individual-level comparisons of employer-reported and household-reported employment status. We accomplish this by linking CPS records to unemployment insurance (UI) wage records for the same individuals. Given the growing interest in the use of UI wage records for labor market analysis and program evaluation, discrepancies between the CPS records and the UI records are of interest in their own right.[3] In addition, since the reports that employers provide in response to survey inquiries should be based on the same records that are used to prepare their UI wage reports, the patterns of any discrepancies we observe between the CPS and UI data may well be reproduced more broadly in discrepancies between employer survey and household survey data. The CES employment estimates, specifically, are benchmarked to the Quarterly Census of Employment and Wages (QCEW) employment counts, which derive from the reports employers must file along with payment of their UI taxes, so that inferences based on comparisons of CPS and UI wage record data seem quite likely to carry over to understanding the differences in behavior of the CPS and CES employment series.

In order to make direct comparisons between the CPS and UI data, we must be able to match workers in the two datasets. A linked individual-level dataset that includes both CPS and UI information has been constructed by the Census Bureau's Longitudinal Employer-Household

---

[3] Examples of studies that use UI wage record data for labor market analysis include Jacobsen, LaLonde and Sullivan (1993) and Schoeni and Dardia (1996) who study the impact of job displacement on workers' earnings; Abowd, McKinney and Vilhuber (2009), who study the impact of job displacement on human capital; Haltiwanger, Lane and Spletzer (2007), who study the relationship between earnings and productivity of workers; Brown, Haltiwanger and Lane (2006) who study the impact of economic turbulence on firm productivity and workers' career paths; Kornfeld and Bloom (1999), who measure the impact of participation in Job Training Partnership (JTPA) programs on low income individuals; and Cancian et al (1999), who examine the subsequent earnings experiences of individuals who exited Aid to Families with Dependent Children (AFDC).

Dynamics (LEHD) program.[4] To create this dataset, individuals' CPS records have been matched to their UI records using Protected Identity Keys (PIKs) that are based on Social Security numbers (SSNs). The data set used in our analysis covers 16 states over the period from 1996 through 2003.

Each quarter, employers in each of these states are required to report the quarterly earnings of each of their employees to the agency that administers the state UI system. Small agricultural employers typically are excluded from state UI 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). 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.[5]

We are able to look at employment in the first quarter of each of the years for which we have data. Although the CPS provides monthly employment information, the UI data contain quarterly employment and earnings records. Further, the PIKs that permit CPS records to be linked to the UI records are available only for March. Because we need to construct a quarterly employment measure from the CPS records that can be matched to the employment information contained in the UI records, our sample must be restricted to persons who responded to the CPS in January, February and March. Further, PIKs are missing for approximately 20 to 30 percent of March CPS respondents. The sample available for analysis consists of approximately 12 to 15 thousand individuals per year who are resident in the 16 states for which we have UI wage records. As explained in detail in the Data Appendix, we use propensity score methods to reweight our data to ensure that they represent the population of interest, at least with respect to

---

[4] A comprehensive description of the LEHD data infrastructure can be found in Abowd et. al. (2009).

their observable characteristics. For comparability with published CPS employment estimates, our sample is restricted to persons age 16 years and older.

Each of the quarterly wage records in the UI data for our matched sample identifies the employing business, the worker, and the earnings paid by that business to the worker over the quarter. Individuals with at least one report of positive in-scope earnings in the UI data during the first quarter are categorized as employed. Individuals for whom more than one employer reported in-scope earnings during the quarter are categorized as multiple job holders. We also create variables that capture the earnings on each worker's highest-earning job and, for those holding more than one job, the earnings on any additional jobs. In addition, by looking at whether the individual had earnings with a given employer in the previous quarter and/or the subsequent quarter, we are able to identify what we term long-duration jobs, by which we mean jobs that continued across at least two successive quarters.

Constructing quarterly employment measures using CPS data is more complicated than constructing the parallel UI employment measures because we cannot always be certain whether a second job reported by a CPS respondent is an in-scope job or whether a person who is employed in two successive months in the CPS had the same or a different employer. Given these uncertainties, we have constructed two measures of CPS multiple-job-holding, one based on applying more restrictive criteria than the other. We also construct several measures intended to capture the characteristics of the jobs that people hold. These 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. An additional variable captures whether the individual works in an industry and occupation with a high proportion of workers who are

---

[5] Note that, except for the exclusion of federal workers, our definition of in-scope employment is similar to the CES employment definition.

self-employed, which we take as an indicator of potential confusion in the reporting of employment status. The Data Appendix provides additional details about all of the UI and CPS variables used in the analysis.

We would like for the weighted estimates of employment in our 16 states based on the restricted linked sample just described to be consistent with estimates based on the full CPS sample and with estimates based on the complete UI wage records database for the same states. In Table 1, we compare employment estimates based on our linked sample to estimates based on the larger datasets, including the number of people employed in the CPS as of March, the number of people in the CPS employed at any point during the first quarter, the number of people for whom jobs are recorded in the UI data during the first quarter, and the number of jobs recorded in the UI data during the first quarter. The larger data set used for the CPS estimates consists either of everyone living in one of our 16 states who responded to the March CPS or to the January, February and March CPS, depending on whether we are looking at March employment or first quarter employment; the larger data set used for the UI estimates is the full set of wage records for our 16 states. As expected, in the CPS data shown in the top panel of the table, the number of people employed in at least one month of the quarter exceeds the number employed in March, by about 6 percent on average. The number of people who report holding at least one in-scope wage and salary job during the first quarter of the year is from 9 to 12 percent larger in the CPS data than in the UI data.

For both sets of CPS numbers, the estimated number of employed people is very similar whether we use our linked sample, which is restricted to those for whom a PIK is available and responded in all three months of the quarter, or the most inclusive sample for which the estimate in question can be produced. This is a reassuring if not surprising indication that our sample

estimation weights are performing as intended. The bottom panel of the table, however, reveals

that our linked sample produces UI employment estimates that fall slightly short of the actual

numbers calculated from the full UI wage records database. This is true both for estimates of the

number of employed people and for estimates of the number of jobs. The most likely

explanation for this discrepancy is that we are not finding all of the UI jobs held by those in the

linked sample, whether because the employer is located outside of the 16 states for which we

have UI wage records or because there is a problem with the PIKs used to carry out the necessary

record linkages. Although the discrepancy is not large, it has potential implications for our

empirical analysis that we consider further in the next section of the paper.

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 or not working in an in-

scope job during the first quarter of the year) and multiple-job-holding status (i.e., holding more

than one in-scope job during the first quarter of the year) in individual-level matched CPS-UI

data. The results we report are based on the pooled micro data for our 16 states over the years

1996 through 2003 and are weighted using adjusted CPS estimation weights as described in the

Data Appendix. In addition to looking at the number of people in the different X and Y cells

described earlier, we also explore the personal and job characteristics that predict membership in

the off-diagonal cells of these tables.

A. Discrepancies in Employment Status

We begin with Table 2, which summarizes the basic patterns of discrepancy in the categorization of individuals as employed versus not employed in an in-scope job in the CPS versus the UI data, using a simple two by two matrix. As a reminder, for our purposes, an in-scope job is defined as a non-agricultural private-sector wage and salary job, a state government job or a local government job. In discussing these tables, we refer to the X_1 through X_4 categories outlined earlier. X_1 workers, for instance, are workers who do not have an in-scope job in either the UI or the CPS. Overall shares, row shares, and column shares are presented in the table along with the associated standard errors.

On average over the eight year period covered by our data, 49.1 percent of individuals aged 16 and older are employed as an in-scope worker in both the CPS and the UI data during the first quarter of the year; 37.1 percent are not in-scope workers in either the CPS or the UI data; 3.4 percent are in-scope workers in the UI but not in the CPS; and 10.5 percent are in-scope workers in the CPS but not in the UI. Looking at conditional relationships, 6.4 percent of in-scope UI workers are not in-scope CPS workers and 17.6 percent of in-scope CPS workers are not in-scope UI workers. Given the large size of the pooled matched dataset, the standard errors of these estimates are low.

The share of CPS workers for whom we can find no UI job is very large. An obvious concern is that we almost certainly are missing some UI jobs held by people in our CPS sample because of problems with the matching of CPS records to the UI wage records. The incomplete geographic coverage of our linked data set is one reason for missing jobs – we have wage records for 16 states and search for jobs held by any CPS respondent across all of those states, but we do not observe UI jobs in the remaining 34 states and the District of Columbia. SSN

errors in either the CPS data or the UI data also would cause us to miss UI jobs held by people in the 16-state CPS sample. Any failure to identify a UI job held by a CPS respondent will raise the share of CPS workers with no UI job and also could affect the share of UI workers with no CPS job.[6] Because we appear to be missing only a relatively small number of UI jobs, however, we do not believe that taking them into account would produce any qualitative change in our conclusions.

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 in our 16 states. Calculations using the full UI wage records database show that the average number of people with first-quarter UI jobs (E_{UI}) in our 16 states over the years 1996-2003 is 55.503 million; the average *estimated* number of UI job-holders ($\hat{E}_{UI}$) based on the weighted values from our linked data set is 53.940 million. Based on the ratio of E_{UI} to $\hat{E}_{UI}$ (1.029), we should have found altogether about 2.9 percent more people with UI jobs than we actually did. If UI jobs held by those with and without CPS jobs are equally likely to have been missed, our estimate that 6.4 percent of UI job holders report no CPS job would be unaffected by accounting

---

[6] More specifically, if we are missing UI jobs held by people who have no CPS employment, X_1 will be too large and X_2 too small; if we are missing UI jobs held by people who are employed in the CPS, X_3 will be too large and X_4 too small. Any adjustment made to account for missing UI jobs thus will lower X_3 and raise X_4, unambiguously lowering X_3 as a share of CPS employment. The effect on X_2 as a share of UI employment will depend on whether the adjustment raises X_2 by proportionally more than, in the same proportion as or proportionally less than it raises X_4.

for missed UI jobs, but the share of CPS job holders with no UI job would fall from 17.6 percent to 15.3 percent.[7] This is worth noting, but does not change the qualitative story told by Table 2.

Seeking further evidence about the potential importance of missing UI jobs, we also replicated our analysis using data only for the three largest states in our linked sample, California, Florida and Texas. According to data from both the 2000 Census and the 2005 American Community Survey (ACS), only about ½ percent of California residents and just over 1 percent of Florida and Texas residents say they work in a different state, and it seems reasonable that, in a large state, a smaller share of residents should commute across state lines to jobs in another state than would be the case in a small state. Results based on data for these three states were very similar to those we report in the text. With regard to keypunching or reporting errors in Social Security numbers (SSNs), as described in the Data Appendix, only validated SSNs reported by CPS respondents are retained to form the PIKs we use 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. While none of this additional evidence is definitive on its own, taken as a whole it seems to corroborate the conclusion from the calculations reported above that missing UI jobs cannot explain the patterns we observe in Table 2 and in particular cannot account for more than a small portion of the many CPS workers with no UI job.

---

[7] Some of the employed people found in the UI wage records for our 16 states likely were residents of one of the other 35 jurisdictions (34 states plus the District of Columbia) for which we do not have data. The implicit assumption underlying our calculations is that the number of people who live in our 16 states but have UI jobs in one of the other 35 jurisdictions 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. Since each set of states accounts for very close to half of national employment over the period we study, we see this as a reasonable approximation. To the extent that those with UI jobs outside of their half of the country also have a UI job inside those boundaries, our calculations will overstate the biases resulting from missed UI jobs.

B. Characteristics of X_2 and X_3 workers

In addition to knowing the relative sizes of the different cells in the employment status matrix, we also would like to know something about the personal and job characteristics that affect the probabilities of being found in the off-diagonal X_2 or X_3 cells. The linear probability models reported in Table 3 and Table 4 address these questions. [8] Results are reported both for models that include only demographic characteristics and models to which measures of job characteristics have been added. We interpret both sets of coefficient estimates with reference to whether characteristics seem likely to be associated with a job being either a marginal job (in the case of the X_2 models) or an off-the-books or Form 1099 job (in the case of the X_3 models). In addition to the demographic and job characteristic variables, all of the models also include year dummies.

The same demographic controls are used in both tables – age, education, sex, marital status, race, an indicator for whether the individual is foreign born, and an indicator for whether the individual is the CPS respondent in the household. Somewhat different information about job characteristics is available for UI jobs and CPS jobs. As predictors of whether a person who has at least one in-scope UI wage record reports no in-scope CPS job, Table 3 includes whether the individual held any long UI job during the quarter; whether the individual held two or more UI jobs during the quarter; and the level of earnings on the individual's primary UI job. Our expectation is that the stable UI jobs with higher earnings should be less likely to be missing in the CPS. As predictors of whether an individual reported as working in the CPS has no UI job, Table 4 includes measures of 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); whether

---

[8] Linear probability model coefficients are reported for ease of interpretation; probit models yield very similar marginal probability effects.

the respondent works in an occupation and industry with a high percentage of self-employed workers; whether any of the CPS jobs held during the quarter were full time jobs; and a set of earnings dummies based on earnings in the main job for those who report employment in their outgoing rotation month (March or April). We expect work interruptions, working in jobs commonly held by contractors or consultants, or working in a part-time position to increase the likelihood that someone who reports a CPS job has no corresponding UI wage record. Our prior with regard to earnings is less clear – we would expect off-the-books workers typically to have low earnings, but independent contractors and consultants could be quite highly paid.

The first column of Table 3 reports results for models of the probability of being an X_2 worker – having a job in the administrative data but no job in the CPS data – that include only demographic factors. Among these variables, being age 65 or older has the largest effect – all else the same, with no job characteristic controls, being in this age group raises the probability of being in the X_2 category by 14 percentage points, and the effect remains large even after job characteristics are added to the model. This is consistent with our expectations, in that older workers may think of themselves as retirees rather than workers, but may nonetheless have some UI earnings. In the model in the second column that includes job characteristics, having very low earnings raises the probability of being an X_2 worker, while having a long-lasting job or more than one UI job reduces the probability of being an X_2 worker.

Table 4 reports results for similar models of the probability of being an X_3 worker – having an in-scope CPS job, but no UI earnings. As in the models for the reverse situation, those over age 65 are especially likely to be found in the off-diagonal cell, though again the effect is weakened by the introduction of job characteristic variables in the second column. Other demographic characteristics that raise the probability of membership in the X_3 cell are being

male, non-black, or foreign born and having either a low or a very high level of education, all of which are insensitive 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, male and having a very high level of education), but the pattern of estimated coefficients as a whole is very consistent with our expectations. The job characteristic variables included in the second column of Table 4 add substantial additional explanatory power to the model. CPS workers who did not work all three months of the quarter, are in an industry and occupation with a high proportion of self-employed workers, had a very low earnings job or reported no earnings in the CPS outgoing rotation month have a substantially higher probability of being absent from the UI wage records. The job characteristic coefficient estimates in this model would seem primarily to be capturing the presence of off-the-books jobs, which one would expect to be more intermittent and lower-paying than other employment, in the X_3 cell.

C. Discrepancies in Multiple-Job-Holding Status

In Table 5, we explore discrepancies in multiple job status, looking at employed individuals who report multiple jobs in the UI but not the CPS data (the Y_2 group) or who report multiple jobs in the CPS but not the UI data (the Y_3 group). Because of the difficulty of identifying workers in the CPS who held more than one job during the 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 the Data Appendix, we have defined alternate indicators of holding multiple CPS jobs. The results reported in the text make use of

the more restrictive of the two sets of criteria for identifying multiple job holders in the CPS, but our main findings are not affected which set of criteria we apply. The at-risk group in Table 5 is individuals who hold both a CPS job and a UI job – in other words, the X_4 group from Table 2. Some 81.3 percent of such workers have just one job both in the CPS and in the UI, 4.6 percent have two or more jobs both in the CPS and in the UI, 10.4 percent have two or more UI jobs but only one CPS job and 3.7 percent have two or more CPS jobs but only one UI job. Row and column percents are again instructive. Some 69.2 percent 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 percent of workers have only one in-scope UI job. In both cases, the conditional discrepancy is quite large. Although Table 5 is based on fewer observations than Table 2, the standard errors for the estimates reported remain small.

Using a less restrictive approach to decide whether a CPS worker has more than one job roughly doubles the share of CPS workers who are categorized as multiple job holders (or at least possible multiple job holders). In our data, applying the less restrictive criteria reduces the share of UI multiple job holders who have just one CPS job from 69.2 percent to 53.9 percent (making Y_2 smaller) and raises the share of CPS multiple job holders who have just one UI job from 44.6 percent to 58.3 percent (making Y_3 larger). The overall share of those with more than one job in the CPS data is much more similar to that in the UI data when the less restrictive criteria are used. Most importantly, however, whichever set of numbers one looks at, there is considerable disagreement between the two data sources about which individuals hold more than one job over the quarter.

D. Characteristics of Y$_2$ and Y$_3$ workers

We now turn to an examination of the factors associated with discrepancies in multiple job holding status. Table 6 reports the results of models that seek to identify the factors associated with holding multiple UI jobs but only a single CPS job (that is, being in the Y$_2$.cell).. The model in the first column contains only demographic factors; characteristics of jobs beyond the first job are added in the second column. For those with multiple UI jobs, the likelihood of having only a single CPS job is lower for more educated and married workers, and higher for black and foreign born workers. We do not have strong priors about the coefficients on these variables, but the positive (negative) coefficients could be associated with holding short-term UI jobs that are not picked up during the CPS reference weeks. With regard to the job characteristic variables added in the second column, the most notable finding is that workers with the highest-paying second UI jobs are much more likely than those with lower-paying second jobs to be Y$_2$ workers. This is not a pattern we had expected. It might be, for example, that individuals in some professions earn significant supplemental earnings that are not regarded by the worker as constituting a second job. We can imagine that people in certain types of jobs– for example, actors, skilled construction trades workers, college faculty, or even doctors or lawyers – might regard themselves as having one job but in fact have income that derives from multiple sources. Results very similar to those reported in Table 6 are obtained when the less restrictive criteria for identifying multiple job holders in the CPS are applied.

Table 7 reports results for models of the probability that CPS multiple job holders hold only one UI job (that is, are found in the Y$_3$ cell). Personal characteristics that raise this probability include being highly educated, age 55 or older (though 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_3 workers. These findings seem most

consistent with the existence of off-the-books employment that is reported in the CPS but for

which no UI wage records exist. Much the same coefficient patterns hold when the more

expansive definition of CPS multiple job holders is applied.

E. Implications for Cross-Sectional Labor Market Analysis

Our results thus far provide compelling evidence of substantial discrepancies between

employment as measured in the CPS and employment as measured in UI wage records even

when the scope of included employment is defined to be the same in both data sources. While

there are sizeable discrepancies of both sorts, it is notable that the share of CPS wage and salary

workers who hold jobs in the private sector, state government or local government for whom no

UI wage record can be identified is substantially larger (17.6 percent) than the share of similar

UI workers for whom the CPS records no corresponding employment (6.4 percent). On net, CPS

estimates of the number of people with in-scope wage and salary jobs are significantly higher

than the corresponding number based on UI data. According to the data reported in Table 1, on

average over the eight years for which we have data, the number of people holding in-scope jobs

in the CPS is more than 10 percent larger than the number holding in-scope jobs in the UI data.

Making an adjustment for UI jobs we may have missed reduces this differential somewhat, but it

remains large.

The finding that there are many CPS workers for whom no UI earnings are reported is consistent with evidence for selected states that employers misclassify a significant share of their employees as independent contractors or pay them off the books. Based on data from in-depth audits of wage record reports filed in 1987 by a randomly-selected sample of 875 Illinois employers, Blakemore, Burgess and Low (1996) conclude that no UI taxes were paid for 13.6 percent 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 worker classification issue, a series of more recent studies that use data from employer audits states must conduct to satisfy federal requirements conclude that there continues to be significant misclassification of workers as independent contractors. Carre and Wilson (2004), for example, find that about 5 percent of all workers in Massachusetts were misclassified as independent contractors over the 2001-2003 period; Belman and Block (2008) report a similar figure for Michigan based during 2003 and 2004. To the extent that the state audits cannot identify off-the-books workers for whom there is no paper trail, these figures understate the extent to which employer reports omit workers who should be classified as employees.

For those who may wish to use either CPS data (or other household survey data) or UI wage record data (or other employer-reported information) for labor market analysis, the differences in how individuals are classified with respect to their employment status may be highly significant. We have already 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 systematic nature of the discrepancies implies that the associated classification error is not classical measurement error but correlated with characteristics that may

be the focus of the analysis. For example, a study focusing on labor market outcomes for low wage workers with low job stability is likely to be more sensitive to such classification error relative to an analysis focused on high wage workers with high job stability.

For a labor market analyst, a closely related but different question is whether the estimated impact of labor market shocks or policies on labor market outcome depends on whether the study uses household or employer based data. This issue depends not only on the measurement error discussed above but, in the current context, how the factors that predict UI employment differ from those that predict CPS employment. If the same personal characteristics affect the propensity to be in the two off-diagonal groups either in a different direction or to a different degree, then the conclusions to be derived based on one data set rather than the are likely to differ in important ways. To address this question, using data for our full linked sample, we have fit models that relate being employed in either the UI data or the CPS data to individuals' personal characteristics.

Table 8 reports the marginal effects of personal characteristics on the probability that a person in the matched dataset is employed in the UI data and in the CPS data. We already know from the findings in Table 2 that the probability of having a CPS job is substantially higher than the probability of having a UI job. Table 8 provides the information necessary to quantify how these probabilities vary by personal characteristics. The Table 8 results show that prime age, educated, male, married, white, and U.S. born individuals are more likely to be employed both in the UI and in the CPS data, but there are notable differences in the magnitude of the estimated effects. For example, the positive marginal effect of having an advanced degree on employment is about 5 percentage points larger in the CPS than in the UI data, and the negative effect of

having less than a 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 vary markedly according to the individuals' personal characteristics. To take two extreme examples, for a black female aged 16-24 with less than a high school education, the predicted probability of employment is very similar across the two data sources (41 percent based on the CPS model versus 42 percent based on the UI model). On the other hand, for a white male aged 35-54 with an advanced degree, the predicted probability of employment is 84 percent based on the CPS model but just 69 percent based on the UI model.[9] These differences imply that, 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.

These same issues are relevant for analyses of the effects of policy (e.g., welfare policy or job training policy) or shocks (e.g., displacement) on labor market outcomes. Consider, for example, the standard treatment effect model:

$$Y_{it} = X_{it}^{'}\beta + \alpha T_{it} + (X_{it} * T_{it})'\delta + \varepsilon_{it}$$

where the outcome of interest for person i in time t is Y_{it}, X_{it} reflects person characteristics and T_{it} is the treatment. It is often of interest to include interactions of treatment with characteristics as treatment impacts are likely to vary with person characteristics. The impact of the treatment effect on the outcome is given by:

$$dY_{it} / dT_{it} = \alpha + X_{it}^{'}\delta$$

---

[9] Both calculations assume that all other characteristics other than those identified are evaluated at the mean of their empirical distribution in the data.

With measurement error in the outcome that is correlated with person characteristics the estimate

of the treatment effect likely will be biased. Our findings in Tables 2 through 7 provide

compelling evidence that such concerns are relevant for outcomes such as employment status or

the number of jobs. In addition, the findings in Table 8 imply that the estimate of the same

treatment effect might be quite different across studies using household and employer based data.

V. Aggregate Time Series Patterns in the Discrepancy Between Household and Employer Based Employment

The analysis in the previous section shows 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. In this section, we consider the aggregate time series implications of these

individual-level findings.

A. Are Aggregate Discrepancies Cyclical? Evidence and Conceptual Underpinnings

We begin by recalling the rather different time series behavior of CPS and CES

employment over the period from 1998 through 2003 that was displayed in Figure 1. Even after

the CPS data have been adjusted to be more comparable 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 divergences between

CES and CPS employment are a cyclical phenomenon. Figure 2 displays a 60 year history of the

ratio of CES employment to CPS nonagricultural wage and salary employment.[10] 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. The clear message of Figure 2 is that whatever story we want to 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_2 and/or Y_2 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 5 temporary staff over the Christmas holidays rather than 2 or 3 such people. 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_2 as the economy tightens (i.e., to expect that X_2 will be procyclical). A similar dynamic might be in play for Y_2. Note that increases in X_2 and/or Y_2 should be associated with an increase in CES employment relative to CPS employment.

---

[10] Among the conceptual differences between the CPS and the CES employment series are that the CPS includes the unincorporated self-employed, unpaid family workers, agricultural and related workers, private household workers, and workers who have a job but were not paid during the survey reference week, all of whom are excluded from the CES. In addition, the CPS counts the number of employed people, whereas the CES counts the number of jobs. To produce the adjusted CPS employment series shown in Figure 1, employment among the various groups excluded from the CES is subtracted from the published CPS employment figures and then, to account for the fact that some CPS respondents hold more than one job, the number of multiple job holders is added to put the CPS employment figure on a jobs basis. The largest adjustments are those for the self-employed and multiple job holding. See Bowler and Morisi (2006) for further details. In the longer historical series shown in Figure 2, it is not possible to make all of these adjustments, but the self-employed, unpaid family workers and agricultural and related workers have been excluded from the CPS series. The most important adjustment not made in Figure 2 is that for multiple job holding.

Off-the-books or Form 1099 employment also may be cyclical. Suppose that, as economic activity strengthens and labor markets become tighter, people tend to leave informal jobs (jobs not recorded on employer payrolls) for formal jobs (jobs that employers report). 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 tightened, we might expect X_3 to fall. In addition, to the extent that those affected are multiple job holders, we might expect Y_3 to fall. Put another way, we might expect both X_3 and Y_3 to be countercyclical. Decreases in X_3 and/or Y_3 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 CPS and UI employment series derived from the linked data set behave similarly to the published CPS and CES employment series. The two sets of numbers are not strictly comparable. The linked sample includes data only for sixteen states and, to match the coverage of the UI wage records, we have restricted our attention to jobs in the private sector, state government or 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 a much smaller sample 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 employment estimate between the first quarter of 1998 and the first quarter of 2001, and then falling towards the CPS employment series. Additional

details may be found in the Data Appendix, but the similarity of the patterns observed in employment estimates based on our linked sample to those based on published estimates gives us reason 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 Jobs

In this subsection, we turn to the question of whether person and job characteristics that help account for the cross sectional variation in the off diagonals of the X and Y matrices also can help account for the aggregate time series patterns. While we acknowledge that our matched sample is not ideally suited to address this question, given that we have only eight years of data covering one cyclical episode, we nonetheless believe that there is useful information to be gleaned from an investigation of changes in the number of marginal (X_2) and off-the-books or Form 1099 (X_3) jobs over this period.

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_2) and in the number of people employed in the CPS but not in the UI (X_3). The X_3 series is larger in magnitude and considerably more volatile. What matters for the gap between the employer- and household-based employment estimates is the relative trends in X_2 and X_3 over time. All else the same, any increase in X_2 relative to X_3 will be associated with growth in UI employment relative to CPS employment; conversely, any decrease in X_2 relative to X_3 will be associated with a decline in UI employment relative to CPS employment. To the extent that X_2 and X_3 grow or shrink together, the gap between the UI and CPS employment series will not be affected.

Over the period from 1996 to 2001, X_2 and X_3 fluctuate relative to one another, but not in a consistent fashion. Over the 2001 to 2003 period X_3 grew by about 900,000 workers while X_2 fell by about 300,000 workers, both movements that would have contributed to the relative increase in CPS employment over this period. The modest decline in X_2 is consistent with a shrinking number of marginal (short duration or low earnings) jobs. The larger increase in X_3 is consistent with more marked growth in the number of off-the-books or independent contractor 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_2) 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_3).[11] The trend line for the Y_3 series is relatively flat but the behavior of the Y_2 series is more interesting, showing that the number of people in the Y_2 category grew markedly between 1996 and 1999 and leveled off thereafter.

Here again what matters for the gap between employer- and household-based employment estimates is whether the difference between Y_2 and Y_3 is changing over time. The gain in the number of people holding multiple jobs in UI but not in CPS (Y_2) from 1996 to 1999 is substantial. Y_2 grows by about 1.6 million people. In contrast, the number of people holding multiple jobs in CPS but not in UI (Y_3) grows by about 200,000. The difference between Y_2 and Y_3 grows substantially over the 1996 to 1999 period — by about 1.4 million people – 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 to 1999 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. Figure 5 includes the actual combined counts and predicted combined counts (where the latter are explained further below). In terms of the actual combined off-diagonal elements, Figure 5 shows the estimate of the jobs in the CPS not in the UI (X_3+Y_3) and an estimate of the jobs in the UI but not in the CPS (X_2+Y_2).[12] It is apparent in Figure 5 that the number of jobs in the UI but not found in the CPS grew more rapidly that the number of jobs in the CPS but not found in the UI over the 1996 to 2001 period. Jobs counted in the UI but not found in the CPS grew by 2.3 million while the number of jobs counted in the CPS but not in the UI grew by about 600,000. These patterns correspond well with the patterns in Figure 1 showing that employer based employment statistics grew more rapidly over this period relative to household based statistics. In contrast, from 2001 to 2003, Figure 4 shows that the number of jobs found in the CPS but not in the UI rose modestly while the number of jobs found in the UI but not in the CPS fell. Jobs counted in the CPS but not found in the UI grew by 800,000 over this period while jobs counted in the UI but not found in

---

[11] As in the previous sections, we focus the analysis on the more restrictive definitions of multiple job holding in the CPS. However, we have conducted the analysis in this section using the less restrictive definition and the main conclusions in this section are robust to the use of this alternative.

the CPS fell by about 500,000. Again, this pattern corresponds well with the pattern in Figure 1 that shows that over this time period employment from employer based data fell relative to employment from the household data.

C. Simulating the Fluctuations in Observed Employment

We now turn to exploring whether changes in the composition of marginal workers and marginal jobs over time can account for the patterns in Figures 3, 4, and 5. To explore this question, we use information on the composition of workers and jobs in each year together with the estimated coefficients from the linear probability models to simulate aggregate values for X_2, X_3, Y_2 and Y_3. The simulated shares of employment falling into the various X and Y categories are computed using the intercept term and the average of the year dummy coefficients from the relevant linear probability model plus the vector product of the characteristic coefficient estimates and the annual characteristic values. It is important to emphasize that we do not use the variation in the estimated year effects (rather only their average) since then the predicted and actual shares in each of the off-diagonal categories by year would be equal by construction.

To simulate the number of workers in the different X and Y categories, we multiply these simulated shares by the number of UI or CPS workers, as appropriate, or by the number of UI or CPS multiple job holders. Our simulated value for X_2, for example, equals the predicted probability that a UI worker is not a CPS worker times the number of UI workers in the year in question. In each case, the coefficient estimates employed are taken from the model that includes both demographic characteristics and job characteristics. Another point to note is that, rather than using our linked sample to compute the shares of worker and job types by year, we

---

[12] Note that this is an underestimate of the true combined totals since the estimates for Y_2 and Y_3 are conditional on the person having at least one job in both the CPS and UI data. A small fraction of individuals have multiple jobs in

use the possible largest dataset for each characteristic variable. Specifically, for the demographic characteristics 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. The reason for using these larger databases is to reduce the sampling variation in the simulated X and Y values; one consequence is that the average of the simulated values will not necessarily equal the average of the predicted values, as would be the case if we had used sample-based characteristic values.

Figure 5 displays the predicted values calculated as just described. The simulated and actual values need not behave especially similarly – it would be entirely possible for the factors that explain the variation in the number of people in the off-diagonal cells to have been captured by the year dummies in the models rather than by the demographic and job characteristic variables. The simulated values are somewhat smoother than the actual values, which is not especially surprising given that the simulation process eliminates random variations. The main message of the figure is that all of the key patterns in the raw data – from 1996 to 2001, the increase in the number of jobs in the UI but not in the CPS, and, from 2001 to 2003, the increase in the number of jobs in the CPS but not in the UI and fall in the number of UI jobs not found in the CPS – are capture reasonably well by the predicted values.

---

the CPS and no jobs in the UI and vice versa. See footnote 2 for details.

VI. Concluding Remarks

Using a large matched data set of CPS respondents to UI wage records, we find large discrepancies in employment status and in the number of jobs for the same set of individuals. Our basic results show that 17.6 percent of CPS workers who are working in a job that should be covered in the UI cannot be found and that 6.4 percent of UI workers who should be reported as working in the CPS are missing. Even larger discrepancies are found for multiple job holders. Some 69.2 percent of multiple job holders in the UI have only one job in the CPS and 44.6 percent of multiple job holders in the CPS only have one job in the UI.

We find further that the large off-diagonal discrepancies are not random but 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 would not consider their main activity and thus might fail to report. CPS workers not found in the UI data have person and job characteristics consistent with off-the-books or independent contractor work relationships. Similar remarks apply to discrepancies for multiple job holders. Taken together, our findings support the hypotheses we develop that 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. For microeconomic analysis, the findings identify types of people and types of jobs for which analysis of labor market outcomes is likely to be especially sensitive to the associated classification error. Moreover, the estimated impact of events and policy on labor market outcomes is likely to 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. With regard to aggregate measures of employment and their different movements

over time, we find some suggestive evidence that understanding these discrepancies lies in

understanding the cyclicality of marginal workers and marginal jobs.   Our quantitative analysis

in this latter respect is only suggestive since our analysis of the aggregate time series patterns is

based on only one cyclical episode. Since the microeconomic findings point to workers and jobs

on the margin as being the most subject to discrepancies, however, it makes intuitive sense that it

should be precisely marginal workers and marginal jobs that are the most sensitive to changing

economic conditions.

Together with other available information, the linked dataset we have developed should

be useful for a variety of labor market analyses. For example, the linked data can be used to

study the factors associated with non-receipt of unemployment benefits, and in particular,

whether lack of qualifying work experience or some other factor are responsible for so many

unemployed individuals not collecting unemployment benefits. For this and other analytic

purposes, it would be desirable to extend the time series of matched CPS and UI data and to

update these series on an ongoing basis. This would particularly useful for developing a better

understanding of differences in the cyclical pattern of aggregate employment estimates based on

household-reported versus employer-reported data. In addition to lengthening the time period

covered by the matched dataset, extending its geographic coverage also would be desirable.

Among other advantages, a national database should yield sufficient regional variation that

cyclical patterns could be investigated taking advantage of differences in the cycle across

regional labor markets.

.

References

Abowd, John M. and Martha H. Stinson. 2005. "Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and Administrative Data." Unpublished working paper. Available at http://courses.cit.cornell.edu/jma7/abowd-stinson-200501.pdf (accessed February 21, 2009). NOT CURRENTLY CITED

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* (Dunne, Jensen and Roberts, ed), Chicago: University of Chicago Press.

Abowd, John M., Kevin L. McKinney, and Lars Vilhuber. 2009. "The Link Between Mass Layoffs, Human Capital and Firm Deaths" in *Producer Dynamics* (Dunne, Jensen and Roberts, ed), Chicago: University of Chicago Press.

Belman, Dale L. and Richard Block. 2008. "The Social and Economic Costs of Economic Misclassification in Michigan." Unpublished manuscript.

Blakemore, Arthur E., Paul L. Burgess and Robert D. St. Louis. 1996. "Employer Tax Evasion in the Unemployment Insurance Program." *Journal of Labor Economics*, April, 14(2), 210-230.

Bowler, Mary and Teresa L. Morisi. 2006. "Understanding the employment measures from the CPS and CES survey." *Monthly Labor Review*, February, 23-38.

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*, vol.9, no. 1, pp. 1-24.

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*, vol.11, no. 3, pp. 345-368.

Brown, Clair, John Haltiwanger and Julia Lane. 2006. *Economic Turbulence: Is A Volatile Economy Good For America?*, Chicago: University of Chicago Press.

Bureau of Labor Statistics. 1997. "Quality Improvement Project: Unemployment Insurance Wage Records." Unpublished report. October.

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.

Cancian, Maria, Robert Haveman, Thomas Kaplan and Barbara Wolfe. 1999. "Post-Exit Earnings and Benefit Receipt Among Those Who Left AFDC in Wisconsin." University of Wisconsin Institute for Research on Poverty Special Report No. 75. Available online http://www.irp.wisc.edu/publications/sr/pdfs/sr75.pdf (accessed February 28, 2009).

Haltiwanger, John, Julia Lane and James Spletzer. 2007. "Wages, Productivity, and the Dynamic Interaction of Businesses and Workers." *Labour Economics,* vol. 14(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*, Washington, DC: National Academy Press.

Jacobsen, Louis S, Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review*, September, 83(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, December, 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*, vol.17, no. 1, pp. 168-197.

Mellow, Wesley and Hal Sider. 1983. "Accuracy of Response in Labor Market Surveys:

    Evidence and Implications." *Journal of Labor Economics*, vol.1, no. 4, pp. 331-344.

Nardone, Thomas, Mary Bowler, Jurgen Kropf, Katie Kirkland and Signe Wetrogan. 2003.

    "Examining the Discrepancy in Employment Growth Between the CPS and the CES."

    Paper prepared for presentation to the Federal Economic Statistics Advisory Committee.

    October.

Roemer, Marc. 2002. "Using Administrative Earnings Records to Assess Wage Data Quality in

    the March Current Population Survey and the Survey of Income and Program

    Participation." LEHD Technical Paper #TP-2002-22. Available at

    http://lehd.did.census.gov/led/library/techpapers/tp-2002-22.pdf (accessed February 21,

    2009).

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.

Schochet, Peter Z., John Burghardt, and Sheena McConnell. 2008. "Does Job Corps Work?

    Impact Findings from the National Job Corps Study."   *American Economic Review*,

    vol.98, no. 5, pp. 1864-1886.

Stevens, David W. 2007. "Employment That Is Not Covered by State Unemployment Insurance

    Laws." LEHD Technical Paper #TP-2007-04. Available at

    http://lehd.did.census.gov/led/library/techpapers/tp-2007-04.pdf (accessed February 24,

    2009).

Figure 1: Household and Payroll Survey Employment, Seasonally Adjusted, 1994-2009

Numbers in thousands
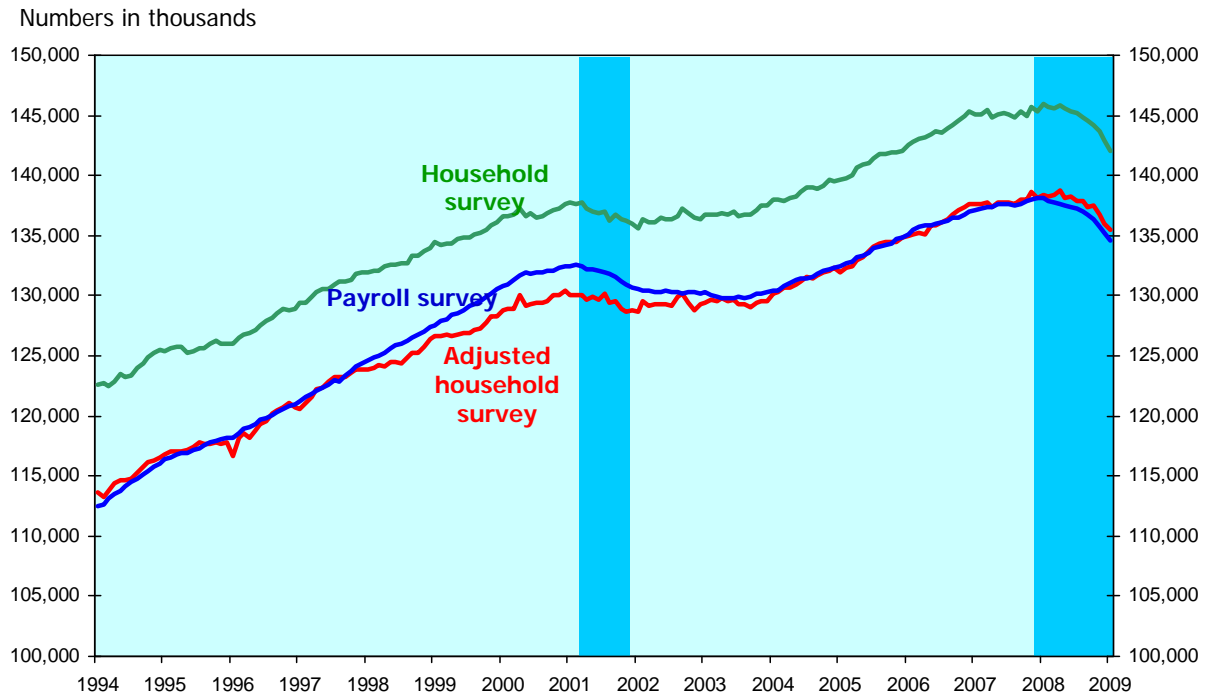
NOTE: The household series presented here has been smoothed for population control revisions. The "adjusted" household series has been adjusted to an employment concept more similar to the payroll survey's and smoothed for population control revisions. Shaded areas represent recessions as determined by the National Bureau of Economic Research (NBER). NBER has not yet determined an endpoint for the recession that began in December 2007.

SOURCE: Bureau of Labor Statistics, February 6, 2009.

Figure 2: Ratio of Establishment Survey Employment to Household Survey Nonagricultural Wage and Salary Employment, 1948-2004
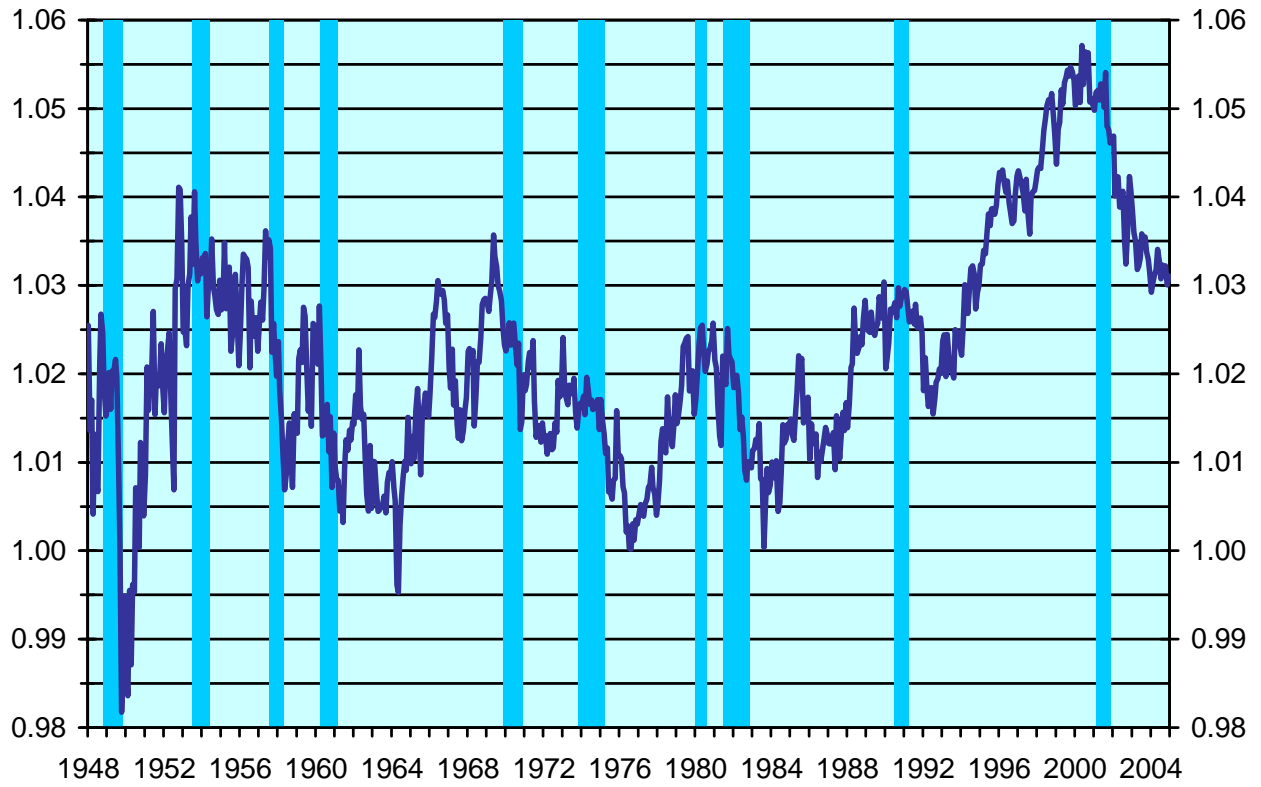
Source: Bowler and Morisi, 2006

Figure 3: Estimated Number of People in Off-Diagonal In-Scope Employment Cells, 1996-2003 (in millions)

	X2: Employed in UI, not in CPS	X3: Employed in CPS, not in UI
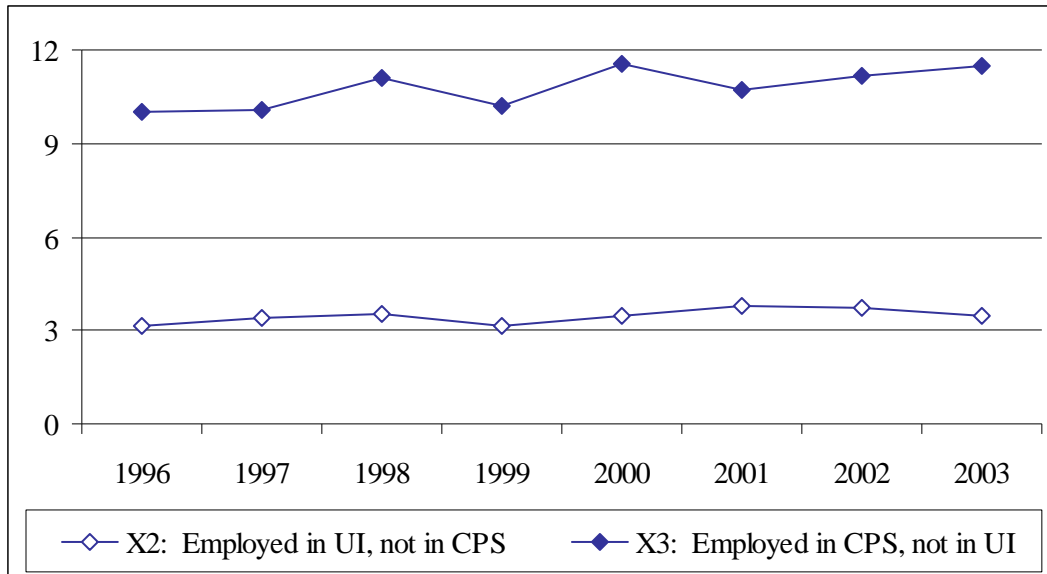
Figure 4: Estimated Number of People in Off-Diagonal Multiple Job Status Cells, 1996-2003 (in millions)

	Y2: Multiple Job Holder in UI, 1 Job in CPS	Y3: Multiple Job Holder in CPS, 1 Job in UI
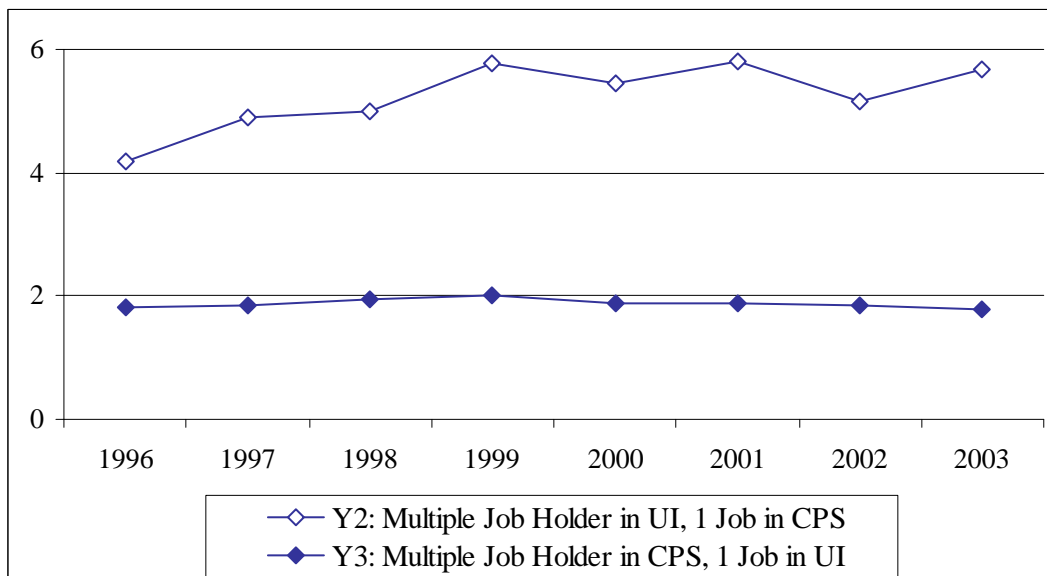
Figure 5: Estimated Number of Jobs in Combined Off-Diagonal Cells (Number of Persons Working plus Number of Multiple Job Holders), 1996-2003 (in millions)
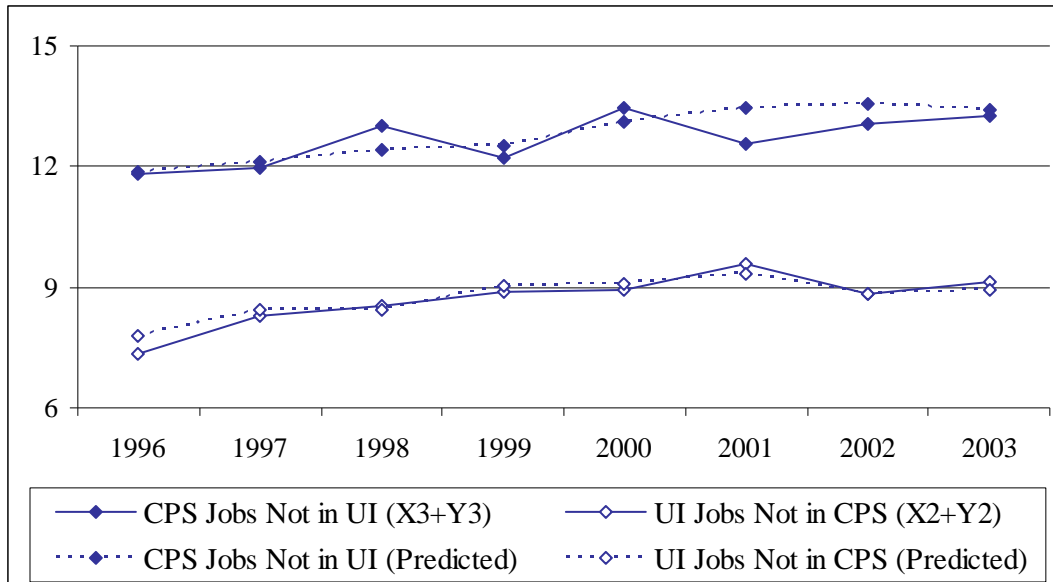
Table 1: Counts of Employed People and Number of Jobs,
CPS Full Sample and UI Universe versus Weighted Linked Sample

Panel 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%

Panel 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%

Table 2: 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		
Overall Share	0.371	0.034
	(0.001)	(0.000)
Row Share	0.917	0.083
	(0.001)	(0.001)
Column Share	0.779	0.064
	(0.001)	(0.001)
In-Scope Worker in CPS		
Overall Share	0.105	0.491
	(0.000)	(0.001)
Row Share	0.176	0.824
	(0.001)	(0.001)
Column Share	0.221	0.936
	(0.001)	(0.001)

Note: Numbers shown are weighted shares of the CPS-UI overlap sample
described in the text. Standard errors are shown in parentheses.
The upper left quadrant corresponds to the X1 group, the upper right quadrant to
the X2 group, the lower left quadrant to the X3 group, and the lower right to
the X4 group.

Table 3: Effects of Person and Job Characteristics on the Probability that a UI Worker is not a CPS Worker (X2)

Age 16 to 24	0.0511	**	-0.0205	**
	(0.0033)		(0.0032)	
Age 25 to 34	0.0026		-0.0047	
	(0.0025)		(0.0024)	
Age 55 to 64	0.0257	**	0.0190	**
	(0.0036)		(0.0034)	
Age 65 plus	0.1442	**	0.0879	**
	(0.0066)		(0.0062)	
Less than High School	0.0474	**	0.0066	*
	(0.0034)		(0.0032)	
Some College	-0.0028		-0.0003	
	(0.0026)		(0.0024)	
College Graduate	-0.0107	**	0.0017	
	(0.0030)		(0.0029)	
More than College	-0.0003		0.0099	*
	(0.0041)		(0.0040)	
Male	-0.0088	**	0.0055	**
	(0.0020)		(0.0019)	
Married	-0.0012		-0.0094	**
	(0.0024)		(0.0022)	
Black	0.0277	**	0.0215	**
	(0.0035)		(0.0033)	
Other Non-white	0.0161	**	0.0088	*
	(0.0047)		(0.0043)	
Foreign Born	0.0149	**	0.0259	**
	(0.0031)		(0.0029)	
Non-proxy Interview	0.0044		0.0063	**
	(0.0024)		(0.0022)	
Any Long Jobs	--		-0.2013	**
			(0.0061)	
Two or More UI jobs	--		-0.0175	**
			(0.0027)	
Qtr UI earn < $1K	--		0.2803	**
			(0.0036)	
$1K <= Qtr UI earn < $2.5K	--		0.0571	**
			(0.0031)	
$12.5K <= Qtr UI Earn < $25K	--		-0.0118	**
			(0.0031)	
Qtr UI earn >= $25K	--		-0.0001	
			(0.0058)	
Observations	56,027		56,027	
R-squared	.024		.160	

* significant at 5% level; ** significant at 1% level

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. Standard errors in parentheses.

Table 4: Effects of Person and Job Characteristics on the Probability that a CPS Worker is not a UI Worker (X3)

Age 16 to 24	0.0043		-0.0401	**
	(0.0050)		(0.0051)	
Age 25 to 34	-0.0167	**	-0.0179	**
	(0.0037)		(0.0037)	
Age 55 to 64	0.0336	**	0.0227	**
	(0.0053)		(0.0053)	
Age 65 plus	0.1423	**	0.0792	**
	(0.0096)		(0.0096)	
Less than High School	0.0345	**	0.0075	
	(0.0051)		(0.0051)	
Some College	0.0052		0.0046	
	(0.0039)		(0.0039)	
College Graduate	0.0119	**	0.0152	**
	(0.0044)		(0.0045)	
More than College	0.0521	**	0.0527	**
	(0.0059)		(0.0061)	
Male	0.0170	**	0.0242	**
	(0.0030)		(0.0031)	
Married	-0.0040		-0.0085	*
	(0.0035)		(0.0035)	
Black	-0.0164	**	-0.0173	**
	(0.0053)		(0.0053)	
Other Non-white	0.0124		0.0046	
	(0.0068)		(0.0068)	
Foreign Born	0.0360	**	0.0420	**
	(0.0046)		(0.0045)	
Non-proxy Interview	-0.0061		-0.0023	
	(0.0035)		(0.0035)	
Work Discontinuity	--		0.0735	**
			(0.0055)	
Probability of Being a Contractor	--		0.0723	**
			(0.0079)	
Any Full Time Jobs	--		-0.0272	**
			(0.0057)	
CPS earnings under $1K			0.1718	**
			(0.0098)	
$1K < CPS earnings < $2.5K			0.0602	**
			(0.0061)	
$12.5K < CPS earnings < $25K			-0.0092	
			(0.0050)	
CPS earning over $25K			0.0311	**
			(0.0112)	
CPS earnings missing			0.1710	**
			(0.0054)	
Observations	63,901		63,901	
R-squared	.010		.045	

* significant at 5% level; ** significant at 1% level

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. Standard errors in parentheses.

Table 5: Discrepancies in Multiple Job Status Between CPS and UI Data

	Number of In-Scope Jobs in UI	
	One	Two Plus
Number of In-Scope Jobs in CPS		
One		
Overall Share	0.813	0.104
	(0.001)	(0.001)
Row Share	0.887	0.113
	(0.001)	(0.001)
Column Share	0.956	0.692
	(0.000)	(0.005)
Two Plus		
Overall Share	0.037	0.046
	(0.000)	(0.000)
Row Share	0.446	0.554
	(0.007)	(0.007)
Column Share	0.044	0.308
	(0.000)	(0.005)

Note: Numbers shown are weighted shares of persons in the CPS-UI overlap sample described in the text who both sources agree have an in-scope job. The upper left quadrant corresponds to the Y1 group, the upper right quadrant to the Y2 group, the lower left to the Y3 group and the lower right to the Y4 group. These estimates use the more restrictive set of criteria described in the Data Appendix to identify persons holding multiple jobs in the CPS. Standard errors are shown in parentheses.

Table 6: Effects of Person and Job Characteristics on the Probability that a UI Multiple Job Holder Has a Single CPS Job (Y2)

Age 16 to 24	-0.0063		0.0023	
	(0.0161)		(0.0161)	
Age 25 to 34	0.0176		0.0313	*
	(0.0133)		(0.0131)	
Age 55 to 64	0.0269		0.0350	
	(0.0223)		(0.0219)	
Age 65 plus	0.0316		0.0621	
	(0.0492)		(0.0483)	
Less than High School	0.0143		0.0125	
	(0.0183)		(0.0179)	
Some College	-0.0482	**	-0.0508	**
	(0.0137)		(0.0134)	
College Graduate	-0.0344	*	-0.0595	**
	(0.0161)		(0.016)	
More than College	-0.0288		-0.0822	**
	(0.0224)		(0.0225)	
Male	0.0171		0.0069	
	(0.0108)		(0.0107)	
Married	-0.0383	**	-0.0278	*
	(0.0122)		(0.0120)	
Black	0.0857	**	0.0972	**
	(0.0160)		(0.0156)	
Other Non-white	0.0044		0.0027	
	(0.0238)		(0.0233)	
Foreign Born	0.0487	**	0.0574	**
	(0.0164)		(0.0161)	
Non-proxy Interview	-0.0186		-0.0255	*
	(0.0125)		(0.0122)	
Any Long 2nd Jobs	--		-0.1103	**
			(0.0144)	
Three or More UI jobs	--		-0.0016	
			(0.0115)	
Qtr UI earn < $1K (2nd job)	--		-0.0136	
			(0.0139)	
$1K <= Qtr UI earn < $2.5K (2nd job)	--		-0.1572	**
			(0.0150)	
$12.5K <= Qtr UI Earn < $25K (2nd job)	--		0.2097	**
			(0.0264)	
Qtr UI earn >= $25K (2nd job)	--		0.2499	**
			(0.052)	
Observations	7,442		7,442	
R-squared	.015		.058	

* significant at 5% level; ** significant at 1% level

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 restricive set of criteria described in the Data Appendix for identifying CPS multiple job holders applied. Standard errors in parentheses.

Table 7: Effects of Person and Job Characteristics on the Probability that a CPS Multiple Job Holder Has a Single UI Job (Y3)

Age 16 to 24	-0.0786	**	-0.1168	**
	(0.0228)		(0.0225)	
Age 25 to 34	-0.0139		-0.0331	
	(0.0187)		(0.0183)	
Age 55 to 64	0.0873	**	0.0868	**
	(0.0297)		(0.0290)	
Age 65 plus	0.0426		0.0124	
	(0.0679)		(0.0664)	
Less than High School	0.0310		0.0143	
	(0.0271)		(0.0265)	
Some College	-0.0099		0.0036	
	(0.0197)		(0.0194)	
College Graduate	0.0368		0.0450	*
	(0.0227)		(0.0223)	
More than College	0.0969	**	0.1167	**
	(0.0294)		(0.0291)	
Male	0.0538	**	0.0468	**
	(0.0153)		(0.0150)	
Married	-0.0252		-0.0178	
	(0.0172)		(0.0169)	
Black	-0.0365		-0.0378	
	(0.0258)		(0.0253)	
Other Non-white	-0.0337		-0.0367	
	(0.0341)		(0.0333)	
Foreign Born	0.0450		0.0310	
	(0.0238)		(0.0234)	
Non-proxy Interview	-0.0571	**	-0.0403	*
	(0.0178)		(0.0175)	
Simultaneous Multiple Jobs	--		-0.1020	**
			(0.0210)	
Multiple Jobs All Three Months	--		-0.0601	**
			(0.0220)	
16+ hours per week 2nd job(s)	--		-0.1236	**
			(0.0220)	
Observations	4,352		4,352	
R-squared	.025		.070	

* significant at 5% level; ** significant at 1% level

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. Standard errors in parentheses.

Table 8: Effects of Person Characteristics on the Probability of UI Employment and CPS Employment

	UI Employment	CPS Employment
Age 16 to 24	-0.042 **	-0.082 **
	(0.005)	(0.004)
Age 25 to 34	0.052 **	0.042 **
	(0.004)	(0.004)
Age 55 to 64	-0.182 **	-0.202 **
	(0.005)	(0.004)
Age 65 plus	-0.514 **	-0.585 **
	(0.004)	(0.004)
Less than High School	-0.143 **	-0.162 **
	(0.004)	(0.004)
Some College	0.024 **	0.033 **
	(0.004)	(0.003)
College Graduate	0.047 **	0.071 **
	(0.004)	(0.004)
More than College	0.038 **	0.088 **
	(0.006)	(0.006)
Male	0.042 **	0.066 **
	(0.003)	(0.003)
Married	0.019 **	0.020 **
	(0.003)	(0.003)
Black	-0.002	-0.032 **
	(0.005)	(0.005)
Other Non-white	-0.058 **	-0.064 **
	(0.006)	(0.006)
Foreign Born	-0.019 **	-0.007
	(0.004)	(0.004)
Non-proxy Interview	-0.027 **	-0.037 **
	(0.003)	(0.003)
Observations	107,731	107,731
R-squared	0.18	0.25

* significant at 5% level; ** significant at 1% level

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. Standard errors in parentheses.

Data Appendix

Analysis Sample

The data file used in our analysis is a linked sample for which CPS records have been matched with state unemployment insurance (UI) wage records 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 agreed to having their survey responses matched to administrative records but did not know or could not 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).

Because we must construct a quarterly employment measure and PIKs are available in the CPS only for March, our attention is restricted to individuals for whom CPS reports were present in all three months of the first quarter of the year. Given the sample rotation pattern for the CPS, this means our sample includes only individuals who were in their 3rd, 4th, 7th, or 8th month in the CPS sample in March of the year in question, 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. Second, the PIK is missing for approximately 20 to 30 percent of March CPS respondents. Finally, although the LEHD program is approaching national coverage, data are available in all years from the mid-1990's through the present only for the following 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 percent or more of the workers resident in the state also are employed in the state, but more than 15 percent 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, leaving us with data for 16 states. These states account for roughly 50 percent of employment nationwide.

In addition to these constraints on the CPS records we can use, we restrict our attention to CPS records for persons aged 16 years or older. This restriction makes the age range of our sample comparable to the age range of the population covered by published CPS employment estimates.

Estimation Weights

One consequence of the restrictions imposed on our linked sample is that we cannot use the standard CPS weights for our analysis. We use a two-step propensity scoring and weight

adjustment procedure to modify the CPS sample weights to make our analysis sample representative of the population in the 16 states for which we have data.

Step 1: The March basic estimation weight is the foundation for our adjusted sample weights. As noted above, 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. March basic estimation weights are used in estimating these models. For each observation in the three-month sample, we then construct a weight adjustment factor equal to the inverse of this predicted probability. Only about half of CPS respondents are in the rotation groups eligible to have been interviewed in all three of the months we care about, and others are lost because they moved or failed to respond in one or more months. The average value of this propensity score adjustment factor is roughly 2.5.

Step 2: Our analysis also requires that a CPS record have available a PIK for linking to the UI data. Roughly 20 to 30 percent 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 available 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 has a much smaller impact on the distribution of weights, adjusting the value of each weight upward by roughly 25 to 40 percent on average (again, the exact adjustment varies with respondent characteristics).

By construction, the estimates of the population for the month of March obtained by applying the final weights just described to our CPS analysis sample match the published CPS population estimates. Estimates based on the first set of propensity-score-adjusted weights we produced yielded employment estimates for the month of March that were noticeably larger than those based on the full CPS sample. This was because these initial models did not account for the fact that employed people are more likely to have a PIK on their CPS records than people who are not employed. Adding a measure of employment to the propensity score models to take this higher probability into account reduced the final weight accorded to employed sample members. Applying these new weights to our analysis sample yielded employment estimates for the month of March that were very close to published CPS employment estimates.

Measuring Employment Status and Job Characteristics

Measuring the number of unique in-scope jobs held by a respondent in each quarter in both the UI and the CPS data is not a completely straightforward task. The strategy we have

adopted is outlined here, together with a brief description of other measures we have created of the characteristics of UI and CPS jobs held during the quarter.

UI Variable Construction. For employers who are covered by a state's unemployment insurance system, the UI data contain a variable identifying the employing business, a variable identifying the worker, and a record of the earnings the worker received at the business in each quarter.[13] Thus, there is a UI record for each employer-employee match present in each quarter. 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. Using this information, we create a first variable to indicate whether a person in the linked sample is observed with positive in-scope UI earnings during the first quarter and a second variable that is equal to the number of different in-scope employers from whom the worker received positive earnings in the quarter.

We also construct variables that capture earnings during the quarter on the worker's primary job, defined simply 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, we code that job as a long-duration job. Using this information, we create variables to indicate whether an individual had at least one long-duration job and whether the individual held any secondary jobs that were of long duration.

CPS Variable Construction. The basic CPS employment data are collected monthly and refer to employment during the survey reference week. Although the CPS records up to four jobs in each month, the class-of-worker information needed to determine whether a job is in-scope is collected as part of the basic CPS interview only for the main job. Class of worker information also is collected for the second job, if there is one, but only in the outgoing rotation months (CPS months-in-sample four and eight). No class of worker information is observed for other jobs.

We define a CPS respondent to have been an in-scope worker in the first quarter if they worked at a main job in any of the three months that was a non-agricultural private sector wage-and-salary job or a wage-and-salary job in state or local government.[14] Constructing a count of the number of unique in-scope jobs held over a quarter requires knowing both the number of in-scope 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 one point in

---

[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. (2006).

[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 months, 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 percent on average.

time. Further, the monthly CPS questionnaire probes only for changes in main job, and even this information is not complete. Given these difficulties, rather than attempting to count the number of in-scope jobs that 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.

A worker may hold more than one job during the quarter either by holding more than one job at the same time and/or by changing jobs. We can identify neither of these situations in the CPS with certainty in all cases. We can, however, separate workers into broad groups within which we have greater or lesser confidence that the worker held more than one in-scope job at the same time or that the worker changed employers. By combining these groups in different ways, we are able to generate both a conservative and a more expansive version of the multiple job indicator variable for each worker, and thence both a conservative and a more expansive estimate of the total number of workers who held multiple in-scope jobs in the quarter.

Simultaneous Jobholders. Some CPS workers can be identified with reasonable confidence as having held two or more in-scope jobs simultaneously at some point during the quarter, but for others this is less clear. The complicating factor is that some of those who report holding multiple jobs may hold a mix of in-scope jobs and jobs in agriculture, jobs in private households, jobs in the federal government and/or self-employment jobs, and complete information about the type of job is not collected as part of the basic monthly interview. We consider two possible cases.

First, if a worker has two or more jobs in January, February and/or March, has two or more jobs in his/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. These people belong to multiple job group 1 (MJ1). In each year, roughly 2 percent of CPS workers fall into this group.

Second, 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 percent and 5 percent of workers fall into this category. Though many in this group may hold more than one in-scope wage-and-salary job, others may not.

Job Changers. Those who changed jobs during the quarter also can be identified with greater or lesser confidence. The basic CPS questionnaire includes a question asked of those employed both in the prior month and in the current month that indicates whether the respondent has changed employer on their main job, but this question is not always asked even when one might think it should be.[15] In addition, respondents are not queried about changes in employer on any jobs other than the main job.

---

[15] Roughly 7 percent of workers who are employed in adjacent months have a missing value for the variable that records the answer to the job change question. When we asked why this variable was missing for so many people, we were told that the question is a screener to determine whether information on industry and occupation needs to be

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. We can be reasonably confident that those who reported a change of employer did indeed change jobs. 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 percent and 3 percent of workers belong to this group.

A high fraction of workers either report no job change or were not asked the job change question. A fair number of these workers, however, have employment patterns that suggest a job change may have occurred. 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 but does report a change in industry, occupation, or class of worker between months, we consider him/her to be a possible job changer and the worker 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, there is a basis for assignment to group JC2. Again depending on the year, between roughly 4 percent and 6 percent of workers belong to this group.

Multiple Job Indicator. Note that it is entirely possible for a CPS worker both to hold simultaneous in-scope jobs and to have changed from one in-scope job to another, though either alone would be enough for the person to be categorized as holding more than one in-scope job over the course of the quarter. We use the indicator variables 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. In addition to variables that tell us whether there is reason to believe that a CPS worker held more than one in-scope job simultaneously at any point during the quarter or changed from one in-scope job to another during the quarter, we construct several other job characteristic measures based on the CPS data. First, we construct indicators of whether the person was without an in-scope job in at least one month of the quarter and, for those holding more than one job during the quarter, whether they ever held multiple jobs simultaneously and whether this was so in all three months of the quarter. Another variable 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 percent for all employment. This variable was constructed using data for 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 was 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

updated and that interviewers have the discretion just to ask those questions if they have any reservations about the quality of the job information collected in the previous interview.

interview, converted to a quarterly equivalent assuming the same earnings in all 13 weeks of the quarter, 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 XX percent 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 Appendix Figure 1 show how accounting in turn for each of the differences between the published CPS and CES estimates affect the behavior of the resulting employment series 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 A1a, 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.

The next panel, Figure A1b, 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 A1c 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, though 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 five 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, then rises between 2002 and 2003.

In Figure A1d, 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 A1d look very similar to those in the top panel.

Finally, in Figure A1e, we switch from estimates based on the full UI database to estimates based on our linked sample of approximately 12 to 15 thousand 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.

Appendix Figure 1: Effects of Sample Restrictions and Employment Concepts on CPS, CES, and UI Employment Trends, 1996-2003

Figure A1a-1: CPS - CES Employment Trends
National, All Sectors, March Employment, SA

Figure A1a-2: CPS - CES Employment Trends
National, All Sectors, March Employment, SA

Figure A1b-1: CPS - CES Employment Trends
16 States, All Sectors, March Employment, NSA

Figure A1b-2: CPS - CES Employment Trends
16 States, All Sectors, March Employment, NSA

Figure A1c-1: CPS - CES Employment Trends
16 States, PSL, March Employment, NSA

Figure A1c-2: CPS - CES Employment Trends
16 States, PSL, March Employment, NSA

Figure A1d-1: CPS - UI Employment Trends
16 States, PSL, 1st Quarter Employment, NSA

Figure A1d-2: CPS - UI Employment Trends
16 States, PSL, 1st Quarter Employment, NSA

Figure A1e-1: CPS - UI Employment Trends
16 States, PSL, 1st Quarter Employment, NSA

Figure A1e-2: CPS - UI Employment Trends
16 States, PSL, 1st Quarter Employment, NSA