THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Bureau Survey and SSA Administrative Data

No. 252

John M. Abowd and Martha H. Stinson
U.S. Census Bureau

May 2011

U.S. Department of Commerce U.S. CENSUS BUREAU

Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Bureau Survey and SSA Administrative Data*

John M. Abowd and Martha H. Stinson†

May 2011

SEHSD Working Paper 2011-19

Abstract

We quantify sources of variation in annual job earnings data collected by the Survey of Income and Program Participation (SIPP) to determine how much of the variation is the result of measurement error. Jobs reported in the SIPP are linked to jobs reported in an administrative database, the Detailed Earnings Records (DER) drawn from the Social Security Administration's Master Earnings File, a universe file of all earnings reported on W-2 tax forms. As a result of the match, each job potentially has two earnings observations per year: survey and administrative. Unlike previous validation studies, both of these earnings measures are viewed as noisy measures of some underlying true amount of annual earnings. While the existence of survey error resulting from respondent mistakes or misinterpretation is widely accepted, the idea that administrative data are also error-prone is new. Possible sources of employer reporting error, employee under-reporting of compensation such as tips, and general differences between how earnings may be reported on tax forms and in surveys, necessitates the discarding of the assumption that administrative data are a "true" measure of the quantity that the survey was designed to collect.

*We would like to thank Joanne Pascale, Tracy Mattingly, Gary Benedetto, Julia Lane, George Jakubson, David Johnson, Donald Kenkel, Simon Woodcock, Kevin McKinney, Kristin Sandusky, Lars Vilhuber, and Marc Roemer for helpful comments and support. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau, Cornell University, or any of the project sponsors. This work was partially supported by the National Science Foundation Grant SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging, and the Alfred P. Sloan Foundation. Abowd also acknowledges direct support from the Census Bureau and other NSF grants. All of the data used in this paper are confidential data. The U.S. Census Bureau supports external researchers' use of some of these data items through the Research Data Center network (www.census.gov/ces). For public-use data please visit www.sipp.census.gov/sipp/ and click "Access SIPP Synthetic Data."

†John M. Abowd is the Edmund Ezra Day Professor of Industrial and Labor Relations at Cornell University, Distinguished Senior Research Fellow at the U.S. Census Bureau, Research Associate NBER, Research Affiliate CREST/INSEE, and Research Fellow IZA. Martha Stinson is an economist at the U.S. Census Bureau. Contacts: john.abowd@cornell.edu, martha.stinson@census.gov

In addition, errors in matching SIPP and DER jobs, a necessary task in any use of administrative data, also contribute to measurement error in both earnings variables. We begin by comparing SIPP and DER earnings for different demographic and education groups of SIPP respondents. We also calculate different measures of changes in earnings for individuals switching jobs. We estimate a standard earnings equation model using SIPP and DER earnings and compare the resulting coefficients. Finally exploiting the presence of individuals with multiple jobs and shared employers over time, we estimate an econometric model that includes random person and firm effects, a common error component shared by SIPP and DER earnings, and two independent error components that represent the variation unique to each earnings measure. We compare the variance components from this model and consider how the DER and SIPP differ across unobservable components.

1 Introduction

This paper linked survey and administrative data to compare two different measures of earnings in order to study the causes of the differences between them. We link job-level earnings reports from the Survey of Income and Program Participation (SIPP) to W-2 records from the Social Security Administration's Detailed Earnings Record (DER) extracted from its Master Earnings File. These matched records provide a unique opportunity to assess differences in employee and employer reports and to consider the impact of these differences on correlations of earnings and other variables of interest in the SIPP survey.

The majority of past studies that compare two data sources designate one of the sources as the "truth," usually the administrative record value. In contrast, we begin with an agnostic view about which data source is true, believing instead that there are legitimate differences between survey and administrative reports. Our goal is to first state the differences between the two data sources and then consider the reasons for these differences.

We believe there are at least three reasons why administrative and SIPP survey reports on earnings might differ. First, there may be matching errors between the records from the two sources. Our matching methodology uses a probabilistic record linkage that produces some false matches. While identifiers for individuals remain constant over time in administrative data, identifiers for firms do not, and this complicates the matching process. Second, the administrative records we use do not include some categories of earnings, such as health insurance premiums, that are usually reported on a pay stub and would probably be reported by a survey respondent. The reverse is also true: there are categories of earnings that would appear on a W-2 form and would not appear on a pay stub, increasing the likelihood that these earnings would not be reported in a survey. Finally, there are differences in reporting that can be labeled mistakes or measurement error. We believe these types of differences arise in both data sources. Employers make mistakes on W-2 forms just as employees make mistakes when they report earnings to a survey collector. The Social Security Administration does make corrections to the Master Earnings File when revised W-2 forms are submitted and when a potential claimant presents credible evidence of errors in the earnings history. Over time mistakes in the Master Earnings File become less prevalent. One might hypothesize that errors in an administrative database are in general less prevalent than in a survey but they still exist.

Our strategy for investigating the differences between the SIPP and DER is to first focus on average differences for demographic sub-groups of interest. If differences between the two data sources vary across sub-groups, this may be evidence of mistakes made more frequently by some types of survey respondents. On the other hand it could also be that some sub-groups have more complicated types of earnings, which give rise to more definitional types of discrepancies between the two data sources. After these initial comparisons, we next focus on differences between the sources that are due to unobservable factors. Within each cell defined by a set of stratifying variables, there are average SIPP and DER earnings and then there is variation due to unobservable person, firm, and time period characteristics. We consider how much of this variation is common between the SIPP and the DER and how much might be unique to one data source or the other.

Our paper is organized as follows. We begin with a brief background on measurement error studies. We then describe the two data sources and our process for matching them. We follow with results from comparing the two data sources along observable characteristics and discussion of potential causes of the differences that we observe. We then explain our model for estimating the variance due to unobservable factors. We report variance components from the SIPP and the DER and then calculate reliability ratios under different assumptions about which components of variation represent true variation. We conclude with some discussion about how the differences we observe might affect analyses done with SIPP data.

2 Background

Economists and statisticians have long recognized that survey data are prone to measurement error. Responses to questions about earnings, education levels, and job characteristics are not measured exactly but instead contain some truth and some error. The classical measurement error model as described by Fuller (2006, chapter 1) defines a dependent variable Y_t that is a linear function of covariates x_t:

$$Y_t = \beta_0 + \beta_1 x_t + e_t$$

However, x_t is not observed directly, and instead we observe

$$X_t = x_t + u_t$$

where x_t is the true value of the covariates and u_t is the measurement error. By assuming that the measurement error, the true values, and the errors are independently distributed as

$$\begin{bmatrix} x_t \\ e_t \\ u_t \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mu_x \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{xx} & 0 & 0 \\ 0 & \sigma_{ee} & 0 \\ 0 & 0 & \sigma_{uu} \end{bmatrix} \right\}$$

the joint distribution of Y and X can be written as

$$\begin{aligned} E[Y, X] &= (\beta_0 + \beta_1 \mu_x, \mu_x) \\ Var[Y, X] &= \begin{bmatrix} \beta_1^2 \sigma_{xx} + \sigma_{ee} & \beta_1 \sigma_{xx} \\ \beta_1 \sigma_{xx} & \sigma_{xx} + \sigma_{uu} \end{bmatrix}. \end{aligned}$$

When Y is regressed on X, the expected value of the estimated regression coefficient $\hat{\beta}_1$ is attenuated.

$$E\left[\hat{\beta}_1\right] = \beta_1 \frac{\sigma_{xx}}{(\sigma_{xx} + \sigma_{uu})}$$

The ratio

$$\kappa_{xx} = \frac{Cov[u, X]}{Var[X]} = \frac{\sigma_{xx}}{(\sigma_{xx} + \sigma_{uu})},$$

often called the reliability ratio, defines the ratio of $\hat{\beta}_1$ to β_1. The proportional attenuation bias resulting from measurement error is defined as $\frac{\beta_1 - \hat{\beta}_1}{\beta_1} = 1 - \kappa_{xx}$.

The bias resulting from measurement error can be exacerbated if one is using first-differenced data. As Angrist and Krueger (1999) describe, the reliability ratio for a variable $\triangle X = (X_t - X_{t-1}) = (x_t - x_{t-1}) + (u_t - u_{t-1})$ is equal to

$$\kappa_{\triangle x \triangle x} = \frac{\sigma_{xx}}{\sigma_{xx} + \sigma_{uu}\left(\frac{(1-\tau)}{(1-\rho)}\right)}$$

where τ is the auto-correlation coefficient of the measurement error and ρ is the auto-correlation coefficient of the true value of earnings. If $\rho > \tau$ then $\frac{(1-\tau)}{(1-\rho)}$ is greater than one and the signal to noise ratio declines. Thus determining the extent to which measurement error persists over time is important in assessing the impact on the estimated coefficient.

If the variance and structure of the measurement error is known, then unbiased or consistent estimators of β_1 can be obtained. Hence, those studying measurement error have focused on estimating κ_{xx} and testing whether the assumptions of classical measurement error were violated. Studies that obtain a second report for the miss-measured variable of interest in order to calculate σ_{uu} and σ_{xx} have been termed validation studies. The most common approach is to treat the second report as "truth" and calculate the measurement errors directly as $u = X - x$. The properties of these errors can then be investigated and researchers have often concluded that the assumptions of classical measurement errors were violated and that the errors were correlated with the true values, i.e. $\sigma_{xu} \neq 0$. However, they acknowledge that their models are driven by the assumption that they obtained a true measure of x. Without this assumption, there would be no way to determine the relation between the errors and the true values. This assumption is fundamentally untestable and is justified solely by the authors' knowledge of the quality of the secondary data source.

One of the first earnings validation studies was done by Mellow and Sider (1983) using a special supplement to the January 1977 Current Population Survey (CPS) that obtained name and address information of employers from the survey respondents.[1] Matched pairs with both employer and employee wage reports totaled 3,612. In these data, employer-reported wages exceeded worker reports by 4.8% on average. In order to test the sensitivity of statistical models to the source of the variables used, the authors estimated four different wage regressions. In the first two wage equations, the logarithms of worker and employer reported wages, respectively, were regressed on respondent-reported variables for union status, industry and occupation. In the second two wage equations, employer-reported union status, industry and occupation replaced the survey self-reports in the conditioning variable list. Returns to education and experience were strikingly constant across these four equations. The nonwhite-white wage differential was smaller when using employer-reported wages while the female differential was higher. The union wage-premium was smaller when using employer-reports of union coverage. Occupation and industry differentials were very similar across the different specifications. The authors concluded that the wage regressions were generally not that sensitive to the source of information: worker versus employer.

During the 1980s, a validation study at a large anonymous manufacturing company was undertaken. Results from this study were reported in Duncan and Hill (1985) and Bound *et al.* (1994). Workers at the company were interviewed using a Panel Study of Income Dynamics (PSID) survey instrument and then information for these workers was obtained from company records. Bound *et al.* provided a comprehensive report on both waves. The first wave of data was collected in the summer of 1983 and included 418 workers and the second wave was conducted in 1987 with 341 of the originally interviewed workers and an additional

[1] Mellow and Sider also evaluate a second matched data set: the Employment Opportunity Pilot Project (EOPP). However this data set contains only general firm data such as industry and union status matched to specific workers and hence it is not possible to compare earnings reports from both the employer and employee using this data set.

151 new workers. The authors treated the company reports of annual earnings as measures of true earnings values and considered any differences between worker and employer reports to be errors on the part of the workers. According to the authors, “We do this because of our confidence in the accuracy and recording of the company records, in part because of the extraordinary cooperation of the company involved. This is crucial, because if there were significant errors in the company records, one would have no way of knowing how they were correlated with other variables.” (page 351) By their own acknowledgement, the results in this paper were completely driven by this assumption.

The authors reported a noise to total variance ratio $\left(\frac{\sigma_{uu}}{(\sigma_{xx}+\sigma_{uu})}\text{in the notation above}\right)$ of 0.302 for annual earnings in 1986 and 0.151 for annual earnings in 1982. They argued that this ratio was misleading because the errors in earnings were correlated with the true levels of earnings. In this case the true variance ratio should be

$$\frac{Cov\left[X,u\right]}{Var\left[X\right]}=\frac{\sigma_{uu}+Cov\left[u,x\right]}{\sigma_{xx}+\sigma_{uu}+2Cov\left[u,x\right]}$$

This ratio was calculated by regressing the measured errors on the employee-reported annual earnings, and was 0.239 in 1986 and 0.076 in 1982. Thus the authors claimed that when earnings measures are used as independent variables in regression analyses, the bias resulting from measurement error will be mitigated by correlation between errors and true values.

Generally, measurement error in a dependent variable will not cause bias in the estimated regression coefficients but will make them less precise because of the increased overall variance of the measured Y. However, the correlation between the true value and the error of a dependent variable will introduce bias even if the independent variables are measured without error. The authors described this result using the following setup for the estimation:

$$\begin{aligned}
Y &= (1+\delta)y+v=x\beta+\varepsilon \\
\widehat{b} &= \frac{(1+\delta)Cov\left[y,x\right]}{Var\left[x\right]} \\
\frac{\widehat{b}}{\beta} &= (1+\delta)
\end{aligned}$$

where y is the true value of Y, $Cov\left[y,v\right]=0$ and $Cov\left[\varepsilon,v\right]=0$. Thus, the proportional attenuation bias in the coefficient is δ, which was estimated as -0.172 for 1986 and -0.104 for 1982. Again, the calculation of these results was completely dependent on the strategy used to identify the errors separately from the true value of earnings.

The authors concluded by estimating two earnings equations, one using employee reported measures of earnings and tenure and the other using company recorded measures of the same variables. Education and experience were also included in the regressions. Since only one measure of education and experience was available (employee interview responses), these variables were considered measured without error. Regression coefficients from the worker-reported equation were measured against the "true" coefficients from the company-reported equation. According to this standard, the interview data overstated the return to education by 40% and understated the return to tenure by 20%.

Bound and Krueger (1991) conducted a similar validation study using linked CPS-Social Security Earnings Records March 1978 CPS respondents were asked to report their Social Security Numbers and, using SSN, name, age, sex, and race, respondents were linked to SSA records. About 50% of respondents who were in both the 1977 and 1978 March CPS were successfully linked to SSA data. This study was complicated by the fact that reported SSA earnings were truncated at the maximum Social Security taxable earnings amount ($16,500 in 1977 and $15,300 in 1976). The authors made the same error-identifying assumptions as Bound, *et al.* Administrative records were viewed as truth with the exception that the truth was sometimes truncated. The authors first estimated the relation between the SSA and CPS earnings using a Tobit maximum likelihood approach, which accounted for the truncation. The results from this estimation were used to calculate the covariance matrix between CPS earnings and true SSA earnings. This matrix in turn was used to compute a covariance matrix between x_t and u_t. The authors reported large negative correlations between measurement error and true earnings for both 1976 and 1977 (-0.46 and -0.42, respectively). They reported reliability ratios that did and did not take account of these correlations as 0.844 and 1.016, respectively, for 1976 and 0.819 and 0.974 for 1977. They also noted that the reporting errors appeared to be positively correlated over time but "with only 2 years of data it is impossible to distinguish an autoregressive process in the measurement error from a person fixed effect or from other time-series processes." (page 15)

Bound, Brown, and Mathiowetz (2001) summarized earnings validation studies and stated that the ideal information for correcting measurement error would be to know the joint distribution of all the true and observed variables, *i.e.*, $f(y, x, Y, X)$. However, the authors recognized that information about this joint distribution has often come at the cost of assuming that validation data are the truth. They write, "Those collecting validation data usually begin with the intention of obtaining 'true' values against which the errors of survey reports can be assessed; more often than not we end up with the realization that the validation data are also imperfect. While much can still be learned from such data, particularly if one is confident the errors in the validation data are uncorrelated with those in the survey reports, this means replacing one assumption (*e.g.* errors are uncorrelated with true values) with another (*e.g.,* errors in survey reports uncorrelated with errors in validation data)." (page 3832)

Bound, Brown, and Mathiowetz also expressed the hope that future validation studies would be able to obtain secondary data reports for multiple consecutive years. Past validation studies have been able to create panels of earnings measures for at most two consecutive years. Thus, it has been difficult to calculate the correlation of errors over time, an important component in assessing the impact of measurement error on longitudinal data. Due to the high cost of validating panel data, the authors foresee the future of validation studies as being critically enhanced by opportunities to "merge administrative data to existing panel data." (page 3832)

A more recent work by Gottschalk and Huynh (2010) uses matched SIPP and SSA data to consider the effect of measurement error on earnings inequality and mobility. The authors conclude that measurement error is mean-reverting, *i.e.*, non-classical, but show that, in their framework, this type of error partially offsets the bias in estimates of inequality in the SIPP. They also conclude that measurement error is correlated over time and this diminishes attenuation bias in the correlation of earnings and hence lessens the impact of measurement error on estimates of earnings mobility in the SIPP.

Our work is unique in this literature because of our view that "true" earnings are unknown and, instead, we observe two separate earnings reports that are correlated with each other and with the truth. The other distinguishing feature of our study is that we compare earnings at the job-level, not the person-level, in a data set we created by matching the jobs using probabilistic record linkage techniques. Analyzing data differences at the job level has the advantage of providing more direct insight into the source of these differences. With the increasing use of matched employer-employee administrative data, understanding both survey and tax job-level data and their potential errors is important to researchers who study many aspects of the labor market.

3 Data Description

The fundamental unit of observation in this paper is a job, defined as a match between an individual and an employer. Data on jobs come from two sources: five Survey of Income and Program Participation (SIPP) Panels conducted during the 1990's[2] and the Detailed Earnings Records (DER) extracted from the Social Security Administration Master Earnings File for the respondents in each of the five panels. In the SIPP, data on earnings were reported on a monthly basis while in the DER, earnings were reported on an annual basis. In both sources there were multiple records per job from repeated interviews and annually filed W-2s. Hence, in order to compare earnings, we first had to identify jobs and group earnings records over time, in each data source. After job records were created, individuals in each data set were linked by Social Security Number and then, for each individual, job records from the SIPP and the DER were matched to each other. We describe each step of this process below.

[2] The five SIPP panels began in 1990, 1991, 1992, 1993, and 1996.

3.1 Creating a SIPP Jobs Data Set

All the SIPP Panels conducted in the 1990s collected detailed labor force information from respondents every 4 months, or approximately 3 times per year, over the course of $2\frac{1}{2}$ to 4 years. Respondents were asked questions about at most 2 jobs held during the previous 4 months, where the term job was loosely defined as working for pay. The employer name, industry, occupation, union status, usual weekly hours, and monthly earnings of each job were recorded, as well as any applicable start and end dates. Combining records from interviews that contained information on the same job allowed us to sum monthly earnings to create annual earnings. To facilitate such linking, during the survey, each job was also assigned a unique identification number, or job ID, with the intent that this identifier be time-invariant. For the first four panels (1990-1993), the Census field representative (FR) used a paper survey form, recorded the employer name, and assigned a job ID for each reported job, even if the job was a continuation of one reported in a previous wave. While the FR was supposed to assign the same job ID to a continuing job and a new job ID to a newly begun job, there was no quality check to ensure that this procedure was followed. Beginning in 1996, a major survey redesign was implemented and information was collected using a Computer Assisted Personal Interview (CAPI) system. As a result, as long as the individual did not miss an interview, during the second and subsequent interviews, the CAPI instrument automatically assigned the same employer name and job ID each time further information about a continuing job was collected. When the respondent reported that a new job had started, the CAPI instrument assigned the next available job ID.

We used the longitudinal SIPP person ID, the wave (interview) number, and the job ID to combine records and create one observation per person per job that contained both time invariant information, such as industry, and time-varying information such as annual earnings. Table 1 shows the total number of respondents in each SIPP panel, the number that report holding at least one job over the course of the SIPP panel, the total number of person-wave-job records, and the total number of jobs reported, using the three identifiers listed above to count jobs. A careful examination of the person-wave-job records revealed serious problems with the SIPP job ID coding process. Because the definition of a "job" was so crucial to our comparison of job earnings from the SIPP and the DER data, we investigated the nature and causes of the job ID coding problems and developed an editing procedure that would resolve some of the inconsistencies we found. Appendix A describes the problems we found and gives a summary of how we repaired the job id variables. The last line of Table 1 shows the number of SIPP jobs after correcting the SIPP job id variable.[3]

Once we had defined a set of jobs for each SIPP panel, we created annual earnings measures by summing

[3] The edited SIPP job ID for the 1990-1993 panels was released by the Census Bureau as an update to the public-use files. The edited job ID is described in Stinson (2003).

monthly earnings reports.[4] It is important to understand the concept of earnings as used during the SIPP interview. During the 1990-1993 SIPP panels, respondents were asked about earnings from a specific employer in the following way: "The next question is about the pay ... received from this job during the 4-month period. We need the most accurate figures you can provide. Please remember that certain months contain 5 paydays for workers paid weekly and 3 paydays for workers paid every 2 weeks. Be sure to include any tips, bonuses, overtime pay, or commissions. What was the total amount of pay that ... received BEFORE deductions on this job in ...?"[5] The field representative reads the name of each month and separately records earnings for that month. A special caveat is added for members of the Armed Forces, "Be sure to include cash housing allowances and any other special types of pay." The intent of the survey question was to collect gross earnings and if the person responded that he or she did not know the earnings amounts, the field representative asked if the person could provide the information during a follow-up phone call.

The 1996 survey instrument asked, "Each time he/she was paid by [Name of Employer] in [MonthX], how much did he/she receive BEFORE deductions?"[6] The field representative then followed up with questions about whether there were any other payments such as tips, bonuses, overtime pay, or commissions. The FR was trained to probe several times to make sure all the payments from an employer in a given month were accurately reported. There were also consistency checks built into the CAPI instrument that were meant to spot earnings amounts that seemed unreasonable and provide the FR with the opportunity to make corrections. Respondents were also asked to refer to earnings records if possible so as to give accurate responses. Thus, in the best case, these earnings reports most likely reflected the gross pay from monthly pay stubs.

3.2 Creating a DER Job-level Data Set

The second source of data, DER from SSA, contained earnings histories for each SIPP respondent in the 1990, 1991, 1992, 1993, and 1996 panels with a validated SSN (for a definition and discussion of validation see section 3.4). These earnings data had as their source the W-2 records filed by employers on behalf of each employee. The Social Security Administration stored annual W-2 reports in their Master Earnings File (MEF) and created the DER extract for research use. A W-2 history for a SIPP respondent consisted of annual earnings, broken down by employer, from 1978-2000. For the purposes of this earnings comparison

[4] When individuals were missing job-level monthly earnings due to item non-response, the Census Bureau processing staff imputed earnings. We used all such imputes in our calculations of annual earnings. When individuals were missing job-level monthly earnings because they missed an interview completely, the Census Bureau did not impute earnings. We treated these missing monthly earnings as zero because we wanted to use only data available to public data users in our calculation of annual earnings. We view these missing data as representing a type of measurement error that will be captured in our models.

[5] SIPP 1993 Wave 1 questionnaire, page 15, available at http://www.census.gov/ sipp/ core_content/ 1993/ quests/ sipp93w1.pdf cited on February 21, 2011.

[6] SIPP 1996 wave 1 questionnaire, Labor Force Amount section, available at http://www.census.gov/ sipp/ core_content/ 1996/ quests/ screens/lf_par2.html, cited on February 21, 2011.

study, non-self-employment jobs held during the time period covered by the survey questions were used.[7] Employers were identified on the W-2s and in the DER by an IRS-assigned Employer Identification Number (EIN). In Table 2 we detail the time period covered by the survey and the total number of jobs and unique EINs for SIPP respondents during this time period. The primary earnings variable came from Box 1 of the W-2 Form: wages, tips, and other compensation. This earnings variable is uncapped and represents all earnings that were taxable under federal income tax.

The EIN linked employers to the Business Register, the master list of all businesses maintained by the Census Bureau that serves as the sampling frame for firm-level surveys. Using this link, we merged information from the Business Register about the industry and name of the employer to each relevant job report in the DER data. Details about this merge can be found in Appendix B. The employer name is the key linking element between the SIPP and DER job data.

3.3 Conceptual Differences between SIPP and DER: Jobs and Earnings Definitions

Before comparing earnings, we discuss some conceptual differences between SIPP and DER reports. These differences fall into two categories: definition of earnings and definition of job. We briefly discuss these differences and summarize how they might affect the comparison between SIPP and DER earnings.

There are at least two parts of earnings that would be reported on an employee's pay stub in "gross earnings" that are not included in Box 1 of the W-2 Form: pre-tax health insurance plan premiums paid by the employee and pre-tax elective contributions made to deferred compensation arrangements such as 401(k) retirement plans. In the later case, these contributions are reported elsewhere on the W-2 form (for example Box 13 in 1999) and the DER file contains reports of these deferred earnings which can be added to Box 1 earnings to approximate gross earnings. While pre-tax health insurance plan premiums are reported on the W-2 Form, they are not contained in the DER extract created for research use. This omission represents one important way in which administrative records may differ from survey records that is not the result of error in the survey data collection process. DER earnings will be lower than SIPP earnings if the respondent reported gross earnings during the survey that included health insurance plan premiums.

There are other possible differences between Box 1 on the W-2 Form and gross earnings reported in the survey, most of which involve some kind of employee benefit that the employee is unlikely to consider wages and may also be unlikely to be reported as such on a pay stub, but which the employer is nonetheless

[7] The DER did contain reports of self-employment earnings. The SIPP also collected information about self-employment, but responses to these questions were treated separately from responses to the questions about jobs with employers. Self-employment reports from either source were not included in this study.

required to report as taxable income. These include educational assistance above a certain monetary level, business expense reimbursement above the amount treated as substantiated by the IRS, payments made by the employer to cover the employee's share of Social Security and Medicare taxes, certain types of fringe benefits such as the use of a company car, golden parachute payments, group-term life insurance over $50,000 paid for by the employer, potentially some portion of employer contributions to Medical Savings Accounts, non-qualified moving expenses, and, in some circumstances, sick pay from an insurance company or other third party payer. In all these cases, DER earnings are likely to be higher than SIPP earnings, due to respondents not reporting these benefits as gross earnings.

A final potential problem with DER employer reports is that EINs do not necessarily remain constant over time. Unlike Social Security Numbers which serve as good longitudinal identifiers for individuals, EINs can change for many reasons that do not necessarily involve a person moving to a new employer. Company reorganizations that consist of mergers, acquisitions, or spin-offs of some parts of the company may result in a worker having two W-2 forms for a tax year, each with a different EIN, without having actually changed jobs. In cases such as these, the DER earnings will be lower than the SIPP earnings because a portion of the earnings for the year are missing. As part of the linking process between DER and SIPP earnings, we attempt to identify these kinds of successor-predecessor problems and merge the two DER jobs determined to be related to a single SIPP job (see Appendix C for details). However, at this early stage of research involving the administrative data, there is no way to know how many cases of this type we miss.

The following list summarizes the potential definitional differences between SIPP and DER earnings.

Health insurance premiums not included in the DER:	DER $<$ SIPP
Employee benefits included in the DER:	DER $>$ SIPP
EIN changes due to change in firm organization or ownership:	DER $<$ SIPP

3.4 Matching SIPP and DER Jobs

After the creation of the SIPP and DER job-level data sets, the next step was to take people who had job reports in both files and try to match each SIPP job record to a DER job record. In Table 3, we list the total number of people and jobs that were potential matches following the job record creation process on both the survey and administrative side.[8] We began the job matching process by first using validated

[8] Except for the 1996 SIPP jobs, the total number of jobs that were potential matches is the same as row 6 in Table 1 for the SIPP jobs and row 2 of Table 2 for DER jobs. In 1996, one final problem necessitated the dropping of a few additional jobs. Respondents were only allowed to report at most two jobs per interview. In cases where people had a series of short or part-time jobs, interviewers recorded a single job which was labeled as "various employers" or "work arrangement." There were 3,908 job records of this type in the 1996 SIPP data, representing possibly triple that many actual jobs. These jobs were essentially impossible to match to the DER because they do not represent earnings from a single employer. Hence, they were

SSNs to link at the person level. Each SIPP respondent was asked to provide an SSN. SSA then validated the self-reported SSNs against SSNs in the Numident, an administrative data base containing demographic information collected when every SSN was issued. Self-reported name, sex, race, and date of birth from the SIPP were compared to their administrative counterparts on the Numident. If a respondent's name and demographics were deemed close enough to the name and demographics associated with the SSN in the Numident, then the SSN was declared valid.[9] Validated SSNs served as the basis for extracting Detailed Earnings Records from the SSA Master Earnings File. Hence, in order for an individual to have any earnings reports in the DER, he or she, by necessity, must have a validated SSN.

The third column of Table 3 shows the number of people who matched between the SIPP and the DER. In all panels, some people were lost from both the SIPP and DER job data sets as a result of this match. On the SIPP side, there were two reasons why a person might not match. First, he or she might not have a validated SSN. The third column of Table 4A shows the number of people affected by this problem. The second possibility was that the person had a validated SSN and reported working in the SIPP, but did not have any earnings reports in the DER. This problem would be caused by the jobs being relatively informal (baby-sitting, yard work, household help) and not generating W-2 forms, or by over-reporting of jobs. On the DER side, the only reason for a person not to match was because the individual did not report any jobs in the SIPP survey. As seen in the third column of Table 4B, it was far more likely for a person to have jobs in the DER and not the SIPP than the reverse. It would appear that overall, the SIPP undercounts employed people.

As shown in Table 3, even for those people who had employment reports in both the SIPP and the DER, the number of jobs reported was much higher in the administrative data compared to the survey data. At least one factor that influenced the job count on each side was the timing of the survey. In every SIPP panel, the survey asked employment questions of at least some respondents in the last few months of the year preceding the official beginning year. For example in the 1990 panel, the first interview reference period included between one and three months of 1989 for 75% of the sample. The last interviews in the 1990 panel were conducted in September 1992, leaving the last quarter of 1992 uncovered for all respondents. The 1991-1993 panels followed similar patterns. In the 1996 panel, the first interview reference period included December 1995 for a quarter of the sample and the last reference period included one or two months of 2000 for half the sample. In order to attempt to match as many SIPP and DER jobs as possible, all DER jobs

dropped, giving a new total of 121,450 jobs.

[9] For respondents who answered "do not know" to the SSN question, an attempt was made to find the missing SSN by locating the person in the Numident based on their reported name and demographic characteristics. When a respondent refused to provide an SSN, no attempt was made to link this person to any administrative data and the SSN was left missing. The method of SSN validation changed substantially after the 1996 SIPP panel. Beginning with the 2001 SIPP panel, the Census PVS system was used to validate and search for SSNs. Although similar in spirit to the earlier SSA system, in PVS a larger administrative data base is used as well as more formal probablistic matching techniques.

from the years either partially or fully covered by the survey were included in the potential match set, as appropriate for each respondent. However, some of these DER jobs could clearly have ended before the survey began or started after the survey ended, thus artificially inflating the DER job counts. In the early SIPP panels (1990-1993), 27-32% of DER jobs ended before the first full survey year or began in the truncated survey end-year. In the 1996 panel, these jobs only accounted for 11% of all DER jobs, largely because the timing of the survey conformed more closely to the calendar year. Another factor which potentially artificially depressed the SIPP job counts is the fact that the survey only collected information about a maximum of two jobs per wave. However, this procedure still allows for six jobs per year and we think it is unlikely that large numbers of respondents had more than six jobs in a year.

After we matched by SSN, a job-to-job match was performed, using probabilistic record linking based on name matching. The matching was performed in several steps, called passes. The primary basis for matching was self-reported name of the employer from the SIPP and administrative name of the employer from the Business Register. Earnings were not used in the match in order to prevent bias in the subsequent comparison of earnings. Appendix C gives the details of this match including which additional matching variables were used and how duplicate matches were handled. The first row of Table 5 gives the number of SIPP jobs that were successfully matched to a counterpart job in the DER. Of the jobs that matched, we then restricted ourselves to comparing earnings only in years fully covered by the survey.

There were some jobs that matched but did not have the same number of years of reported earnings. For example a SIPP job could have earnings reports for 1996 and 1997 but not 1998 while the DER job could have reports for all three years. We did not require the timing of the earnings in the SIPP and the DER to be identical. We compared SIPP and DER earnings when there was at least one DER earnings report and one SIPP earnings report for a full survey year but we did not require these reports to be in the same year. As a result, the SIPP and DER sample sizes were slightly different for each year. Missing values were modeled in the maximization routine as conditionally missing at random (ignorable, Rubin 1976) and hence the panel was not required to be balanced. The decision not to require exact matching in the earnings years was based on the fact that earnings essentially reported as zero in one source and positive in another source was a type of measurement error that we did not wish to exclude. The third row of Table 5 shows the final total number of jobs per panel that were used in the analysis. At this point jobs from all panels were combined to give a total of 197,337 jobs, 133,849 people, and 110,454 unique employers.

Tables 6 and 7 describe the covariance structure of the SIPP and DER earnings over time. Variances are shown on the diagonals, the covariances are listed below the diagonal, and correlations are listed above. In the SIPP data, the correlations between adjacent years range from 0.54 to 0.74. In the DER, they are higher, ranging from 0.79 to 0.81. The variance of earnings is also higher in the DER than in the SIPP.

Table 8 gives the correlations between each year of DER and SIPP data. The correlations between SIPP and DER earnings in the same year range from 0.74 to 0.86. The correlations between adjacent years of SIPP and DER data are not as high as between adjacent years of DER data but the correlations are quite similar to adjacent years of SIPP data. They range from 0.58 to 0.71. In general, correlations in the early 1990s are lower than correlations in the later years of the decade, as might be expected given the improvements of the 1996 panel.

4 Results from Comparing Data Sources: Observables

Once the matching process was completed, we were able to compare SIPP and DER earnings at the job level. We began our comparisons with simple tables of means, stratified by SIPP demographic and economic variables. These tables show average differences between SIPP and DER earnings and allow us to consider which groups have the most pronounced differences. We look at all individual-job matches and then look separately at individuals who changed jobs. We report coefficients from a regression of the logarithm of DER annualized wages on race, gender, education, labor force experience, and a linear time trend. We compare the DER regression to an identical equation using a SIPP-based annualized wage as the dependent variable.

4.1 Race-Gender-Education Subgroups

In Table 9, we present average earnings for white males, white females, non-white males, and non-white females, stratified by a five category education variable. The education categories are no high school diploma, high school diploma only, some college, college degree, and graduate degree. For every sub-group in this table, average earnings are higher in the DER (column 2) than in the SIPP (column 1). For most groups, the differences become more pronounced as the level of education increases. For example, white males with no high school diploma report earning approximately $12,000 on average in the SIPP, while in the DER, this same group earns approximately $14,000 on average. This $2,000 difference is about 16% of SIPP earnings (columns 3 and 4). In contrast, white males with a graduate degree earn about $44,000 on average according to the SIPP while the DER average is almost $59,000. This difference is 33% of SIPP earnings. Thus it appears that the largest discrepancies between the two data sources occur for more, not less, educated individuals. This pattern is also present for white females and non-white females, although to a lesser extent. For white females, the differences range from 10.6% to 12.7% of SIPP earnings and for non-white females the differences are between 14.5% and 18.1% of SIPP earnings. For non-white males, the education groups on either end of scale have the largest differences between SIPP and DER earnings, on average. Individuals without a high school diploma have earnings 21.1% lower in the SIPP while those with a graduate degree have earnings 20.8% lower in the SIPP.

In Figure 1 we chart the differences in SIPP and DER earnings by gender, race, and education sub-groups. When comparing the four lines in this figure across race and gender groups, it appears that SIPP and DER differences are often higher for men than women and for blacks than whites. However this is not always true nor are all of the differences statistically significant. When comparing across white and non-white men, the only significant differences are in the college degree and graduate degree categories, where white men have higher differences between SIPP and DER earnings than non-white men (23.4 versus 21.9, significant at the 10% level, and 33.3 versus 20.8, significant at the 1% level). When comparing white and non-white women, the only significant difference is for women without a high school diploma (10.6 versus 14.5, significant at the 1% level). There are more significant differences when comparing men to women. White men have larger differences between self-reported and administrative earnings than white women for every education category and these differences are all significant.[10] Non-white males have higher differences than non-white females, but only in the college degree category is the difference significant.

Standard deviations are shown in the last column of Table 9 and follow similar patterns to average earnings; they are universally higher in the DER and this difference varies by education. There seems to be more dispersion in earnings as education levels increase and this is more pronounced in the DER than in the SIPP. The standard deviations are large due to the presence of a number of high earners in the right tail of the distribution.

These results are somewhat surprising given that concern about under-reporting earnings has often focused on lower-income and less educated individuals. It appears that the largest systematic differences between the SIPP and the DER happen for the most educated. One possible reason for this result might be that highly educated professionals receive a larger part of their compensation in the form of end-of-the-year bonuses and the SIPP frequently misses these one-time payments. Another reason might be definitional differences of earnings between the SIPP and the DER. Highly educated/highly compensated individuals may report a measure of earnings in the SIPP that does not include deferred compensation or some other form of compensation that these individuals consider separate from wages. However, these earnings are still reported in Box 1 or Box 13 of the W-2 and so are counted in total compensation in the DER. If these definitional differences are more likely in some types of jobs than others, this might partially explain the differences across education groups. Other possible explanations are proxy reporting and missing data imputation. Women may be more likely than men to respond themselves instead of via a proxy. Also, it is possible that imputed earnings are too low in a systematically different way across education groups.

[10] The differences between white men and white women are all significant at the 1% level except for the no high school diploma category, where the difference is significant at the 10% level.

To shed some light on the causes of these SIPP/DER discrepancies, we compare industry and occupation distributions for white men and women with graduate degrees. White women with graduate degrees are similar to all other white women with respect to their SIPP/DER earnings differential while white men with graduate degrees have substantially higher differentials than other white men. If white men with graduate degrees work in different industries and occupations than white women with graduate degrees, this might be evidence for missed bonuses and definitional differences. The results in Table 10 that show that white men and women with graduate degrees are distributed differently across industries and occupations. Highly educated white women are more likely to work in the professional services and retail industries whereas the same group of men are more likely to be in finance/insurance/real estate, manufacturing, and public administration. The occupation distribution shows that white women with graduate degrees are more highly concentrated in the teaching profession than men (almost 40% versus almost 20%) and less likely than men to be in executive/management occupations (25% versus 15%). Interestingly, when the executive/management field is broken down into finer categories, we see that there are almost equal numbers of men and women in education management but fewer women in financial management. Women out-number men in the health professions overall, but a higher percentage of men with advanced degrees are doctors and dentists than women with advanced degrees. Men are more likely to be lawyers and work in sales, including securities and financial services sales. These findings are consistent with both the hypothesis that men have more complicated compensation arrangements which translate into difficulties reporting to a survey and that the jobs of men may more often give rise to definitional differences between SIPP and DER earnings.

We also compare the rate of proxy response for these two groups and find that for white men with graduate degrees, proxies respond for at least one month of the year almost 54% of the time. For the comparable group of women, this rate is 37%. Among those whose earnings were reported by a proxy at some point during the year, 43% of men had 9-12 months of proxy reports, compared to 30% for women. Thus, for highly educated white men, we are more likely to observe the use of a proxy and for a longer period of time.

In summary, it seems there are many factors that work together to produce our results. There is likely some reporting error due to proxy respondents and difficulties reporting bonus payments, in addition to differences in how a highly paid, highly educated white male SIPP respondent views his earnings compared to how the IRS views the earnings from the same job.

4.2 Job Changers

In Table 11 we report earnings changes for individual who switch jobs. We define a job switch as an individual who reports two consecutive employers that do not overlap. There may be a gap between the ending date of job 1 and the starting date of job 2, but the starting date of job 2 cannot come before the ending date of job 1. We allow at most one job switch per individual. Some individuals have only one job and others have jobs that overlap and these individuals are not included in our table. The first row shows that on average the SIPP-reported change in earnings after switching jobs is small, about $40, and is not significantly different from zero. In contrast, in the DER, job-changers earn around $900 less on average in the second job. These results differ substantially by industry of the first and second jobs. The first group of 15 rows in Table 11 reports earning changes classified by respondent-reported industry of the first job. Four industries are similar to the overall average with small earnings gains reported in the SIPP and losses reported in the DER: Construction, Retail Trade, Personal Services, and Professional Services. Retail Trade is slightly different from the others in that the DER change is only -$6 while the SIPP change is over $600. Three other industries report positive changes in earnings from switching jobs – Agriculture, Business and Repair Services, and Entertainment and Recreation Services. The SIPP reports a higher positive change for the first two of these industries and is just slightly lower for the last one. These three industries seem to have the closest SIPP and DER reporting. The remaining seven industries have negative earnings changes between jobs in both the SIPP and the DER, but the changes are substantially smaller in the SIPP than in the DER. For example, individuals with a job in the Finance, Insurance, and Real Estate (FIRE) industry who switched jobs reported a decline in earnings of just over $400 in the SIPP, but their W-2 records show a decline of almost $2,000, on average. Figure 2 summarizes these results.

The next group of 14 lines in Table 11 shows earnings changes for individuals classified by the industry of their second job. Here, five industries have positive SIPP changes and negative DER changes, three industries have positive SIPP and DER changes, and six industries have negative earnings changes in both data sources. When earnings changes are positive in both sources, the SIPP report is substantially higher than the DER report (Public Administration, FIRE, and Manufacturing Non-durable). When both sources report negative earnings changes, the DER is substantially more negative, with the exception of Personal Services where the change is essentially the same in the two sources. It is interesting that some industries have different signs on earnings changes between the Job 1 table and the Job 2 table, for example FIRE. This is may be due to the fact that many individuals with FIRE as their original job industry, have a different industry at their second job and they earn less in this new industry. For individuals with FIRE as their second job industry, they may have switched out of a lower-paying industry. This effect is captured in both the SIPP and the DER although the magnitude of the change is different between the two data sources.

The last two lines of Table 11 show earnings changes for individuals who do not switch industries and those who do switch industries. Overall, there do not appear to be significant differences among industry changers and stayers. Their SIPP-reported changes in earnings are very similar as well as their DER earnings changes. They follow the overall pattern of the data with slightly positive earnings gains in the SIPP and larger negative earnings losses in the DER.

Overall it appears that when earnings increase at the time of a job switch, the increase is stronger in the SIPP than in the DER and when earnings decrease, the decrease is stronger in the DER than in the SIPP. This leads to the SIPP reporting earnings gains on average while the DER reports earnings losses on average.

4.3 Earnings Regressions

In Table 12, we present results from two mixed effects earnings regressions, one using SIPP data for the dependent variable and the other using DER data. The dependent variable is an employer-specific log annualized wage rate. For each year, we chose a dominant employer based on earnings and then kept all years with positive earnings for that employer. In general this leads to only one observation per person per year but sometimes when dominant employers from separate years both had earnings in a common year, there were multiple observations per year. We calculated the annualized wage rate by dividing total annual earnings by total annual hours worked. The information on hours comes from the SIPP survey and is used to create both the SIPP and DER wage rate. The explanatory variables are SIPP variables and are identical in both regressions. As fixed effects, we include an intercept, indicators for the source SIPP panel (1996 is the excluded category), a linear time trend, interactions of race and gender (white males are the excluded category), interactions of race, gender, and five levels of education (no high school diploma is the excluded category), and interactions of race, gender, and a piecewise linear spline in experience. As random effects, we include a person effect and a person/labor force experience interaction and allow these effects to be correlated. These two terms allow for individual deviations from the overall intercept and from the overall labor force experience slope due to unobservable individual characteristics. In addition, we include a random employer effect and specify an AR(1) process for the error term. This person-firm effect model is similar in spirit to Abowd, Kramarz, and Margolis (1999). We estimate the mixed effects model using Restricted Maximum Likelihood (REML) which does not impose orthogonality between the independent variables and the design matrix of the random effects, alleviating the usual concerns about random effects estimators.

The intercepts in Table 12 are remarkably similar for the SIPP and DER regressions but the panel indicators are substantially higher in the DER than in the SIPP. This result means that there is a much larger difference in the overall average of SIPP and DER earnings in the panels conducted in the early 1990s than in the panel that began in 1996, with differences between 5% and 10%. Given the fact that the SIPP was re-designed and switched to computer-assisted interviews for the 1996 panel, it is not surprising that the differences between the SIPP and DER are smaller for this panel. The linear time effect and race and gender interactions are also very similar. The coefficient on non-white males is small and positive in the DER and just barely negative in the SIPP, but it is not significantly different from zero in either data source. The education coefficients compare differently depending on the demographic group. For white men, the education coefficients are slightly larger in the DER than in the SIPP by 1% to 2.5%. The differences for college and graduate degrees are higher than the differences for high school diploma and some college. For non-white males, the education coefficients are 1% to 4% smaller in the DER than in the SIPP. White females are similar to white males, with earnings 1% to 3% higher in the DER. Non-white females with a graduate degree have the biggest gap, with the DER coefficient being 7% higher than the SIPP coefficient.

To better visualize the differences between the labor force experience splines, we have graphed the experience profiles of the four demographic groups using both the SIPP and DER spline coefficients. The intercept for each profile is predicted earnings for an individual in the 1996 SIPP panel with a high school diploma and no accumulated labor force experience. Figure 3 shows results for white men and women. Initially for men, the increase in earnings due to labor force experience accumulation is slower in the DER than in the SIPP. However, by around 14 years of experience, the DER earnings have caught up and from then on, surpass the SIPP earnings. The gap continues to widen over the later years of an individual's career. Profiles for white women follow a similar pattern. However, the cross-over point for the DER and SIPP is not until approximately 25 years of experience. As shown in Figure 4, non-white men have DER earnings well below SIPP earnings for much of the profile, reflecting a lower intercept for non-white men with a high school degree in the DER than in the SIPP and then lower growth for the first two years. After that, the DER growth rate is higher than the SIPP rate, and earnings cross at around 25 years again. For non-white women, the DER growth rate is again lower for the first two years but after this, the DER catches up very quickly and surpasses the SIPP by year 6.

In general, it appears that returns to experience in the first two years are higher in the SIPP than in the DER. For years 3-5, returns are higher in the DER. For years 6-10, the rate of return is quite close for all groups except non-white males, where the DER is substantially higher. For years 11-25 and 26+, the DER is again uniformly higher. Hence, it appears that there is a range over which the effect of labor force experience is very similar between the SIPP and the DER. However there are also portions of the profile where the estimated effect is quite different.

Finally, at the end of Table 12, we list values for the variance components of the mixed effects models. The main person effect, or person intercept as we label it in the table, is three times as large in the DER as in the SIPP. The person slope, or interaction of person and labor force experience, is 2.5 times larger in the DER than the SIPP. The firm effect is three times larger in the DER than in the SIPP. Finally, the variance of the error term is significantly higher in the DER but the AR(1) correlation coefficient is similar in both regressions. DER earnings overall have higher variance and this is reflected in the random effects.

5 Results from Comparing Data Sources: Unobservables

After exploring differences in means, we next consider a variance components model with both fixed and random effects (*i.e.*, mixed effects model) that accounts for the observable differences discussed above and quantifies the remaining differences in unobservable characteristics. Our modeling follows the spirit of Abowd and Card (1989). They examined the covariance matrix of first-differenced log earnings and tested the fit of various structural models, all of which included a measurement error component. Here, our model will rely on random person and firm effects instead of first-differencing and has the advantage of a second source of data to identify the effects, but the parsing of variance among structural components is similar. We describe our model in detail and then present results.

5.1 Model

We estimate the following SIPP earnings equation:

$$\begin{aligned} \ln\left(SIPPEARN_{ist}\right) \quad = \quad & \beta_{oSIPP} + \beta_{1SIPP} Race.Gender + \beta_{2SIPP} Race.Gender.Educ + \\ & \beta_{3SIPP} Race.Gender.Exp_{it} + \\ & \beta_{4SIPP} Time_{it} + \beta_{5SIPP}\left[P_{1990}, P_{1991}, P_{1992}, P_{1993}\right] + \\ & \theta_i + \theta_{iSIPPDEV} + \psi_j + \psi_{jSIPPDEV} + \eta_{ist} + \omega_{ist} \end{aligned} \tag{1}$$

and the DER earnings equation for the same individual is identical except for the last component:

$$\begin{aligned} \ln(DEREARN_{ist}) \quad = \quad & \beta_{oDER} + \beta_{1DER} Race.Gender + \beta_{2DER} Race.Gender.Educ + \\ & \beta_{3DER} Race.Gender.Exp_{it} + \\ & \beta_{4DER} Time_{it} + \beta_{5DER}\left[P_{1990}, P_{1991}, P_{1992}, P_{1993}\right] + \\ & \theta_i + \theta_{iDERDEV} + \psi_j + \psi_{jDERDEV} + \eta_{ist} + \upsilon_{ist} \end{aligned} \tag{2}$$

where i subscripts the individual, j subscripts the employer, s subscripts the person-firm match or job, and t subscripts the year. The variables are defined as follows:

$P_{1990}, P_{1991}, P_{1992}, P_{1993}$ = vector of 4 indicator variables specifying the SIPP panel of the individual; the 1996 panel is the excluded group

$Race.Gender$ = full interaction of male and white produces four categories: white male, non-white male, white female, non-white female

$Educ$ = four levels of education fully interacted with race and gender: high school diploma, some college, college degree, graduate degree separately for each demographic group; less than high school diploma is the excluded group

Exp_{it} = general labor market experience, interacted with race and gender: actual experience calculated using employment history collected in the SIPP; experience enters as a piecewise linear spline with nodes at 2 years, 5 years, 10 years, and 25 years; separate effects for each demographic group

$Time_{it}$ = calendar time; base year is 1990

Person heterogeneity = $\theta \sim N(0, G_1)$

Source-specific person heterogeneity = $(\theta_{SIPPDEV}, \theta_{DERDEV}) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, R\right)$

Firm heterogeneity = $\psi \sim N(0, G_2)$

Common error component = $\eta \sim N(0, G_3)$

Measurement error, SIPP and DER = $(\omega, \upsilon) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, R\right)$

G_1, G_2, G_3, R are defined below.

This model accounts for average differences in the SIPP and DER using fixed effects for the intercept, race and gender interactions, education, experience, and a time trend. The effects, θ and ψ, are random person and firm effects, respectively, and capture unobservable effects of individual and employer heterogeneity. We also interact the person and firm effects with the data source indicator in order to tell whether there is source specific variation at the person and firm heterogeneity. Thus $\theta_{SIPPDEV}$, θ_{DERDEV}, $\psi_{SIPPDEV}$, and ψ_{DERDEV} represent deviations from the main person and firm random effects. The effect η is a shared random error component that can be thought of as a nested individual-job-time period random effect. This effect is estimable due to the presence of two earnings observations for each year of the panel. It represents "economic" noise, or fluctuations in annual earnings due to unobservable economic factors that influence both earnings measures, presumably by influencing "true" underlying earnings. The final terms in the model, ω and υ, are residuals that capture any remaining variation. Strictly interpreted, these terms capture variation across time within a job that is unique to each data source.

The total number of jobs held by all individuals is N, the total number of individuals is I, the total number of firms employing individuals in the sample is J, the number of covariates included in X is k, and the total number of time periods is 10. The maximum number of time periods a job may be observed depends upon the origin SIPP panel. In the 1990, 1991, and 1993 SIPP panels, there are two years of complete earnings data. In the 1992 panel there are three years and in the 1996 panel, four years. Thus, a job may be observed anywhere from one to four years depending on the tenure of the job and the source panel.

Written in matrix notation, the model is

$$Y = X\beta + Zu + e$$

where Y is an $(N \times 10 \times 2) \times 1$ vector of stacked SIPP and DER earnings, X is an $(N \times 10 \times 2) \times k$ design matrix of covariates treated as fixed effects, β is a $k \times 1$ vector of fixed effect coefficients, Z is an $(N \times 10 \times 2) \times (I + J + N \times 10)$ design matrix of the random effects, u is a $(I + J + N \times 10) \times 1$ vector of random effects, and e is an $(N \times 10 \times 2) \times 1$ vector of residuals.

The fixed effects represent shifts in the conditional mean of the distribution of SIPP or DER earnings. For example, the β_{0SIPP} term is the mean of the entire SIPP earnings distribution and β_{0DER} is the mean of the DER earnings distribution. The vector β_{5SIPP}=[$\beta_{5SIPP1990}$, $\beta_{5SIPP1991}$, $\beta_{5SIPP1992}$, $\beta_{5SIPP1993}$] captures shifts in the mean of the panel-specific earnings distributions due to differences across SIPP panels. The equivalent vector β_{5DER} reflects shifts in the panel-specific DER earnings distributions due to the same cause.

The random effects capture variation in the data due to individual, firm or time period heterogeneity that remains after controlling for observed characteristics. In other words, there is variation around the conditional mean earnings due to unobservable characteristics of the person, employer or time period for every category of individual defined by the effects treated as fixed (X). The random effects quantify the amount of variance due to the different sources. The random effects vector, u, contains the stacked random effects, $\theta_1 ... \theta_I$, $\theta_{1SIPPDEV} ... \theta_{ISIPPDEV}$, $\theta_{1DERDEV} ... \theta_{IDERDEV}$, $\psi_1 ... \psi_J$, $\psi_{1SIPPDEV} ... \psi_{JSIPPDEV}$, $\psi_{1DERDEV} ... \psi_{JDERDEV}$, $\eta_{111990} ... \eta_{IN1999}$. The variance matrices for the random person, firm and shared error component effects, respectively, can be written as

$$
\begin{aligned}
G_1 &= I_{I\times I} \otimes \sigma_\theta^2 \\
G_{1DEV} &= I_{I\times I} \otimes \begin{bmatrix} \sigma^2_{\theta SIPPDEV} & 0 \\ 0 & \sigma^2_{\theta DERDEV} \end{bmatrix} \\
G_2 &= I_{J\times J} \otimes \sigma_\psi^2 \\
G_{2DEV} &= I_{JxJ} \otimes \begin{bmatrix} \sigma^2_{\psi SIPPDEV} & 0 \\ 0 & \sigma^2_{\psi DERDEV} \end{bmatrix} \\
G_3 &= I_{NxN} \otimes \sigma_\eta^2 \begin{bmatrix} 1 & \rho & \rho^2 & ... & \rho^9 \\ \rho & 1 & \rho & ... & ... \\ \rho^2 & \rho & ... & ... & \rho^2 \\ ... & ... & ... & 1 & \rho \\ \rho^9 & ... & \rho^2 & \rho & 1 \end{bmatrix}_{10\times 10} \\
\text{where } \sigma_\eta^2 &= \frac{\sigma_\varsigma^2}{(1-\rho^2)}.
\end{aligned}
$$

The shared error component is modeled as an $AR(1)$ process where errors are correlated within the same job for a given individual but not across jobs nor across individuals. The $i.i.d.$ shock in the $AR(1)$, ς_{ijt}, has variance σ_ς^2. The person and firm deviation effects reflect that some additional, uncorrelated variation might exist in either the SIPP or the DER or both.

The error vector, e, contains the stacked terms, $\omega_{111990} ... \omega_{IN1999}$,$\upsilon_{111990} ... \upsilon_{IN1999}$. The SIPP and DER errors follow separate $AR(1)$ processes with the covariance between them constrained to be zero. These errors are identified by differences in the SIPP and DER earnings reports for each year, given all other effects in the model. The variance matrix for the residuals can be written as

$$R = I_N \otimes \begin{bmatrix} \sigma^2_\omega \begin{bmatrix} 1 & \rho_{sipp} & \rho^2_{sipp} & ... & \rho^9_{sipp} \\ \rho_{sipp} & 1 & \rho_{sipp} & ... & ... \\ \rho^2_{sipp} & \rho_{sipp} & ... & ... & \rho^2_{sipp} \\ ... & ... & ... & 1 & \rho_{sipp} \\ \rho^9_{sipp} & ... & \rho^2_{sipp} & \rho_{sipp} & 1 \end{bmatrix} & 0 \\ 0 & \sigma^2_\upsilon \begin{bmatrix} 1 & \rho_{der} & \rho^2_{der} & ... & \rho^9_{der} \\ \rho_{der} & 1 & \rho_{der} & ... & ... \\ \rho^2_{der} & \rho_{der} & ... & ... & \rho^2_{der} \\ ... & ... & ... & 1 & \rho_{der} \\ \rho^9_{der} & ... & \rho^2_{der} & \rho_{der} & 1 \end{bmatrix} \end{bmatrix}$$

where ρ_{sipp} and ρ_{der} are the autocorrelation terms of the SIPP (ω) and DER (υ) errors, respectively, and the submatrices are all (10×10).

Estimates of β_{0SIPP} to β_{5SIPP}, β_{0DER} to β_{5DER}, the variance components (σ^2_θ, $\sigma^2_{\theta SIPPDEV}$, $\sigma^2_{\theta DERDEV}$, σ^2_ψ, $\sigma^2_{\psi SIPPDEV}$, $\sigma^2_{\psi DERDEV}$, σ^2_η, ρ, σ^2_ω, ρ_{sipp}, σ^2_υ, ρ_{der}) and realizations of the random effects (θ, $\theta_{SIPPDEV}$, $\theta_{SIPPDEV}$, ψ, $\psi_{SIPPDEV}$, ψ_{DERDEV}, η) and the residuals (υ, ω) can be obtained by solving the restricted maximum likelihood (REML) problem and the mixed model equations:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\beta} \\ \widehat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ Z'R^{-1}Y \end{bmatrix}.$$

The estimation is done by REML using an average information (AI) algorithm, developed and programmed by Gilmour, Thompson, and Cullis (1995). This method closely follows the Fisher scoring algorithm proposed by Patterson and Thompson (1971). Parameters are chosen to maximize the log likelihood function by satisfying a set of first order conditions, or score equations. Solutions to the score equations are calculated iteratively. The user furnishes a set of starting values for the variance components and the algorithm calculates the log likelihood and produces initial estimates of the fixed effects (β) and the realized random effects (u). The information matrix is calculated using an averaging method that simplifies the process for large data sets with multiple random effects. The information matrix is then used to update the variance component estimates. The process is repeated until the estimates converge.

This model is similar to the earnings regressions described in Section 4.3 but with some important differences. First, the sample of people-job matches used in the estimation is different. All person-job matches that have SIPP and DER earnings in some years are used to estimate variance components. We do not select a dominant employer because we want to quantify the sources of variance for all the data, not estimate economic relationships. Second, we use earnings not wages because it is variation in earnings that is of interest. Finally, we jointly estimate equations 1 and 2 so that there are some common variance components and some components particular to either the SIPP or the DER.

5.2 Variance Components and Reliability Ratios

In Table 13, we show the estimated variance components. The main person and firm effects are 0.28 and 0.32, respectively. The interactions of person, firm and data source produce variance components that go to zero for the SIPP. Essentially, there is no variation left in the SIPP at the person and firm levels after taking account of the variation that is common to both the SIPP and the DER. However, there is additional variation in the DER, $Var\left[\theta_{DERDEV}\right] = 0.04$ and $Var\left[\psi_{DERDEV}\right] = 0.1$. These magnitudes imply that about 25% of the variation due to unobservable firm characteristics in the DER is not found in the SIPP and about 10% of the variation due to unobservable person characteristics. The variance in the SIPP measurement error term is also lower than the DER measurement error term and the SIPP error is less correlated over time. The common time period component has a higher variance than either measurement error term. The magnitudes imply that for time-period-specific variation in the SIPP, 23% is unique to the SIPP and 33% is unique to the DER.

Our estimation of this model allows us to parse variation due to unobservables into common variation and source-specific variation. However, we cannot talk about measurement error without making an additional assumption: namely what is true variation? One possible assumption is that all the common components are true variation and the deviations $\theta_{SIPPDEV}$, θ_{DERDEV}, $\psi_{SIPPDEV}$, ψ_{DERDEV}, ω, and υ are measurement error. Using this assumption, we can calculate the reliability ratio, as commonly used in the literature, that compares true variation to total variation. If only the common variation is considered true, then the formulas are as follows:

$$
\begin{aligned}
\kappa_{SIPP} &= \frac{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_\psi^2}{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_{\theta SIPPDEV}^2 + \sigma_\psi^2 + \sigma_{\psi SIPPDEV}^2 + \sigma_\omega^2} \\
\kappa_{DER1} &= \frac{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_\psi^2}{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_{\theta DERDEV}^2 + \sigma_\psi^2 + \sigma_{\psi DERDEV}^2 + \sigma_\upsilon^2}
\end{aligned} \tag{3}
$$

Another possible assumption is that the common variation and the DER person and firm deviations are true and the SIPP deviations are measurement error. Under this assumption, the reliability ratio for the SIPP remains the same but for the DER it becomes:

$$\kappa_{DER2} = \frac{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_{\theta DERDEV}^2 + \sigma_\psi^2 + \sigma_{\psi DERDEV}^2}{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_{\theta DERDEV}^2 + \sigma_\psi^2 + \sigma_{\psi DERDEV}^2 + \sigma_\upsilon^2}$$

A third possible assumption is that all variation in the DER is true in which case the reliability ratio for the DER becomes:

$$\kappa_{DER3} = \frac{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_{\theta DERDEV}^2 + \sigma_\psi^2 + \sigma_{\psi DERDEV}^2 + \sigma_\upsilon^2}{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_{\theta DERDEV}^2 + \sigma_\psi^2 + \sigma_{\psi DERDEV}^2 + \sigma_\upsilon^2} = 1.$$

These assumptions provide a range of reliability ratios for the DER data. If the range is small, because κ_{DER1} is large, this will indicate that most of the variation found in the DER is also found in the SIPP. If κ_{DER1} is small, then we conclude that there is a substantial amount of variation that is unique to the DER. In the final three rows of Table 13, we present calculations of κ_{SIPP} , κ_{DER1}, and κ_{DER2} . The magnitudes of κ_{SIPP} and κ_{DER1} indicate that approximately 70% of total DER variation due to unobservables is common to both the SIPP and the DER. In the SIPP 86% of total variation is common to both sources. If we assume that $\sigma_{\theta DERDEV}^2$ and $\sigma_{\psi DERDEV}^2$ are true variation, then the appropriate reliability measure for the DER is $\kappa_{DER2} = 0.80$. Thus the range of ratios for the DER, depending on how much of the DER variation one is willing to term truth, is 0.72 to 1.

The fact that $\kappa_{SIPP} > \kappa_{DER1}$ reflects two things. First, overall variation in the SIPP is lower than in the DER. Second, σ_ω is smaller than σ_υ. If we believe that the SIPP is missing variation that is actually true variation, then a higher reliability ratio is not an indication of less measurement error. If one chooses to adopt the hypothesis that $\sigma_\eta^2 + \sigma_\theta^2 + \sigma_\psi^2$ represents true variation and all other variation is measurement error of some kind, then the overall level of SIPP variation is 14% too high. However, if one chooses instead to adopt the hypothesis that $\sigma_\eta^2 + \sigma_\theta^2 + \sigma_{\theta SIPPDEV}^2 + \sigma_\psi^2 + \sigma_{\psi SIPPDEV}^2$ is true variation, then overall variation in the SIPP is too low. To see this point, consider the ratio of common to true variance under the second hypothesis

$$\frac{\text{Common variance}}{\text{True variance}} = \frac{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_\psi^2}{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_{\theta DERDEV}^2 + \sigma_\psi^2 + \sigma_{\psi DERDEV}^2}.$$

This ratio is 0.91 which indicates that about 9% of true variation is missing from the SIPP. In other words, 14% of SIPP variation found at the person-job-time period level is measurement error, but 9% of variation due to person and firm heterogeneity is missing. If one chooses to believe that all variation in the DER is truth, then the ratio becomes:

$$\frac{\text{Common variance}}{\text{True variance}} = \frac{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_\psi^2}{\sigma_\eta^2 + \sigma_\theta^2 + \sigma_{\theta DERDEV}^2 + \sigma_\psi^2 + \sigma_{\psi DERDEV}^2 + \sigma_\upsilon}$$

which is equal to 0.72. Under this hypothesis, about 28% of true variation due to person, firm, and time period heterogeneity is missing from the SIPP. This later ratio is exactly equal to κ_{DER1} but the interpretation is different because of the different assumption made about the DER.

Which hypothesis about the DER to adopt is not answered by the data. However, it is interesting to note that σ_υ and ρ_υ are both relatively large in magnitude. Thus, the DER-specific time period effect accounts for more variation than the total DER person effect and, unlike classical measurement error, it does not immediately die out in the next time period. Because of these estimates, we hesitate to label υ as strictly measurement error and hypothesize that the true reliability ratio of the DER is somewhere between 0.8 and 1, and that the SIPP is missing between 9% and 28% of true variation.

6 Conclusion

In comparing the SIPP and the DER we have found two consistent results. DER earnings are on average higher than SIPP earnings and there is more variation due to unobservables in the DER than the SIPP. Of the definitional differences discussed earlier, it appears that lack of health insurance premiums and EIN changes are not dominant factors since these would give rise to lower DER earnings on average. We cannot say for certain how much of the differences we find might be due to employee benefits appearing on a W-2 form and not on a pay stub. However, our opinion is that these differences are unlikely to be solely the result of the SIPP and the DER measuring different quantities. In particular, it seems likely to us that highly educated SIPP respondents with high incomes do under-report their earnings to some extent. Those wishing to study high earners might be cautioned against using SIPP data without the link to administrative earnings. Our data on job changers are particularly interesting in that the major differences between the SIPP and the DER are in earnings changes not levels.

Our results examining differences due to unobservables lead us to believe that there is too little variation in SIPP earnings. Without further research, we cannot give a definite reason why this might be the case. However, we hypothesize that difficulty in capturing with-in year fluctuations in pay and Census hot-deck imputation procedures contribute to the lower SIPP variation. We also believe that definitional differences in earnings play a substantial role. The SIPP earnings distribution is truncated on the right because high-earners do not report some types of pay that in fact appear on their W-2 forms.

The SIPP collects earnings at the monthly level, so to some extent there will always be difficulties in comparisons to an administrative data source that is annual. In light of these results, it might be useful to consider ways in which the SIPP could better capture an annual earnings measure that included bonuses and irregular extra earnings. Also, an imputation procedure which used the DER to help model SIPP earnings could reduce the differences between the two sources, by allowing draws for SIPP earnings values to be taken from a distribution with greater variance.

In spite of the differences we find, we feel that there are reasons to be confident in the use of SIPP data. Of the variation that is found in the SIPP, 86% of it can also be found in the DER. Earnings regressions using SIPP and DER data produce similar coefficients. Researchers studying returns to experience would draw similar conclusions from SIPP and DER data. As our understanding of these two data sources continues to develop, different measures of earnings may emerge that combine information from both survey and administrative records to create something that might be closer to "true" earnings.

References

[1] Abowd, John M. and David Card, "On the Covariance Structure of Earnings and Hours Changes," *Econometrica,* Volume 57, Number 2 (March 1989): 411-446.

[2] Abowd, John M., Francis Kramarz, David N. Margolis, "High Wage Workers, High Wage Firms," *Econometrica,* Volume 67, Number 2 (March 1999): 251-333.

[3] Abowd, John M. and Lars Vilhuber, "The Sensitivity of Economic Statistics to Coding Errors in Personal Identifiers," *Journal of Business and Economic Statistics,* Volume 23, Number 2 (April 2005): 133-152.

[4] Angrist, Joshua and Alan B. Krueger, "Empirical Strategies in Labor Economics" in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Elsevier, 1999, Volume 3, Part 1, pp. 1277-1366.

[5] Bound, John and Alan B. Krueger, "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?," *Journal of Labor Economics,* Volume 9, Number 1 (January 1991): 1-24.

[6] Bound, John, Charles Brown, Greg J. Duncan, Willard L. Rodgers, "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data," *Journal of Labor Economics,* Volume 12, Number 3 (July 1994): 345-368.

[7] Bound, John, Charles Brown and Nancy Mathiowetz, "Measurement Error in Survey Data" in J.J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Elsevier, Volume 5, 2001, pp. 3705-3843.

[8] Duncan, Greg J. and Daniel H. Hill, "An Investigation of the Extent and Consequences of Measurement Error in Labor-economic Survey Data," *Journal of Labor Economics,* Volume 3, Number 4 (October 1985): 508-532.

[9] Fellegi, Ivan P. and Alan B. Sunter, "A Theory for Record Linkage," *Journal of the American Statistical Association,* Volume 64, Number 328 (December 1969): 1183-1210.

[10] Fuller, Wayne, *Measurement Error Models*. Wiley, 1987.

[11] Gilmour, Arthur R., Robin Thompson, Brian R. Cullis, "Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models," *Biometrics,* Volume 51, Number 4 (December 1995):1440-1450.

[12] Gottschalk, Peter and Minh Huynh, "Are Earnings Inequality and Mobility Overstated? The Impact of Non-Classical Measurement Error," *Review of Economics and Statistics*, Volume 92, Number 2 (May 2010): 302-315.

[13] Mellow, Wesley and Hal Sider, "Accuracy of Response in Labor Market Surveys: Evidence and Implications," *Journal of Labor Economics*, Volume 1, Number 4 (October 1983): 331-344

[14] Newcombe, H.B., J.M. Kennedy, S.J. Axford, and A.P. James, "Automatic Linkage of Vital Records," *Science,* Volume 130, Number 3381 (October 1959): 954-959.

[15] Patterson, H.D. and R. Thompson, "Recovery of Interblock Information when Block Sizes are Unequal," *Biometrika*, Volume 58, Number 3 (1971): 545-554.

[16] Rubin, Donald B. "Inference and Missing Data," *Biometrika*, Volume 63, Number 3 (1976): 581-92.

[17] Stinson, Martha H. 2003. "Technical Description of SIPP Job Identification Number Editing, 1990-1993 SIPP Panels." *SIPP Technical Paper*, U.S. Census Bureau, available at http://www.census.gov/sipp/core_content/core_notes/ DescriptionSIPPJOBIDEDITING.pdf, cited February 21, 2011.

7 Appendix A: Editing of SIPP JOB ID Variable

As described in Section 3.1, during our initial work to create job records by linking SIPP wave data, we discovered that identifiers created to track jobs over time in fact had difficulty correctly linking wave reports that belonged to the same job. In the 1996 SIPP panel, the largest problem with jobs arose when the jobs had start dates prior to the beginning of the first wave in which they were reported and prior to the beginning of the previously held job. Table A1 gives one generic example of the cause of this problem. In this case, the individual was interviewed in waves 1 through 4 and reported a job which began February 1, 1996. However, the individual missed the fifth interview. When the next interview was conducted in wave 6, a new job was reported but the start date was prior to the beginning of wave 6 and prior to the beginning of job 1. The CAPI system was not designed to allow job IDs to be carried forward through missed interviews. Consequently, when this person temporarily dropped out of the panel, she was automatically given a new job ID at the time of the next interview, regardless of whether the job had actually begun in wave 6 or not. However, there were no restrictions placed on the start date she reported and hence this discrepancy arose. The case illustrated in Table A1 was the most common cause of the early start date problem. However it was not the only cause. The problem affected 21.6% of all jobs (29,520) and about 40% of the time there appears to have been a missing wave problem, while the rest of the time, the cause could not be determined. Whatever the reason, it was clear that the survey job IDs sometimes failed to link job records correctly.

The problems encountered in the early SIPP panels (1990-1993) were considerably more complicated. There were two major types of problems – improper re-use of job IDs and improper assigning of new job IDs. Tables A2 and A3 give generic examples of these problems. In Table A2, the SIPP respondent held the same job throughout the first four waves of the survey; however, in wave 3, the job ID was incorrectly changed, causing a false job transition. This error is identifiable because the name of the employer stays the same across the waves. Table A3 shows the second type of problem. In this case, the person changed jobs between waves 3 and 4 but the job ID was not changed. Thus, it appears that the person remained at the same job through all four waves and, consequently, a job transition was missed. Again, the true work history is apparent only through scrutinizing the employer names.

We developed an editing procedure that used employer name and person-level total job counts from the DER data to identify and correct SIPP job ID coding errors. In the early SIPP panels, the problems were mostly the result of field representatives being required to collect information about an on-going job over and over again. Inconsistencies crept in over time as the name of the employer was collected and written down separately at the time of each interview. Wave-specific names differed both across and within the original SIPP job IDs. Different spellings, use of abbreviations in later waves, and slightly different wording were the most common differences within job IDs. In contrast, the 1996 panel only recorded employer names when new jobs were begun and hence employer names differed only across job IDs and not within.

For the 1990-1993 panels, the goal was to create an entirely new set of job IDs that was not derived from the old job IDs because these were deemed too unreliable. Hence, it was necessary to compare all person-job-wave records for a given individual and group those with the same name.

In the 1996 panel, however, the goal was simply to check jobs for an individual to determine if they should be linked because an individual may have missed a survey wave and been incorrectly assigned a new job ID when he or she was next interviewed. Hence, person-job-wave observations with the same job ID assigned were accepted as belonging to the same job and job-level records were created. These job-level records were then compared and those with names deemed to be the same were grouped together.

Because of the spelling and wording differences across observations, we used probabilistic record matching methods as developed by Newcombe, Kennedy, Axford, and James (1959) and Fellegi and Sunter (1969), and implemented in a commercial software program called Integrity. These methods have been used extensively at the U.S. Census Bureau to solve problems of miss-coded identifiers (for an example of an application of probabilistic person name matching to fix SSN miscodes, see Abowd and Vilhuber (2005)). This method involves grouping records into "blocks" of possible matches and then computing matching weights, sometimes called disagreement indices, for pairs of records within the "block". Pairs with matching weights above a certain threshold, or cutoff point, are deemed to be matches and those with weights below another threshold are deemed to be non-matches. Those pairs with matching weights in between the two thresholds are termed uncertain and clerical review is suggested.

A matching weight for a pair of records is a composite score that is created by comparing the records across a variety of fields, assigning a weight to each field based on a determination of whether the field agrees or disagrees, and then summing the weights from all the fields involved in the comparison. Each field used in the matching is assigned an m and u probability. The m probability is the probability that the same field on two separate records agrees given that the two records were indeed a match. When this probability is set to less than one, it is assumed that there are some errors in the fields and that even if two records are a match, there is still some probability that the field is miscoded on one of the records and the two fields will disagree. The u probability is the probability that the same field on two separate records agrees given that the records are not a match. This is the probability that a field agrees at random. Given the m and u probabilities, agreement and disagreement weights for each field are calculated using the following formulas:

$$\begin{aligned}\text{agreement weight} &= \log_2(\frac{m}{u}) \\ \text{disagreement weight} &= -(\log_2(\frac{1-m}{1-u}))\end{aligned}$$

The decision of whether a field agrees or disagrees, and hence whether it receives the agreement or disagreement weight, can be implemented in a variety of different ways. One can be quite strict and insist on absolute identity in order to declare agreement or one can allow some level of discrepancy between fields without declaring disagreement. This flexibility is especially useful for name matching because it allows the user to take account of potential misspelling of words.

In our application, we blocked on the SIPP person identifier and hence only job records for the same person were compared. To create the fields for comparison we parsed the reported name into several pieces. Common words such as "Inc," "Company," or "Firm" were saved in one set of fields while geography words such as state names were saved in another set. The remaining words from the name were thought most likely to be unique to a particular employer and were saved in a third set of fields. We performed several sets of comparisons, or passes, using different fields in each pass. The choice of m and u probabilities and cutoff levels was determined both by knowledge about the fields and by experimentation. For the fields that contained unique name words, a high m probability and low u probability were chosen. Since these words were deemed to be the part of the employer name that was unique to that firm, matching values were essential to matching records, thus requiring the high m probability. At the same time, these words were unlikely to agree at random and hence produce false matches, so a low u probability was chosen. The result of these choices was that matching values of the unique word names received very high agreement weights and also very high disagreement weights. The fields that contained common words and geography words, on the other hand, had higher u probabilities. Agreement in one of these fields produced a lower agreement weight because matches were more likely to happen at random while disagreement produced a more negative disagreement weight because non-matches meant the companies were unlikely to be the same. Cutoff values were chosen by examining certain and uncertain matches and determining the range of their weights. Appendix Table A4 gives the exact blocking and matching fields used along with their m and u probabilities.

A first pass of record linking produced a preliminary set of new job IDs. Using these new IDs, we counted the total number of jobs held over the course of the survey for each individual. We performed a similar count in the administrative DER data. We compared these two counts to identify cases where the name matching software had failed to correct or had introduced new job history errors. In the 1996 panel, the job count comparison showed that the name matching step had corrected the most obvious problems and further editing was deemed unlikely to provide enough improvements to be worth the resource cost involved.

For the early SIPP panels, however, probabilistic name matching alone proved inadequate for creating a consistent set of job IDs. While the name matching procedure both separated jobs records originally assigned the same job ID and connected job records originally assigned different job IDs, the former was the most common outcome. This can be seen in the first row of Table A5 where the number of total jobs rose substantially after name matching. This result was due to the fact that the most common problems in the survey were the re-use of job IDs, as described in Table A3, and the high degree of irregularity in the spelling of job names and the common use of abbreviations in later waves. Universities and government agencies with common acronyms were especially problematic. For example, the Integrity software could not recognize the names "University of X" and "UofX" or "Department of Y" and "DOY" as being the same. Hence, in these panels, a second pass with the name matching software was performed and then a considerable amount of clerical review was undertaken in order to separate cases where Integrity correctly and incorrectly split job records, as determined by the trained human editors. In cases where there were discrepancies between the total job count in the SIPP and the total job count in the DER after both name matching passes, job records were output and reviewed by two separate individuals. When one of the reviewers discovered two jobs in a respondent's job history that appeared to be the same, she manually changed the job ID to reflect this determination. The second reviewer re-checked all these changes as a quality assurance measure. After this extensive manual review, a few final edits were performed to locate any final obvious cases where Integrity had erroneously failed to link job records. The work history of any person who had one job that consisted of at least four linked job records and a second job that consisted of only one job record was examined to see whether the single job record in fact belonged to the job with at least four linked records. Corrections to job IDs were made to link job records that were determined to belong to the same job.[11]

Tables A5 and A6 provide a summary of the process for all the SIPP panels, showing how many unique jobs resulted from each step of the editing procedure and how many records were affected at each step. Row 5 of Table A5 and row 2 of Table A6 show the final number of jobs defined by the revised set of job IDs and row 6 and row 3, respectively, show the number of jobs belonging to people who still have discrepancies between job counts in the DER and the SIPP. We are confident that the majority of these cases are the result of reporting differences between the survey and the administrative data and not failure to link job records in the SIPP. Recognizing that a similar edit could not be performed by researchers lacking access to the DER data, the Census Bureau has publicly released the revised SIPP job IDs for the 1990-1993 SIPP panels on the SIPP website, http://www.bls.census.gov/ sipp_ftp.html#sipp_jobid (cited on February 21, 2011).

8 Appendix B: Linking DER and the Business Register

The merge between the DER and the Business Register was somewhat complex because the Business Register has two parts. The first part, called the Single-unit (SU) file, contains records for all EINs that were either single-unit companies or sub-masters. Single-unit companies were firms with only one establishment that had a single EIN. Sub-masters were companies with multiple establishments that shared an EIN, *i.e.* multi-unit (MU) companies. For SU companies, the names and industries found on the SU Business Register were likely to correspond to the names and industries of employers reported in the SIPP. However for sub-masters, the name and industry were potentially quite different because these represented some aggregate concept – name of parent company or major industry out of a group of industries represented within a multi-unit company. Hence for sub-masters, we also searched for information about the EIN in the second part of the Business Register, the MU file. Here, we obtained multiple records for each EIN representing the names and

[11] A full description of the process for the 1990-1993 panels can be found in Stinson (2003).

industries of all the different establishments associated with a sub-master record. For these MU companies, we kept one record for each unique three-digit industry. Establishments within the same industry tended to have extremely similar names and hence this choice resulted in both a manageable number of observations to match to SIPP jobs while still providing additional information that might assist in the match.

The research Business Register files are maintained on an annual basis. Initially, an EIN from a job was sought in the Business Register year that corresponds to the first year the job was reported in the DER. If a job was already in progress at the time of the beginning of the survey, the start year was coded to be the first survey year since this was the first year the job was at risk to match to the SIPP. If the job was not found in the Business Register year corresponding to the start year, it was sought in the following two Business Register years. Appendix Table B1 presents a summary of the match rates between the DER and the Business Register. There are several interesting things to notice in this table. First, the match rates are extremely high, 98% for every panel except 1996. The low match rate for the 1996 panel relative to the other four panels can be explained by the fact that the latest year for which the Business Register was available at the time this research was conducted was 1999. Thus any job in the DER for a SIPP respondent from the 1996 panel that began in the year 2000 could not be matched to the Business Register. For the purposes of this study, this lack of data did not present a serious problem because so little SIPP data was collected in 2000 that annual earnings from jobs beginning in 2000 could not be accurately constructed for SIPP jobs. As described in Section 3.3, jobs beginning in the year 2000 were dropped from both the SIPP and the DER before comparing earnings. The second interesting thing to note is that although only 27% to 32% of all EINs were multi-unit companies, these EINs accounted for 39% to 44% of all jobs. SIPP respondents disproportionately work for multi-unit companies, which tend to be larger than single-unit employers. Third, a small percentage of EINs and jobs were found in the MU file but not in the SU file. The cause of this is unknown at this time and will need further research.

9 Appendix C: Linking Job-level SIPP and DER Records

The job-level match between the SIPP and DER data compared employer name, calendar year indicators, and industry in order to link records from each source. On one side of the match were all the SIPP jobs deemed to be reports of employment at a single employer. On the other side of the match were all the records associated with the DER jobs deemed to have taken place during the at-risk time frame. Each DER record contained the name and industry of the EIN as found on the SU part of the Business Register. When the EIN was also found on the MU part of the Business Register, the record contained a second name and industry representing information about a particular establishment of this EIN. When an EIN was associated with multiple establishments with different industries on the MU file, multiple records were created for this DER job. Each record contained the same SU name and industry information but different MU name and industry information. This was done in an attempt to maximize the number of job matches obtained by using all possible name information associated with multi-unit companies. For example a person might report working for company X in the SIPP and have a job report in the DER with EIN A that is a multi-unit. The main company name of EIN A may be Y but one of the subsidiary establishments may be called X. By attaching names X and Y to EIN A in the DER, we increase the likelihood that this job will match correctly to the SIPP job reported at company X.

Appendix Table C1 gives the blocking and matching fields for each pass along with the accompanying m and u probabilities for the matching fields. Several variables were also used in multiple passes, with the requirements for matching gradually relaxed. For example, in the third pass, three-digit SU industry was used as a blocking variable and the four year-indicators were used as matching variables. Pass five was quite similar except that instead of requiring records to match on all four year-indicators, only start year was required to match. Start year was a field that indicated the first year that a record was found for this job with the first possible year being the year that data was first collected in the survey and the last possible year being the last year data was collected in the survey. Likewise in pass seven, only one-digit Single-unit industry was used as a blocking variable. This process enabled the detection of high-probability matches in early passes and then the addition of lower-probability matches in later passes.

Appendix Table C2 shows the results of the matching. Of the SIPP jobs, between 77% to 79% were successfully matched to a DER job. Of these matches, 86%-88% were deemed high probability matches that surpassed the clerical editing threshold, while the remaining matches were between the clerical threshold and the no-match cutoff point or were duplicate matches. The majority of the matching took place in the first pass (between 75% and 83% of all matches). The next most successful passes were 3 (5% to 9%) and 7 (5% to 6%).

Appendix Tables C3 and C4 highlight two problems that resulted from the matching. First, two different SIPP jobs could match to the same DER job. An example of this case is illustrated in Table C3. There were several possible causes of the problem. First, it was possible that the two SIPP jobs were indeed the same and the SIPP job creation phase erroneously failed to link them. In this case the duplicate record was a "true" duplicate and both jobs were correctly matched to one DER job. However, another possibility was that the matching software mistakenly matched a second SIPP job to the same DER job due to lack of differentiating information for the SIPP jobs. This was particularly likely in the later passes where matches were based on year and industry indicators alone. In this case, the duplicate was a false match and only one of the two matches was correct. Careful clerical inspection of duplicate cases led to the adoption of the following rule: if the two SIPP jobs had been matched to the one DER job in either the first or second pass and there were two or fewer residual DER jobs left that had not matched to any SIPP job, then the second SIPP job was declared a true duplicate. It was combined with the first SIPP job to become one single SIPP job matched to the one DER job. Otherwise, if the two SIPP jobs had matched to the one DER job in pass 3 or later or they had matched in pass 1 or 2 and there were three or more residual DER jobs, the duplicate was declared false and only the master record match was kept. The duplicate SIPP job was changed to a residual, non-matching, SIPP job. The total number of duplicates that were determined to be "true" and hence were subsequently combined is shown in row 3 of Table 9.

The second problem was the reverse duplication issue: two different DER jobs sometimes matched to the same SIPP job. Table C4 gives an example. This type of duplication was more common and it was more difficult to know the causes. The first possibility was that a company changed its EIN due to a change in ownership structure or some other reason. This is the successor-predeccessor problem described earlier. Another possibility was that SIPP respondents reported "lump" jobs, meaning that one SIPP job was really a combination of several jobs. Since administrative records pertained to the legal source of the earnings, it was possible that some individuals considered themselves as holding only one job but were paid from several different source EINs. It was also possible that individuals consciously grouped jobs in order to ease the burden of responding to the survey. These issues warrant further research.

We made a first attempt to tell whether two DER jobs that matched one SIPP job were indeed the "same" job by using some additional information from the Business Register. Previously, we had augmented our list of EINs from the DER to include parent company information and name and industry information from one establishment of every unique three-digit industry group within the parent company. We then added annual geography information (a geocode created from the exact address) to each EIN at both the parent company level and the establishment level. We compared this geography information for each year of the survey across the two EINs and if the geocode was ever the same for the parent company or the establishment, we declared the two jobs to be duplicates. The intent of this geocode comparison exercise was to find cases where an EIN changed but the physical location did not change and hence it was likely that the SIPP respondent still considered himself to be at the same job. Since we did not keep every establishment within an industry group, we clearly did not compare every possible geocode. Hence, our determination of how many DER jobs were duplicates and should be combined is probably an undercount.

We also added parent company identifiers to the EINs so that we could tell if two EINs had some kind of ownership relationship. Two DER jobs that matched to one SIPP job but had the same parent company identifier were also declared to be a match. In this case it seemed possible that the SIPP respondent had kept the "same" job but had moved within the company or had simply experienced a company re-organization where the EIN tax reporting structure had changed. Row 4 of Table C2 shows how many DER job duplicates were determined to be legitimate.

Table 1: Original SIPP Job Summary					
SIPP Panel	1990	1991	1992	1993	1996
Total SIPP respondents	69,432	44,373	62,412	62,721	116,636
Respondents who ever report a job	37,291	23,520	33,920	32,972	63,600
Person-job-wave observations	216,851	136,693	228,214	208,748	498,553
Jobs defined by original SIPP jobid	57,800	35,515	55,453	52,591	136,550
Jobs defined by revised SIPP jobid	66,991	40,818	65,278	61,094	125,358

Table 2: Jobs from the DER						
	SIPP Panel	1990	1991	1992	1993	1996
1	Years covered by survey	1989-1992	1990-1993	1991-1995	1992-1995	1995-2000
2	DER jobs in survey time	96,086	58,020	99,524	81,320	192,720
3	EINs	60,131	38,628	62,406	51,880	105,095

Table 3: Match Rates for People with Jobs in the SIPP and DER					
				Both	
SIPP Panel		SIPP	DER	SIPP	DER
1990	People with jobs	37,291	35,032	30,993	
	Total jobs held	66,991	96,086	55,087	88,324
1991	People with jobs	23,520	21,729	19,056	
	Total jobs held	40,818	58,020	32,447	52,797
1992	People with jobs	33,920	31,557	27,394	
	Total jobs held	65,278	99,524	51,650	90,360
1993	People with jobs	32,972	29,831	26,267	
	Total jobs held	61,094	81,320	47,723	74,317
1996	People with jobs	63,116	55,894	48,542	
	Total jobs held	121,450	192,720	97,149	173,623

Table 4: Causes of Match Failures for SIPP and DER Records

Panel A: Reasons SIPP Workers Do Not Match DER

SIPP Panel	Total SIPP People	SIPP People without Valid SSNs	People Who Have Only SIPP Jobs	People in SIPP and DER
1990	37,291	4,856	1,442	30,993
1991	23,520	3,629	835	19,056
1992	33,920	5,477	1,049	27,394
1993	32,972	5,535	1,170	26,267
1996	63,116	12,425	2,149	48,542

Panel B: Reasons DER Workers Do Not Match SIPP

SIPP Panel	Total DER people	DER People without Valid SSNs	People Who Have Only DER Jobs	People in SIPP and DER
1990	35,032	0	4,039	30,993
1991	21,729	0	2,673	19,056
1992	31,557	0	4,163	27,394
1993	29,831	0	3,564	26,267
1996	55,894	0	7,352	48,542

Table 5: Final Sample of Matched Jobs						
SIPP Panel	1990	1991	1992	1993	1996	Total
Number of matched jobs after combining duplicates	41,885	25,258	39,729	36,469	75,110	218,451
Jobs w/out SIPP and DER earnings in sample years	5,716	3,497	2,706	6,904	2,291	21,114
New matched job total	36,169	21,761	37,023	29,565	72,819	197,337

Table 6: Covariance/Correlation Matrix for Ln(SIPP Job Annual Earnings)

	1990	1991	1992	1993	1994	1996	1997	1998	1999
1990	2.0293	0.61							
1991	1.0170	2.0643	0.54						
1992		0.8618	1.8204	0.57	0.51				
1993			0.9260	1.8653	0.74				
1994			0.7042	1.0943	1.9877				
1996						2.0875	0.72	0.66	0.63
1997						1.1456	2.0732	0.72	0.66
1998						0.8343	1.0889	2.0162	0.72
1999						0.7060	0.8084	1.0457	1.8932

Notes: Covariances on and below the diagonal; correlations above the diagonal.

Table 7: Covariance/Correlation Matrix for Ln(DER Job Annual Earnings)

	1990	1991	1992	1993	1994	1996	1997	1998	1999
1990	1.9604	0.81							
1991	1.1795	2.0640	0.80						
1992		1.2313	2.0615	0.80	0.74				
1993			1.2486	2.1542	0.79				
1994			1.0129	1.2330	2.1987				
1996						2.2320	0.80	0.75	0.71
1997						1.3094	2.2750	0.80	0.75
1998						1.0805	1.3083	2.3040	0.80
1999						0.9456	1.0736	1.3284	2.2816

Notes: Covariances on and below the diagonal; correlations above the diagonal.

Table 8: Correlation Matrix of SIPP/DER Job Annual Earnings

	Ln(SIPP Job Annual Earnings)								
Ln(DER Job Annual Earnings)	1990	1991	1992	1993	1994	1996	1997	1998	1999
1990	0.74	0.58							
1991	0.59	0.75	0.58						
1992		0.63	0.77	0.61	0.54				
1993			0.71	0.85	0.69				
1994			0.67	0.71	0.86				
1996						0.85	0.70	0.65	0.61
1997						0.68	0.84	0.68	0.63
1998						0.63	0.68	0.84	0.67
1999						0.60	0.63	0.68	0.83

Table 9: Annual Earnings at a Job by Demographic and Education Groups								
			Average Annual Earnings				Standard Deviation	
Demographic		N	SIPP	DER	DIFF	%of SIPP	SIPP	DER
Group	Education Level		(1)	(2)	(3)	(4)	(5)	(6)
white male	no high school	20,023	12,204	14,217	2,013.7	16.5%	12,736	15,816
	high school	46,175	19,953	23,468	3,515.7	17.6%	17,254	21,134
	some college	46,870	21,351	24,887	3,536.2	16.6%	22,863	31,646
	college degree	21,658	36,979	45,640	8,660.9	23.4%	39,136	72,317
	graduate degree	17,993	43,998	58,640	14,641.8	33.3%	41,487	229,519
white female	no high school	14,815	6,956	7,695	738.1	10.6%	7,611	8,675
	high school	45,084	12,530	13,993	1,462.9	11.7%	12,381	13,078
	some college	50,330	13,347	14,852	1,505.2	11.3%	13,958	15,687
	college degree	21,004	21,395	23,980	2,584.9	12.1%	20,045	23,918
	graduate degree	15,804	27,348	30,833	3,485.4	12.7%	27,035	41,567
non-white male	no high school	3,362	10,792	13,064	2,272.2	21.1%	10,316	13,100
	high school	6,605	15,738	18,652	2,913.9	18.5%	13,978	16,006
	some college	6,300	17,080	19,876	2,796.0	16.4%	17,117	19,095
	college degree	2,319	28,926	35,251	6,324.9	21.9%	26,319	61,885
	graduate degree	2,031	36,680	44,314	7,634.5	20.8%	31,895	54,168
non-white female	no high school	3,685	7,800	8,932	1,131.9	14.5%	7,658	9,097
	high school	7,767	11,452	13,301	1,849.0	16.1%	10,107	12,961
	some college	8,960	13,247	15,179	1,932.9	14.6%	13,059	14,344
	college degree	3,216	22,126	25,626	3,499.7	15.8%	19,986	21,229
	graduate degree	2,033	28,448	33,606	5,158.8	18.1%	22,654	26,357

An observation is an annual earnings report for a person-job match.
There are multiple earnings reports for person-job matches that spanned multiple years.
Earnings are reported in real 1999 dollars using the annual average CPI-U to deflate.

Table 10: Industrial, Occupational and Proxy Interview Differences for Individuals with a Graduate Degree, by Sex, Whites Only				
	White male		*White female*	
Total person-job-year obs	17,993		15,804	
Panel A: Percentage in Industrial Categories				
Percentage in industry categories				
Agriculture	0.72		0.51	
Mining	0.48		0.05	
Construction	1.49		0.27	
Manufacturing Nondurable	5.21		2.40	
Manufacturing Durable	10.34		2.28	
Transp., Comm., Public Ut.	4.55		2.02	
Wholesale Trade	2.81		1.16	
Retail Trade	5.74		6.06	
Finance, Insur., Real Estate	6.63		4.13	
Business & Repair Services	4.65		3.28	
Personal Services	0.94		0.84	
Entertain. & Recreation Ser.	1.32		1.30	
Professional Services	44.95		71.20	
Public Administration	9.15		4.32	
missing industry	1.04		0.16	
Panel B: Percentage in Occupational Categories				
Executive, Administrative, Managerial	24.77		14.89	
(financial)		1.35		0.47
(education management)		2.68		2.61
Professional Specialties - Math/Science	12.83		3.08	
Health	5.61		10.84	
(doctor/dentist)		3.40		1.42
Teachers - including post-secondary	19.05		38.04	
Professional Specialties - Social Science	1.56		1.80	
Social Workers/Clergy	3.08		3.73	
Lawyers/Judges	4.52		2.41	
Writers,Artists,Entertainment,Athletes	2.27		2.76	
Technicians,Related Support	3.79		3.25	
Sales (including FIRE)	7.31		5.08	
(financial)		0.69		0.22
Administrative Support	4.72		8.95	
Service	3.87		3.76	
Farm,Forestry,Fishing	0.51		0.21	
Precision Production,Craft,Repair	2.35		0.24	
Operators,Fabricators,Laborers	2.62		0.76	
Military	1.14		0.19	
Panel C: Proxy Interviews				
Interviewed by proxy during the year	53.87		37.67	
Months of proxy response during the year				
(conditional on ever having proxy response)				
0-4 months	30		40	
5-8 months	27		30	
9-12 months	43		30	

Table 11: Job Changers - Earnings Comparisons at Old and New Jobs

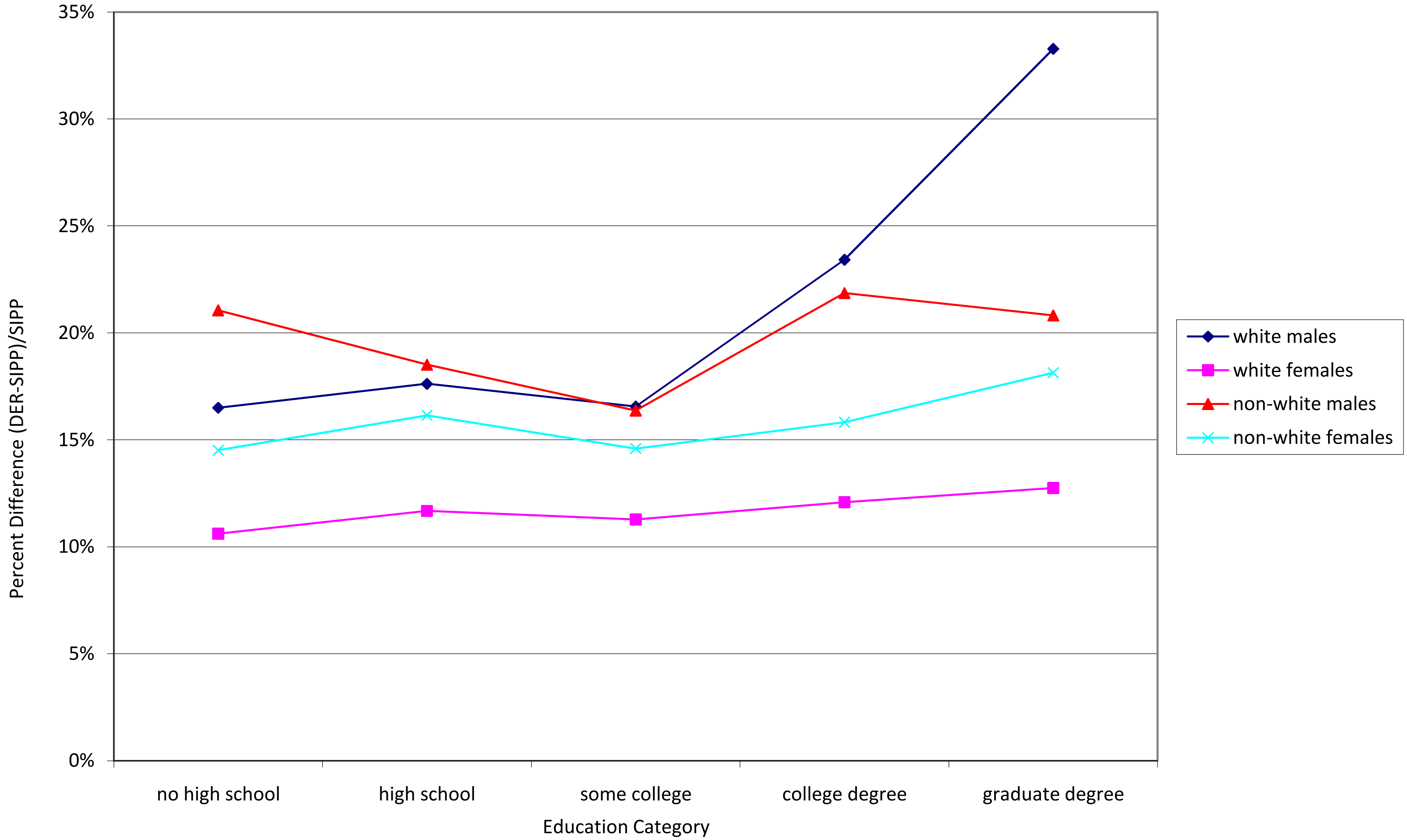
		SIPP			*DER*		
	N	Earn Job1	Earn Job2	**Change**	Earn Job1	Earn Job2	**Change**
OVERALL	26,241	8,265	8,307	**41.5**	9,685	8,769	**-916.3**
Panel A: Respondent reported industry for job 1							
Agriculture	481	4,687	5,399	**711.9**	4,736	5,243	**506.4**
Mining	121	14,812	14,464	**-348.1**	19,363	14,923	**-4,439.8**
Construction	1,487	10,147	10,159	**11.4**	10,354	9,983	**-370.8**
Manufacturing Nondurable	1,457	10,383	9,625	**-758.0**	12,983	10,373	**-2,609.0**
Manufacturing Durable	2,036	12,724	11,024	**-1,699.9**	16,350	12,117	**-4,232.3**
Transp., Comm., Public Ut.	1,104	11,745	11,036	**-709.0**	14,143	11,777	**-2,365.6**
Wholesale Trade	936	12,145	11,536	**-609.1**	14,277	12,080	**-2,197.4**
Retail Trade	8,154	4,447	5,063	**616.5**	5,226	5,219	**-6.6**
Finance, Insur., Real Estate	1,270	13,683	13,249	**-433.8**	16,720	14,811	**-1,908.6**
Business & Repair Services	2,212	8,064	9,022	**957.9**	8,701	9,337	**636.2**
Personal Services	886	5,338	5,381	**43.3**	5,490	5,243	**-246.9**
Entertain. & Recreation Ser.	666	5,137	6,078	**941.0**	5,580	6,586	**1,005.9**
Professional Services	4,798	9,336	9,499	**162.5**	10,712	10,051	**-661.2**
Public Administration	555	11,997	10,542	**-1,454.4**	14,672	12,224	**-2,447.5**
Panel B: Respondent reported Industry for job 2							
Agriculture	422	5,146	5,312	**166.3**	5,339	5,053	**-286.0**
Mining	115	16,482	16,399	**-83.0**	18,910	17,564	**-1,345.5**
Construction	1,624	9,674	10,127	**453.1**	10,193	9,948	**-245.8**
Manufacturing Nondurable	1,349	9,335	9,997	**661.9**	11,089	11,265	**176.8**
Manufacturing Durable	2,039	11,456	12,120	**664.1**	14,139	13,314	**-825.2**
Transp., Comm., Public Ut.	1,279	10,069	10,575	**505.7**	11,913	11,303	**-609.8**
Wholesale Trade	953	10,854	12,140	**1,285.4**	13,194	13,058	**-135.5**
Retail Trade	7,307	4,921	4,457	**-464.8**	5,895	4,585	**-1,310.2**
Finance, Insur., Real Estate	1,354	12,221	13,794	**1,574.0**	14,668	15,300	**632.6**
Business & Repair Services	2,646	9,052	8,495	**-556.6**	10,628	8,432	**-2,195.3**
Personal Services	818	5,657	5,155	**-502.5**	6,002	5,537	**-465.2**
Entertain. & Recreation Ser.	620	6,034	5,964	**-70.6**	6,442	6,125	**-317.6**
Professional Services	5,114	9,217	8,936	**-281.4**	10,659	9,516	**-1,143.1**
Public Administration	582	9,866	11,335	**1,469.7**	11,692	11,818	**125.7**
Panel C: Did respondent switch Industries between job 1 and job 2?							
No switch	11,494	9,348	9,416	**67.8**	11,041	10,166	**-875.0**
Switch	14,656	7,371	7,430	**59.0**	8,555	7,671	**-884.5**

An observation is an individual who reports two consecutive, non-overlapping jobs during the SIPP survey.
Only one job-switch per person is included in the table.
Earnings are last annual earnings for job 1 and first annual earnings for job 2.
Earnings are reported in real 1999 dollars using the annual average CPI-U to deflate.

Table 12: Earnings Regressions Comparisons					
		SIPP	DER	SIPP	DER
FIXED EFFECTS		Coefficients		Standard Errors	
non-white female	lf exp years 1-2	0.067	0.009	0.0170	0.0287
	lf exp years 3-5	0.044	0.068	0.0072	0.0119
	lf exp years 6-10	0.028	0.030	0.0037	0.0059
	lf exp years 11-25	0.011	0.014	0.0011	0.0018
	lf exp years 25+	-0.004	-0.001	0.0011	0.0018
white female	lf exp years 1-2	0.077	0.053	0.0071	0.0119
	lf exp years 3-5	0.049	0.053	0.0031	0.0051
	lf exp years 6-10	0.036	0.034	0.0016	0.0026
	lf exp years 11-25	0.011	0.015	0.0005	0.0008
	lf exp years 25+	-0.008	-0.005	0.0005	0.0007
non-white male	lf exp years 1-2	0.100	0.036	0.0191	0.0323
	lf exp years 3-5	0.055	0.076	0.0082	0.0134
	lf exp years 6-10	0.038	0.045	0.0042	0.0067
	lf exp years 11-25	0.012	0.017	0.0013	0.0020
	lf exp years 25+	-0.006	-0.004	0.0011	0.0017
white male	lf exp years 1-2	0.118	0.063	0.0074	0.0125
	lf exp years 3-5	0.062	0.086	0.0033	0.0054
	lf exp years 6-10	0.053	0.054	0.0017	0.0027
	lf exp years 11-25	0.019	0.022	0.0005	0.0008
	lf exp years 25+	-0.010	-0.008	0.0004	0.0006
non-white female	high school	0.124	0.158	0.0229	0.0145
	some college	0.276	0.293	0.0225	0.0143
	college degree	0.623	0.645	0.0287	0.0183
	graduate degree	0.790	0.863	0.0320	0.0205
white female	high school	0.164	0.188	0.0111	0.0070
	some college	0.288	0.298	0.0110	0.0069
	college degree	0.596	0.617	0.0131	0.0083
	graduate degree	0.748	0.767	0.0139	0.0088
non-white male	high school	0.175	0.137	0.0244	0.0154
	some college	0.259	0.241	0.0248	0.0156
	college degree	0.568	0.551	0.0319	0.0203
	graduate degree	0.760	0.728	0.0327	0.0208
white male	high school	0.163	0.176	0.0101	0.0064
	some college	0.247	0.248	0.0101	0.0064
	college degree	0.561	0.586	0.0120	0.0076
	graduate degree	0.677	0.700	0.0125	0.0079
non-white female		-0.005	-0.008	0.0518	0.0303
white female		-0.029	-0.030	0.0287	0.0169
non-white male		-0.001	0.016	0.0569	0.0333
linear time effect		0.021	0.033	0.0014	0.0008
panel1993		0.083	0.139	0.0082	0.0051
panel1992		0.100	0.173	0.0083	0.0051
panel1991		0.130	0.233	0.0106	0.0065
panel1990		0.151	0.253	0.0110	0.0067
intercept		1.266	1.250	0.0235	0.0138
Number of observations		1,641,180	1,617,320		
RANDOM EFFECTS					
person intercept		0.180	0.540		
person slope		0.076	0.180		
correlation		-0.099	-0.270		
firm effect		0.069	0.210		
variance of residual		0.160	0.400		
AR1 correlation coefficient of residual		0.500	0.450		

Table 13: Comparison of Unobservable Heterogeneity: Random Effects	
Person effect	
main	0.2814
SIPPdev	0.0000
DERdev	0.0367
Employer (firm) effect	
main	0.3163
SIPPdev	0.0000
DERdev	0.1013
Common time period	
variance	0.7486
AR1 correlation	0.5737
Residual	
SIPP variance	0.2171
DER variance	0.3753
SIPP AR1 correlation	0.2705
DER AR1 correlation	0.6365
Reliability ratio	
SIPP	0.8612
DER1	0.7240
DER2	0.7982

Figure 1:
Percentage Differences in SIPP and DER Earnings



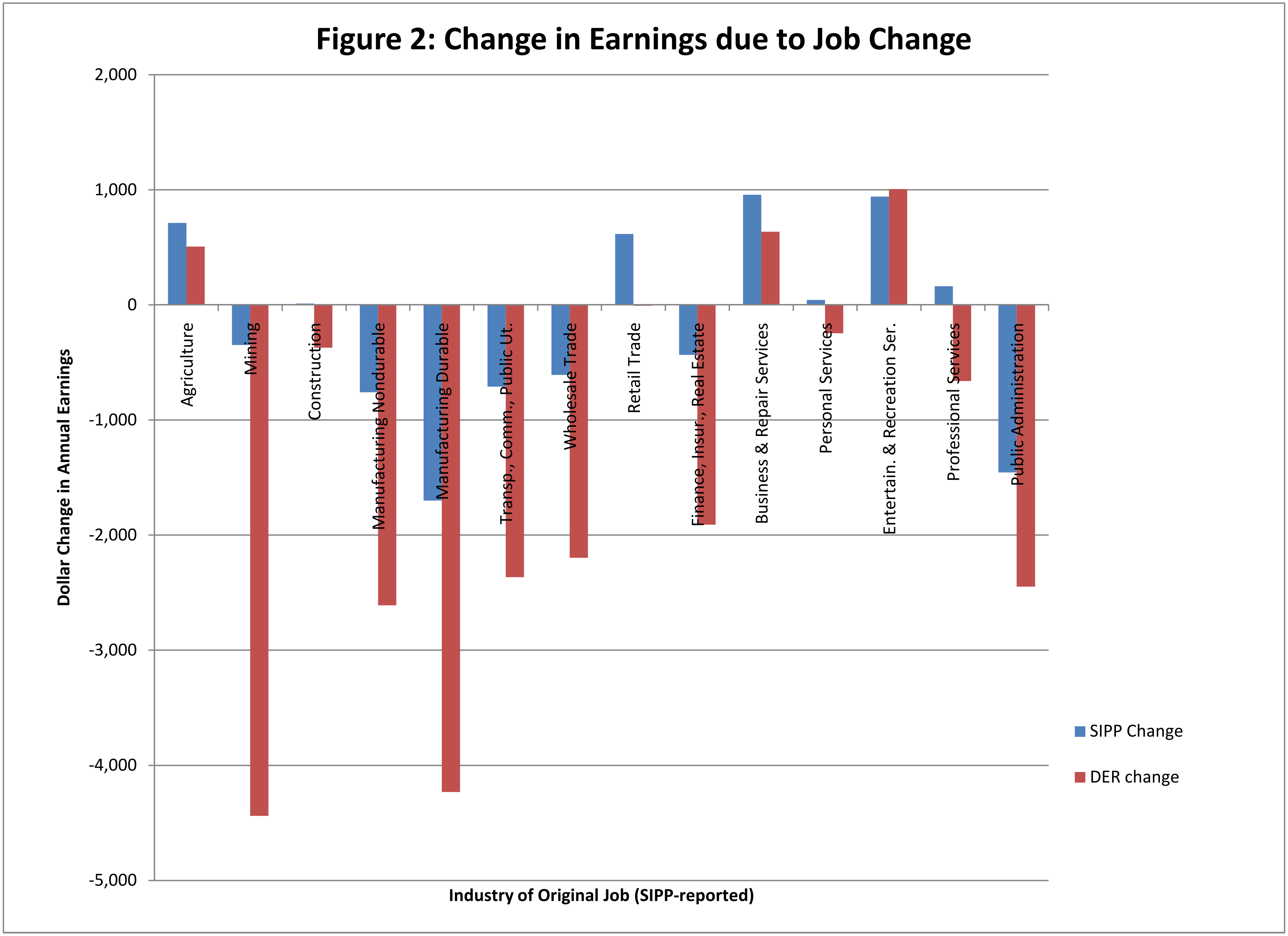


Figure 2: Change in Earnings due to Job Change
2,000
1,000
0
-1,000
-2,000
-3,000
-4,000
-5,000
Dollar Change in Annual Earnings
Agriculture
Mining
Construction
Manufacturing Nondurable
Manufacturing Durable
Transp., Comm., Public Ut.
Wholesale Trade
Retail Trade
Finance, Insur., Real Estate
Business & Repair Services
Personal Services
Entertain. & Recreation Ser.
Professional Services
Public Administration
Industry of Original Job (SIPP-reported)
SIPP Change
DER change

Figure 3:
Labor Force Experience Profiles for
White Men and Women

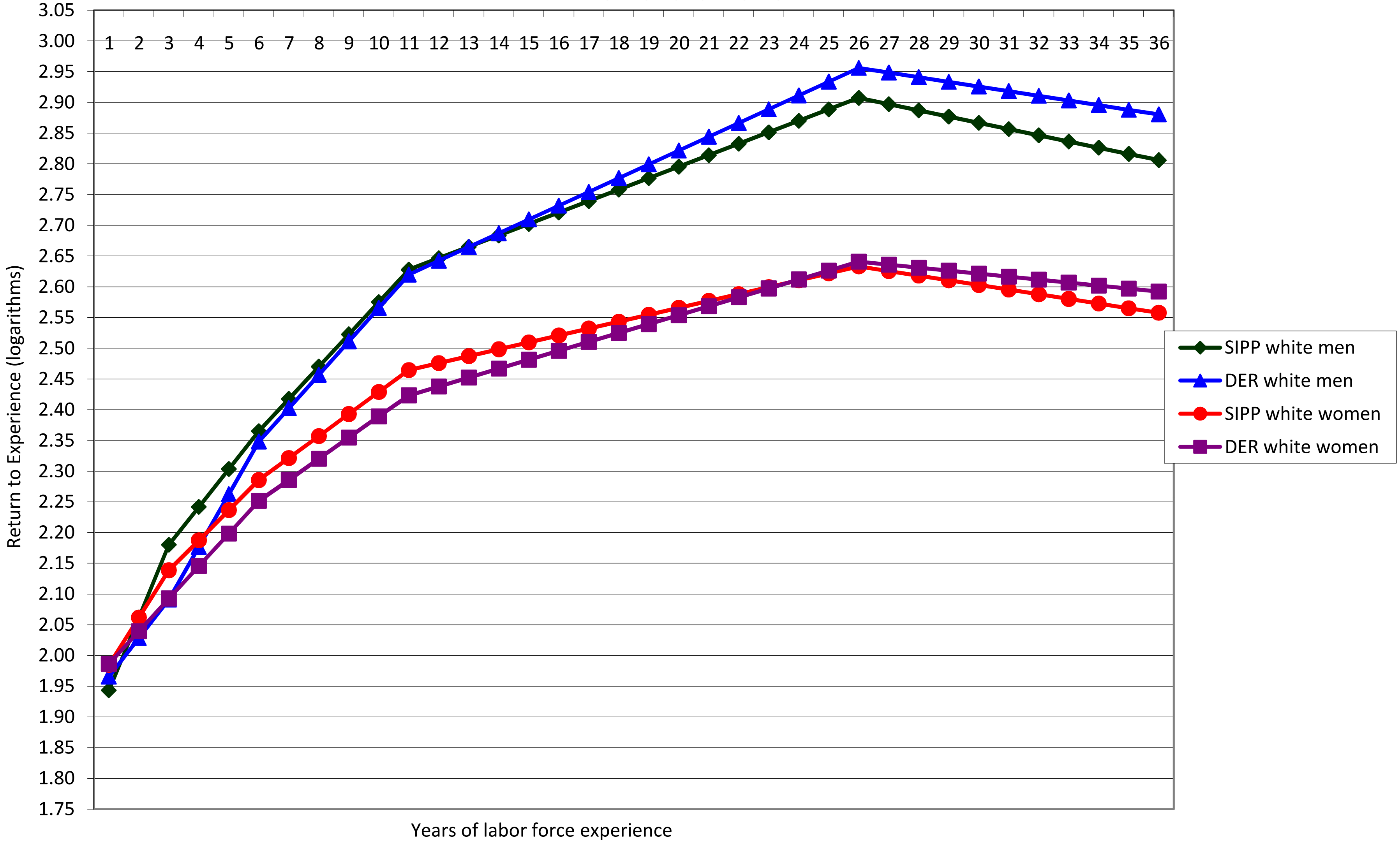


Figure 4:
Labor Force Experience Profiles, Job Level for Non-white Men and Women

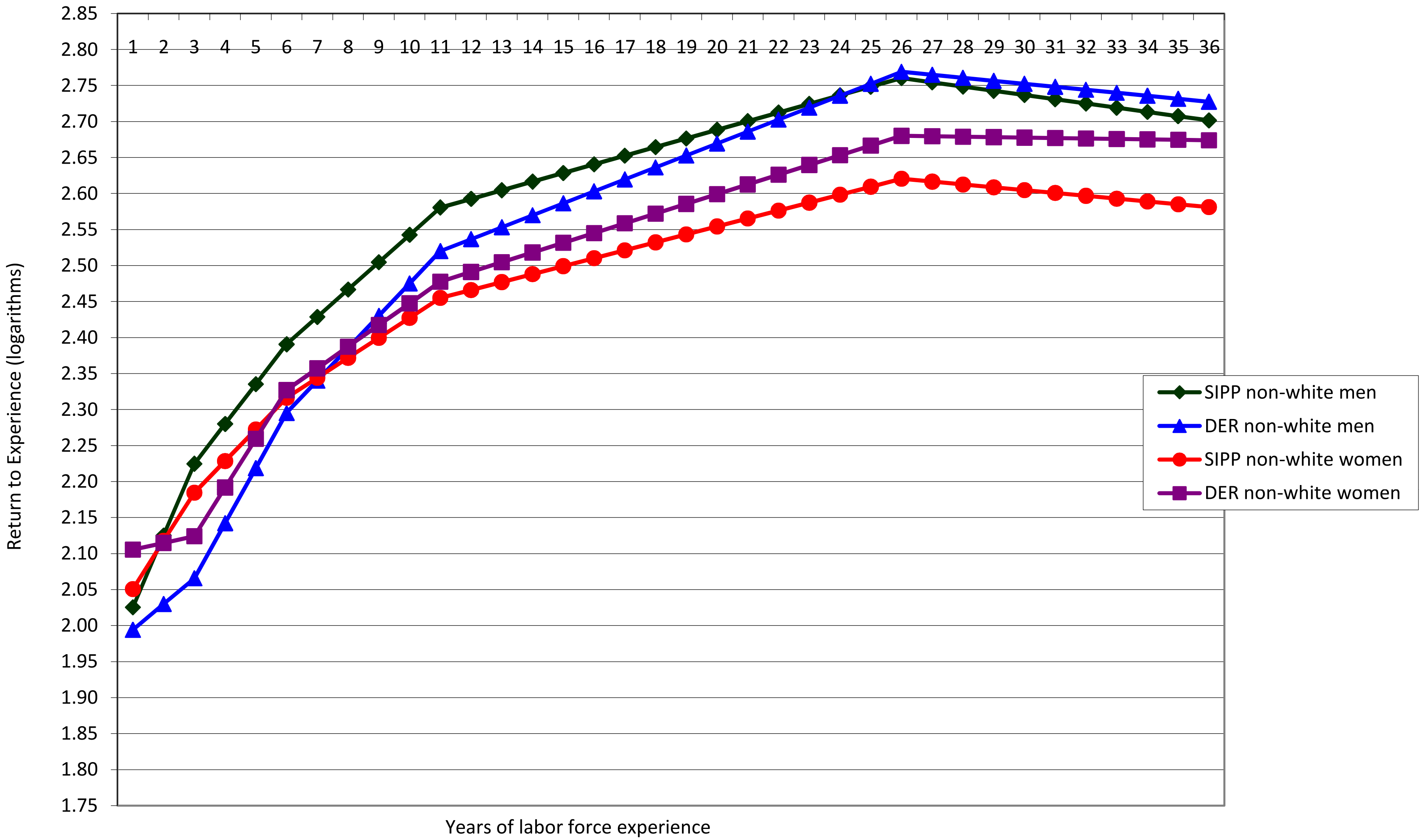


Table A1: SIPP Job ID Problems, 1996

Wave	Start date	Jobid
1	Feb. 1, 1996	1
2	Feb. 1, 1996	1
3	Feb. 1, 1996	1
4	Feb. 1, 1996	1
5		
6	Jan. 1, 1996	2

Table A2: SIPP Job ID Problems, 1990-1993 Panels

Failure to link job across waves		
Wave	Firm Name	SIPP Jobid
1	AAAA	1
2	AAAA	1
3	AAAA	2
4	AAAA	1

Table A3: SIPP Job ID Problems, 1990-1993 Panels

Failure to separate jobs across waves		
Wave	Firm Name	SIPP Jobid
1	AAAA	1
2	AAAA	1
3	AAAA	1
4	BBBB	1

Table A4:
First Round of Job Name Matching, SIPP Panels 1990-1993
Description of Person-Job-Wave observation matching by pass

	Blocking Variables	Matching Variables	m, u prob.	cutoffs
Pass 1	Person ID	Full employer name*	.9, .02	2, .3
Pass 2	Person ID	Fields from employer name*:		2, .3
		word one	.9, .15	
		word two	.95, .6	
		word three	.95, .6	
		word four	.95, .6	
		qualifier word one	.95, .6	
		qualifier word two	.95, .6	
		type word one	.95, .6	
		type word two	.95, .6	
		SIPP original job id number	.6, .5	
Pass 3	Person ID	Fields from employer name*:		5, .05
		Array: first 4 words	.95, .6	
		word one	.9, .15	
		qualifier word one	.95, .6	
		qualifier word two	.95, .6	
		type word one	.95, .6	
		type word two	.95, .6	

*Jobs with missing names were excluded from this round of name matching.

Table A4 (continued):
Second Round of Job Name Matching, SIPP Panels 1990-1993
Description of Person-Job-Wave observation matching by pass

	Blocking Variables	Matching Variables	m, u prob.	cutoffs
Pass 1	Person ID	Fields from employer name:		.1, .1
		full name**		
		array: first 4 words	.9, .1	
		Array: first 2 qualifier words	.9, .15	
		Array: first 2 type words	.9, .15	
		Geo word	.7, .5	
Pass 2	Person ID	Full employer name**	.95, .1	.1, .1

Job Name Matching, SIPP Panel 1996
Description of Person-Job record matching by pass***

	Blocking Variables	Matching Variables	m, u prob.	cutoffs
Pass 1	Person ID	Full employer name**		2, .15
Pass 2	Person ID	Fields from employer name:		2, .15
		full name**		
		word one	.9, .15	
		word two	.95, .6	
		word three	.95, .6	
		word four	.95, .6	
		qualifier word one	.95, .6	
		qualifier word two	.95, .6	
		type word one	.95, .6	
		type word two	.95, .6	
		geo word	.7, .5	
Pass 3	Person ID	Fields from employer name:		2, .15
		full name**		
		Array: first 4 words	.95, .6	
		word one	.9, .15	
		qualifier word one	.95, .6	
		qualifier word two	.95, .6	
		type word one	.95, .6	
		type word two	.95, .6	
		geo word	.7, .5	

**When no weights were assigned, complete employer name was included but given zero weight unless it was blank and then the full disagreement weight was assigned. This was used to prevent jobs with blank names from matching. If weights were assigned, full disagreement weight was also assigned if name was missing.
***Job records with observations in the same wave were disqualified from matching to each other because the same job could not be reported on twice in the same wave.

Table A5: Summary of the Job ID Editing Process, 1990-1993 SIPP Panels					
	SIPP Panel	1990	1991	1992	1993
1	Number of Jobs, post name matching pass 1	78,225	46,316	74,078	68,803
2	Jobs belonging to people with conflict with DER	45,725	24,149	38,752	35,352
3	Number of Jobs, post name matching pass 2	69,138	41,814	66,602	62,251
4	Jobs belonging to people with conflict with DER	10,011	5,106	8,131	8,330
5	Number of Jobs, post clerical edits	66,991	40,818	65,278	61,094
6	Jobs belonging to people with conflict with DER	7,089	3,800	6,448	6,670

Table A6: Summary of Job ID editing process, 1996 SIPP Panel		
	SIPP Panel	1996
1	Jobs with startdate probs	29,520
2	Number of Jobs, post name matching	125,358
3	Jobs belonging to people with conflict with DER	15,331
4	Jobs with startdate probs	22,353

Table B1: DER Match to the Business Register								
			Match to					
SIPP Panel		DER Total	Business Register		Single-Unit File		Multi-Unit File	
1990	EINs	60,131	58,991	98.10%	58,255	96.88%	16,990	28.25%
	Jobs	96,086	93,520	97.33%	92,379	96.14%	37,807	39.35%
1991	EINs	38,628	38,096	98.62%	37,686	97.56%	12,497	32.35%
	Jobs	58,020	56,725	97.77%	56,118	96.72%	24,526	42.27%
1992	EINs	62,406	61,391	98.37%	60,777	97.39%	19,361	31.02%
	Jobs	99,524	96,982	97.45%	96,029	96.49%	43,197	43.40%
1993	EINs	51,880	50,839	97.99%	50,376	97.10%	17,029	32.82%
	Jobs	81,320	78,933	97.06%	78,198	96.16%	35,977	44.24%
1996	EINs	105,095	95,122	90.51%	94,438	89.86%	28,923	27.52%
	Jobs	192,720	172,832	89.68%	171,585	89.03%	82,546	42.83%

Table C1:
Description of SIPP Job to DER Job Matching Algorithm by Pass

	Blocking Variables	Matching Variables	m, u prob.	cutoffs
Pass 1	Person ID	Fields from SU name:		2, .3
		Array: first 4 words	.95, .1	
		Array: first 2 qualifier words	.9, .3	
		Array: first 2 type words	.9, .3	
		Geo word	.7, .5	
		year indicators*	.75, .3	
		Complete SU name**		
Pass 2	Person ID	Fields from MU name:		2, .3
		Array: first 4 words	.95, .1	
		Array: first 2 qualifier words	.9, .3	
		Array: first 2 type words	.9, .3	
		Geo word	.7, .5	
		year indicators*	.75, .3	
		Complete MU name**		
Pass 3	Person ID	year indicators*	.9, .3	2, .3
	3-digit SU Industry			
Pass 4	Person ID	year indicators*	.9, .3	2, .3
	3-digit MU Industry			
Pass 5	Person ID	start year***	.9, .3	2, .3
	3-digit SU Industry			
Pass 6	Person ID	year indicators*	.9, .3	2, .3
	1-digit SU Industry			
Pass 7	Person ID	year indicators*	.9, .1	2, .3
		3-digit SU Industry	.9, .1	2, .3

*Year Indicators by Panel
1990: 1990, 1991, 1992
1991: 1991, 1992, 1993
1992: 1992, 1993, 1994, 1995
1993: 1993, 1994, 1995
1996: 1996, 1997, 1998, 1999

**Complete employer name was included but given zero weight unless it was blank and then the full disagreement weight was assigned. This was used to prevent jobs with blank names from matching in the first 2 passes.

***Start year was first year during survey time frame when job was observed in the SIPP or DER.

Table C2: SIPP Jobs matched to DER Jobs

		1990		1991		1992		1993		1996	
		SIPP Jobs	DER Jobs	SIPP Jobs	DER Jobs	SIPP Jobs	DER Jobs	SIPP Jobs	DER Jobs	SIPP Jobs	DER Jobs
1	Master Match	37,207	37,207	22,513	22,513	34,593	34,593	32,289	32,289	66,366	66,366
2	Clerical Match	4,678	4,678	2,745	2,745	5,136	5,136	4,180	4,180	8,476	8,476
3	Duplicate Match on SIPP side	529		294		490		436		971	
4	Duplicate Match on DER side		907		609		1,007		771		1,920
5	Total Matches	42,414	42,792	25,552	25,867	40,219	40,736	36,905	37,240	75,813	76,762
6	Match Rate	76.99%	48.45%	78.75%	48.99%	77.87%	45.08%	77.33%	50.11%	78.04%	44.21%
7	Percent of Matches that are Master	87.72%	86.95%	88.11%	87.03%	86.01%	84.92%	87.49%	86.71%	87.54%	86.46%
8	Residual Job (non-match)	12,673	45,532	6,895	26,930	11,431	49,624	10,818	37,077	21,336	96,861
9	Total Jobs	55,087	88,324	32,447	52,797	51,650	90,360	47,723	74,317	97,149	173,623

Table C3: Example of Duplicate Match on SIPP Side

Type of Match	DER EIN	SIPP Jobnum
Master Match	A	1
Duplicate Match on SIPP side	A	2

Table C4: Example of Duplicate Match on DER Side

Type of Match	DER EIN	SIPP Jobnum
Master Match	A	1
Duplicate Match on DER side	B	1